

## Small non-coding RNA profiling in human biofluids and surrogate tissues from healthy individuals: description of the diverse and most represented species

Giulio Ferrero<sup>1,2,\*</sup>, Francesca Cordero<sup>1,3,\*</sup>, Sonia Tarallo<sup>3</sup>, Maddalena Arigoni<sup>4</sup>, Federica Riccardo<sup>4</sup>, Gaetano Gallo<sup>5,6</sup>, Guglielmo Ronco<sup>7</sup>, Marco Allasia<sup>8</sup>, Neha Kulkarni<sup>4</sup>, Giuseppe Matullo<sup>3,9</sup>, Paolo Vineis<sup>3,10</sup>, Raffaele A. Calogero<sup>4</sup>, Barbara Pardini<sup>3,9,\*</sup> and Alessio Naccarati<sup>3,11,\*</sup>

<sup>1</sup>Department of Computer Science, University of Turin, Turin, Italy

<sup>2</sup>Department of Clinical and Biological Sciences, University of Turin, Turin, Italy

<sup>3</sup>Italian Institute for Genomic Medicine, IIGM (formerly Human Genetics Foundation, HuGeF), Turin, Italy

<sup>4</sup>Molecular Biotechnology Center, Department of Biotechnology and Health Sciences, University of Turin, Turin, Italy

<sup>5</sup>Department of Medical and Surgical Sciences, University of Catanzaro, Catanzaro, Italy

<sup>6</sup>Department of Colorectal Surgery, Clinica S. Rita, Vercelli, Italy

<sup>7</sup>Center for Cancer Epidemiology and Prevention, AO City of Health and Science, Turin, Italy

<sup>8</sup>Department of Surgical Sciences, University of Turin and Città della Salute e della Scienza, Turin, Italy

<sup>9</sup>Department of Medical Sciences, University of Turin, Turin, Italy

<sup>10</sup>MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom

<sup>11</sup>Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Prague, Czech Republic

\*These authors contributed equally to this work

**Correspondence to:** Alessio Naccarati, **email:** alessio.naccarati@iigm.it

**Keywords:** next-generation sequencing; small non-coding RNA profiling; microRNAs; non-invasive biomarkers; surrogate tissues

**Received:** September 06, 2017

**Accepted:** November 15, 2017

**Published:** December 14, 2017

**Copyright:** Ferrero et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

The role of non-coding RNAs in different biological processes and diseases is continuously expanding. Next-generation sequencing together with the parallel improvement of bioinformatics analyses allows the accurate detection and quantification of an increasing number of RNA species. With the aim of exploring new potential biomarkers for disease classification, a clear overview of the expression levels of common/unique small RNA species among different biospecimens is necessary. However, except for miRNAs in plasma, there are no substantial indications about the pattern of expression of various small RNAs in multiple specimens among healthy humans.

By analysing small RNA-sequencing data from 243 samples, we have identified and compared the most abundantly and uniformly expressed miRNAs and non-miRNA species of comparable size with the library preparation in four different specimens (plasma exosomes, stool, urine, and cervical scrapes).

Eleven miRNAs were commonly detected among all different specimens while 231 miRNAs were globally unique across them. Classification analysis using these miRNAs provided an accuracy of 99.6% to recognize the sample types. piRNAs and tRNAs were the most represented non-miRNA small RNAs detected in all specimen types that were analysed, particularly in urine samples. With the present data, the

**most uniformly expressed small RNAs in each sample type were also identified. A signature of small RNAs for each specimen could represent a reference gene set in validation studies by RT-qPCR.**

**Overall, the data reported hereby provide an insight of the constitution of the human miRNome and of other small non-coding RNAs in various specimens of healthy individuals.**

## INTRODUCTION

The discovery of many stable extracellular small RNAs has changed our view of gene expression regulation, including the role that these molecules may play in several complex processes previously partially understood such as cell-to-cell communication [1]. In this respect, with an astonishing number of publications in the last decade, microRNAs (miRNAs) represent the most explored small non-coding RNA (sncRNA) species in humans [2]. A large number of studies has demonstrated that cellular and extracellular miRNA altered expression is associated with a wide variety of diseases, including cancer [3, 4]. However, little is known about the presence within the same matrix of other common species of sncRNAs such as piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), tRNAs etc. All these versatile RNA species are known to be key components of molecular interactions and gene regulation in eukaryotes [5].

The field of circulating extracellular RNA molecules is rapidly growing thanks to the implementation of Next-Generation Sequencing (NGS) technologies and bioinformatics solutions that analyze the huge amount of data released from sequencing. With such high-throughput approach, all extracellular RNAs can be quantified and tested as potential sources of new diagnostic and therapeutic biomarkers in many different types of biological samples [6]. To achieve this, RNA-Sequencing (RNA-Seq) has emerged as a powerful tool in transcriptomics, gene expression profiling and biomarker discovery. Sequencing cell-free nucleic acids from liquid biopsies additionally provides exciting possibilities for molecular diagnostics, and might help establish disease-specific biomarker signatures [7]. Lower complexity, not known post-processing modifications, simple detection and amplification methods, tissue-restricted expression profiles, and sequence conservation between humans and model organisms make extracellular miRNAs and other sncRNAs ideal candidates for non-invasive biomarkers to reflect and study various physiopathological conditions in the body [8]. It is possible to extract and quantify high-quality sncRNAs from a wide range of cell and tissue sources, including cell lines, fresh and formalin-fixed paraffin-embedded tissues, plasma, serum, urine and other body fluids [8–10]. Despite this increasing interest, the field is still largely in an exploratory and descriptive phase. There are no standardized methods for sample collection, isolation, or analysis. There is also no general agreement on the terms for a good quality sample definition, and each specimen (body fluid or surrogate tissue) under

various disease/injury conditions are likely to have diverse contents and different criteria for quality assessment [7, 11]. A growing number of isolation methods for profiling circulating extracellular RNA molecules have been developed but still, there is no gold standard for the most efficient inclusive or selective protocols [6]. However, the complexity of the small RNA-Seq workflow bears challenges and biases that researchers need to be aware of, in order to generate high-quality data [12].

The creation of large repositories including data from different human specimens, isolation methods, detection platforms, and analysis tools is essential to increase our understanding of the extent and types of extracellular RNA material present in different body fluids/surrogate tissues. At present, there are few large datasets describing the extracellular contents in biofluid samples from healthy controls [13–17]. Besides, previous studies on extracellular sncRNAs have investigated very small numbers or pooled samples with the purpose of identifying a specific class of RNAs [18]. The largest investigations of samples focused almost exclusively on miRNAs, with the main limitation of measuring either only targeted miRNAs in large numbers of individuals or the whole known miRNome in very small populations. In a recent work it has been described the largest group of plasma-based miRNAs and the first broadest variety of extracellular (non-miRNA) sncRNAs in a large population [15]. In another similar work, authors profiled the small RNA (16–32 nts) payload of human biofluids by NGS. Extracellular RNAs were isolated from plasma, urine and saliva samples from 55 young male athletes and sequenced to establish a sncRNA pattern at steady state [6].

In the present study, we investigated pattern and expression levels of miRNAs and other sncRNAs of comparable size in four different biospecimens representing ideal surrogate tissues for diagnostic and screening programs. Specifically, we analysed data from small RNA-Seq from 125 plasma-derived exosomes, 48 urine, 31 cervical scrapes, and 39 stool samples collected from healthy subjects. For cervical scrapes and stool, this is the first study investigating sncRNAs by NGS. In addition, urine and stool samples were paired with those from plasma collected from the same subjects.

## RESULTS

### Overview of study samples and pipeline analysis

We analysed small RNA-Seq data of RNA extracted from exosomes from 125 plasma samples of healthy

donors derived from three different studies (respectively 39 for the Study 1, 46 for the Study 2, and 40 for the Study 3) (Materials and Methods). Additionally, sequencing was performed on RNA from 39 faecal samples (Study 1), 48 urine samples (Study 2), and cervical scrapes from 31 Human Papilloma Virus (HPV) negative women. Some of the plasma sample donors provided at the same occasion a sample of stool (39 from Study 1) or urine (46 from Study 2).

Total RNA was isolated from samples with specific kits for each type of specimens while library preparation for small RNA-Seq was performed adopting the same kit and protocol. Libraries were run at the same sequencing facility. Finally, all bioinformatics analyses (i.e. pre-processing of raw data) were performed following the same pipeline by the same operator.

To explore the landscape of sncRNA expression levels in different biospecimens, we designed a computational strategy for small RNA-Seq data analysis (Figure 1A). We updated the miRNA analysis pipeline published by our group [19] by adding a second phase focused on the analysis of small RNA-Seq reads unmapped against the human miRNome (Materials and Methods).

Initially, small RNA-Seq datasets were pre-processed and quality controlled to remove adapter sequences and low-quality reads. The processing information about the 243 datasets analysed is provided in Supplementary Table 1A and 1B. Quality check confirmed that there were no reads shorter than 15 nucleotides and the rate of low quality reads (quality score < 30) was on average below 8%, with urine and stool samples providing the best rates (<1%).

### Identification of miRNAs and non-miRNA sncRNAs

miRNA mapping analysis showed remarkable differences among specimens for read alignment rates (Figure 1B, 1C and Supplementary Figure 1A). Consistently with the highest rates of read alignment (Figure 1C), urine samples were generally associated with a high number of reads (median reads = 12.38 million) followed by plasma exosomes (median reads = 11.34 million), stool (median reads = 4.88 million), and cervical scrapes (median reads = 4.13 million) (Supplementary Figure 2A).

Datasets from plasma exosome and urine samples were characterized by the highest miRNA alignment rates (16.3% and 11.0%, of reads aligned, respectively) while datasets from stool and cervical scrape samples were associated on average with low miRNome alignment rates (0.7% and 1.2% of reads aligned, respectively).

Of the 1,823 miRNA annotations from miRBase, a range from 19.9% (cervical scrapes) to 73.8% (plasma exosomes *study 2*) of human miRNAs were detected in all the investigated specimens. A median of 58.61% of miRBase annotations were detected across the four

specimen types. Specifically, miR-486-5p was the most expressed miRNA in plasma exosomes samples (median reads = 180,173 reads) while miR-320a (median reads = 198 reads), miR-6813-5p (median reads = 5,911 reads), and miR-30a-5p (median reads = 25,910 reads) were the highest expressed in cervical scrapes, stool, and urine, respectively (Supplementary Figure 2B).

Since a large fraction of sequencing reads did not map on miRNome (Supplementary Table 1A), the alignment analysis was extended to other candidate sncRNA annotations by initially remapping reads on the human genome. Then, mapped reads were assigned to sncRNA annotations quantifiable using our size selection criterion. These annotations included sncRNAs annotated in GENCODE v24 database [20] (transcript length  $\leq 70$  bp) as well as piRNA (average length  $31 \pm 1$  bp) and tRNA (average length  $74 \pm 7$  bp) species annotated in the Database of Small Human non-coding RNAs (DASHR) release 1 [21] (Supplementary Table 1C). The alignment rates observed were higher for cervical scrapes (88.4%) followed by urine (81.1%), and plasma exosome samples (69.5%). As expected, stool datasets were associated with the lowest alignment rate on the human genome (28.1%) consistently with the presence of microbiome RNAs and other RNAs introduced by the diet, contributing to the large fraction of faecal RNA content (Supplementary Table 1A and Supplementary Figure 1A). In urine, most reads were assigned to piRNA (44.5%) or tRNA annotations (45.1%). Conversely, in the other specimens, a low assignment rate was observed ranging 1.8–3.4% for piRNAs and 1.0–3.3% for tRNAs, respectively (Supplementary Figure 1B). Homologous piRNAs annotated to different loci were associated with the same number of reads across samples.

### Common and specific miRNAs among different specimens

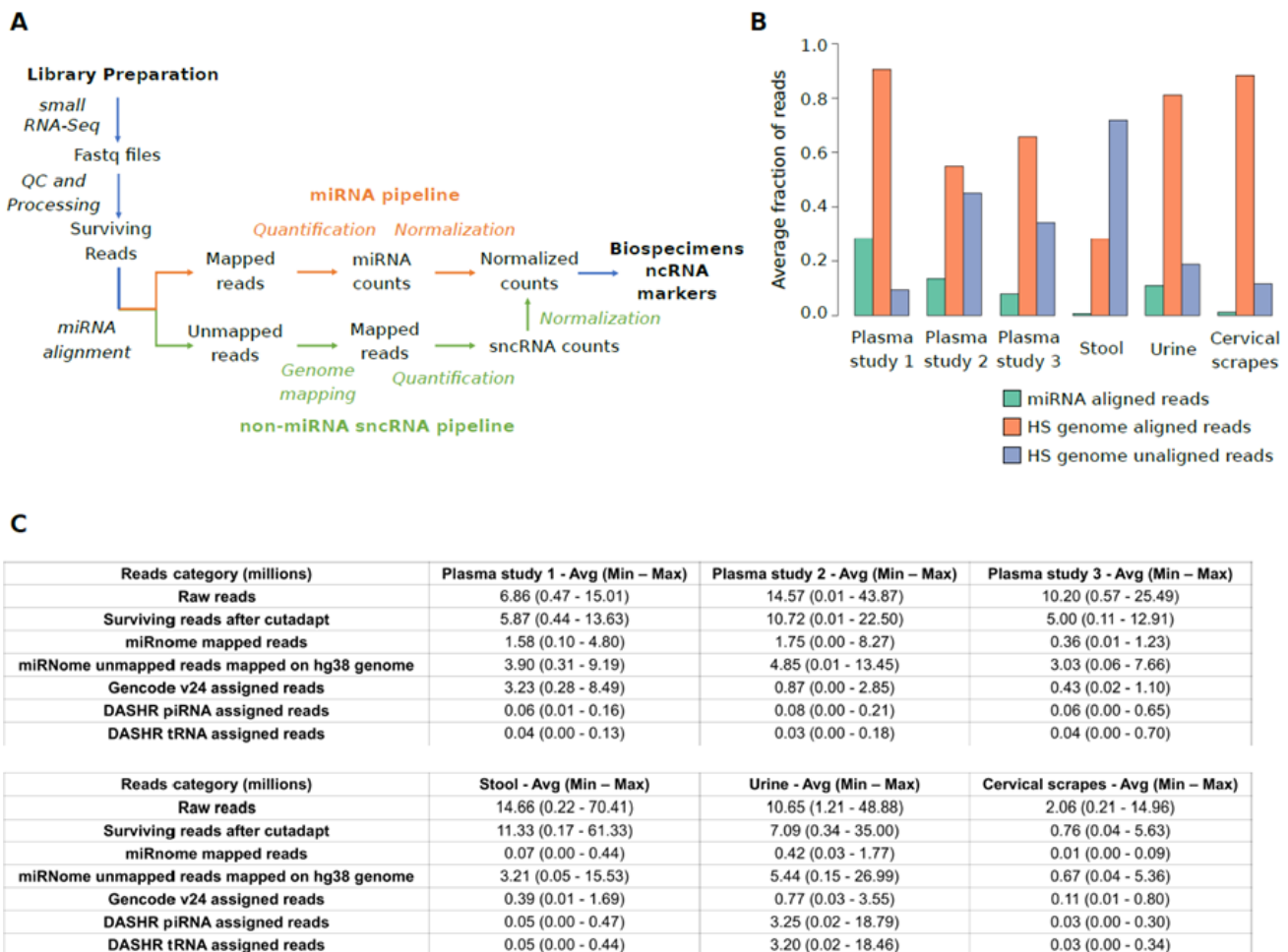
Considering the individual datasets from plasma exosome samples, it was evident a study-specific influence on the read alignment distribution with samples from the *study 1* characterized, on average, by the overall highest alignment on miRNome annotations (28.3% aligned reads). However, PCA on miRNAs and other sncRNA annotations expressed in at least one study (within study median number of reads >20) showed a distinct cluster formed by all plasma exosome samples with respect to other biospecimens (Supplementary Figure 1C). A comparable result was obtained by computing a pairwise correlation analysis: datasets from the three plasma exosome studies clustered together and were clearly separated from the others (Supplementary Figure 1D). Given the results from the PCA and correlation analyses, plasma exosome samples from the three studies were merged into a single group after read count correction with Surrogate Variable Analysis (SVA). The identification of

pattern of miRNAs detectable in the different specimens was performed by considering miRNAs characterized by a median of normalized reads higher than 20 in at least one specimen. Using this threshold, cumulatively, 394 miRNAs were quantified in at least one specimen (Figure 2A, Supplementary Table 2A). Eleven miRNAs were identified as commonly detectable in all types of specimens: miR-320a, miR-589-5p, miR-636, miR-1273a, miR-3960, miR-4419a, miR-4497, miR-4709-5p, miR-4792, miR-7641-1, and miR-7641-2.

Functional enrichment analysis of validated target genes of the 11 shared miRNAs revealed biological processes related to mRNA translation and transcription including translational initiation (GO:0006413,  $p = 1.9 \times 10^{-8}$ ) or positive regulation of transcription, DNA-templated (GO:0045893,  $p = 4.2 \times 10^{-7}$ ) (Supplementary Table 2B).

Plasma exosome samples were characterized by the highest number of specimen-specific miRNAs (155 miRNAs) followed by stool (55 miRNAs), urine (22 miRNAs), and cervical scrape samples (one miRNA) (Figure 2A, 2B). Considering only the specimen-specific miRNAs, miR-122-5p was the most expressed in plasma exosome samples (median reads = 32,512 reads) while miR-655-5p (median reads = 792 reads), miR-204-5p (median reads = 750 reads) and miR-4741 (median reads = 28 reads) were the most abundantly expressed in stool, urine, and cervical scrapes, respectively (Supplementary Figure 2C).

PCA analysis of the highly-expressed sets of miRNAs showed a good accuracy in the classification of different biospecimens (Figure 2C). To identify the discriminative miRNAs in the specimen classification, we also performed a classification and attribute selection



**Figure 1:** (A) Schematic representation of the computational pipeline applied in the analysis of small RNA-Seq dataset from healthy individuals. The modules of the pipeline designed for miRNAs and other sncRNAs are depicted in orange and green, respectively. (B) Bar plot showing for each specimen, the average number of sequencing reads aligned to miRNA annotations (green), unmapped on miRNA annotations but mapped on human genome (red), and unmapped on both miRNA annotations and the human genome (blue). (C) Table reporting the average, minimum, and maximum number of reads (in million) composing the starting datasets, aligned in the different analysis phases, or assigned to specific RNA annotations. HS= Homo sapiens.

analysis. Using a Random Forest classifier, we obtained an accuracy of 99.6% with only one sample incorrectly classified (Supplementary Table 2C). All the miRNAs analysed were associated with a high chi-square statistic (merit) in the attribute selection analysis with miR-204-5p, miR-5698, and miR-335-3p associated with the highest merit (Supplementary Table 2D).

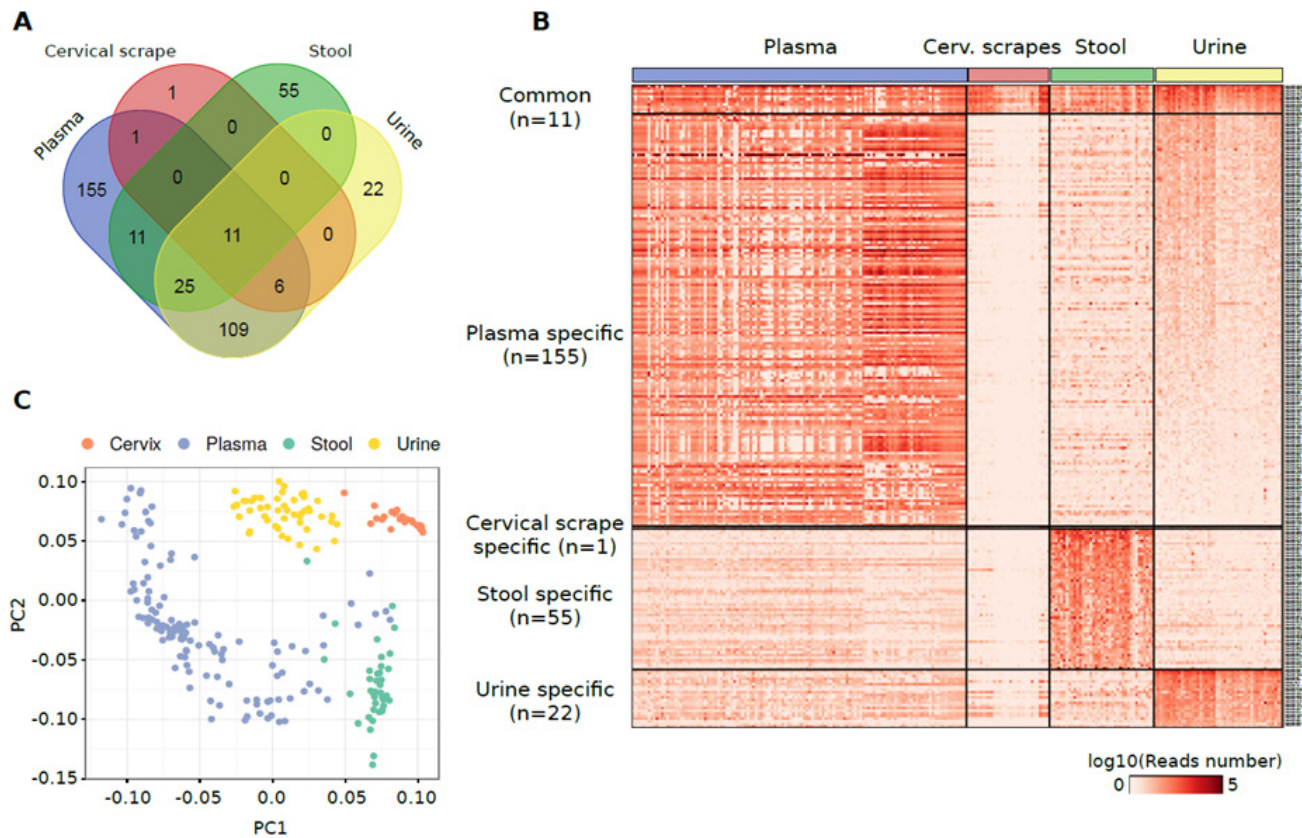
For a subset of patients, paired data from plasma and stool samples or from plasma and urine samples were available allowing a comparison between expression levels of sncRNAs in the different specimens from the same subject. As reported in Supplementary Table 2E, 2F, a low co-expression was generally observed either between plasma-stool or plasma-urine samples. The only exception was miR-3665 which was characterized by a positive correlation between plasma and urine samples ( $r = 0.59, p = 2.0 \times 10^{-5}$ ).

Prediction of candidate miRNA isoforms (isomiRs) was also performed using our datasets. As reported in Supplementary Table 2G, 832 isomiRs associated with more than 20 supporting reads in at least one specimen type were detected. Overall, 94.4% of isomiRs were detected in plasma exosome or urine samples consistently with the higher number of aligned reads in these samples.

The isomiRs with the highest number of supporting reads were a 3' variant of miR-486/miR-486-2 in plasma samples, a 5' variant of miR-934 in urine sample, a 5' variant of miR-7704 in cervical scrapes, and a 3' variant of miR-583 in stool samples. Among the previously identified 11 common miRNAs, eight were associated to an isomiR predicted in only one or two types of specimens (particularly in plasma or urine samples) (Supplementary Table 2H).

### Expression pattern of other sncRNAs

Cumulatively, 615 non-miRNA sncRNAs were quantified in at least one specimen. Of this set of annotations, 112 sncRNAs were commonly detected in all the analysed sample types (Figure 3A and Supplementary Table 3A). Coherently with the highest alignment rates, piRNAs were the most represented type of sncRNAs in urine, plasma exosomes, and stool (Supplementary Figure 2D). Urine samples emerged as the specimen characterized by the highest piRNA and tRNA contents (Supplementary Figure 1B). Among the other sncRNAs identified there were tRNAs, mitochondrial RNAs, and snoRNAs particularly in plasma exosomes. Consistently,



**Figure 2:** (A) Venn diagram reporting the number of miRNAs detected in different specimens from healthy individuals and their overlap. (B) Heat map showing the log<sub>10</sub> number of normalized reads supporting the miRNAs specifically detected in one specimen or commonly detected among them. (C) PCA plot showing the small RNA-Seq datasets separation obtained using miRNAs detected in samples analysed.

considering those sncRNAs specific of each specimen, the highest number of sncRNAs was identified in urine ( $n = 127$ ) and the same were grouped substantially apart from the other datasets in a PCA analysis using the sncRNA expression levels (Figure 3B).

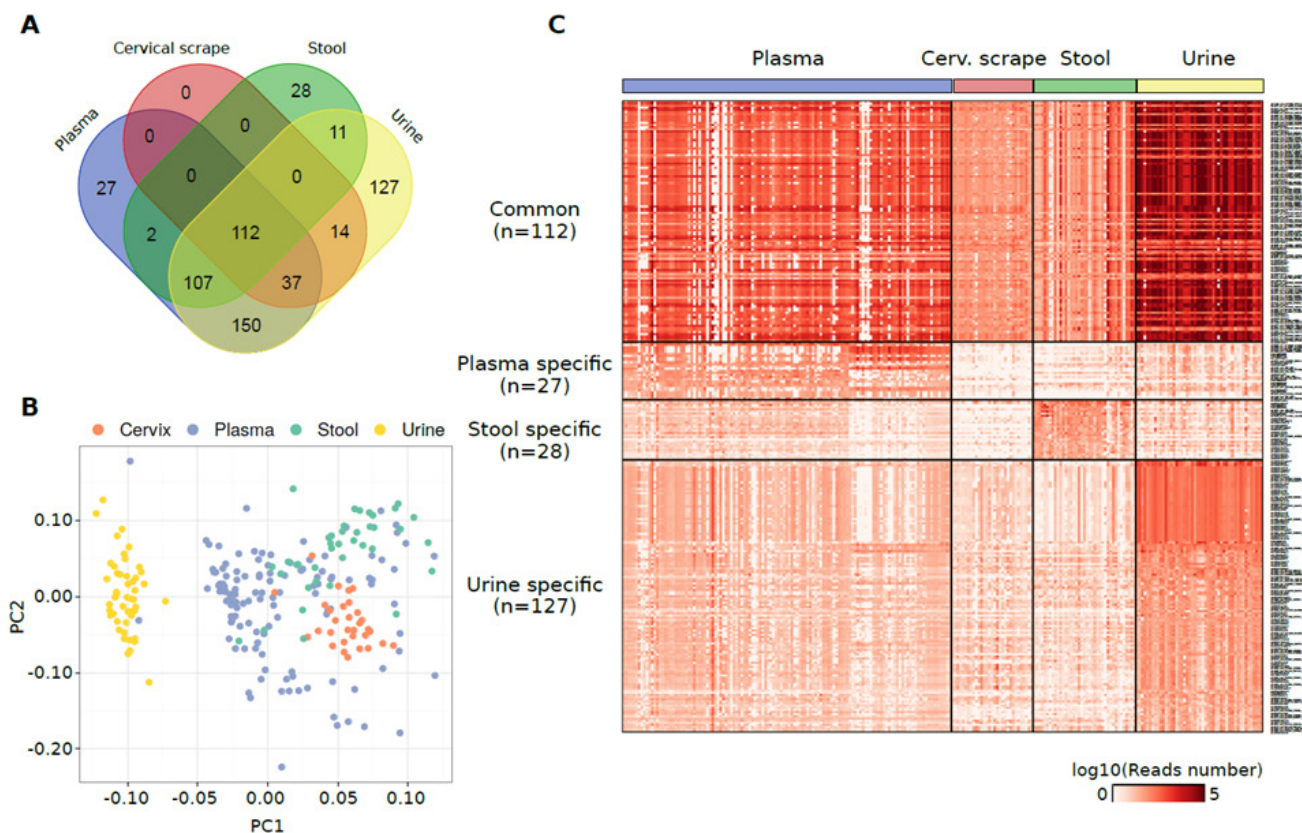
PiR-31068 was the most abundant molecule in urine samples (Supplementary Figure 2D). The tRNA chr1.tRNA2-GlyCCC showed the highest expression levels among the sncRNAs specific in urine samples (median reads = 419 reads) while piR-43137 was the most abundant plasma exosome-specific sncRNA (median reads = 366 reads), and piR-36705 the most abundant stool-specific sncRNA (median reads = 131 reads) (Figure 3C and Supplementary Figure 2F). No specific sncRNAs of cervical scrapes were identified.

The specificity of these sets of sncRNAs was confirmed using a Random Forest classification algorithm which exactly classified 236 samples out of 243 (97.1%) (Supplementary Table 3B). The attribute selection analysis evidenced tRNAs chr19.tRNA2-GlyTCC, chr2.tRNA12-PseudoCTC, and chr6.tRNA150-MetCAT as the sncRNAs with the highest *merit* in the classification (Supplementary Table 3C).

Regression analysis between paired plasma exosome and stool samples or plasma exosome and urine samples from the same individuals showed a low coherent expression for sncRNAs detected (Supplementary Table 3D, 3E).

### Assessing inter-individual variability in sncRNA expression in each specimen type

Independently of the extensive intrinsic variability among subject's extracellular RNA levels for each specimen, we selected the highly abundant sncRNAs with the lowest variable expression levels (i.e. potential reference sncRNAs) across all subjects. To achieve this, the highly-expressed miRNAs and sncRNAs specifically detected in plasma exosomes, stool, or urine (Figure 2A and 3A) characterized by the smallest expression variation in each specimen were identified by computing the median and the Median Absolute Deviation (MAD) of the expression levels (Supplementary Tables 2I and 3F). Specifically, the analysis highlighted miR-142-5p, miR-655-5p, and miR-196a-1-5p as potential *reference* miRNAs in plasma exosomes, stool, and urine,



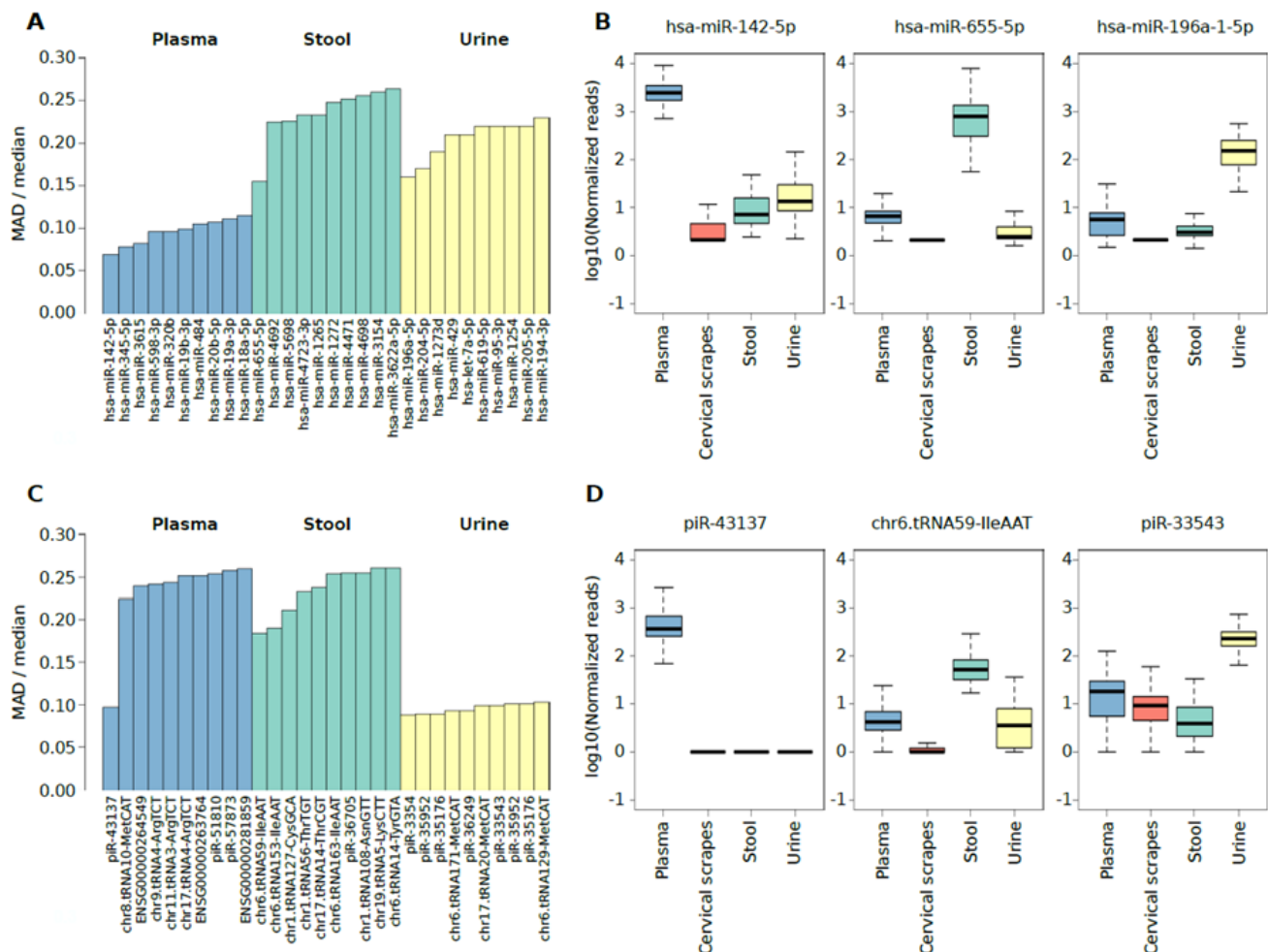
**Figure 3:** (A) Venn diagram reporting the number of non-miRNA sncRNA species detected in different specimens from healthy individuals and their overlap. (B) PCA plot showing the small RNA-Seq datasets separation obtained using the non-miRNA RNA species detected in the samples analysed. (C) Heat map showing the log<sub>10</sub> number of normalized reads supporting the non-miRNA RNA species detected in one specimen only or commonly detected among them.

respectively (Figure 4A, 4B). Considering the isomiRs predicted for the reference miRNAs reported in Figure 4A, all the isomiRs predicted for reference miRNAs in plasma and urine were also identified in these sample types while no isomiRs were predicted for reference miRNAs in stool samples (Supplementary Table 2D). The analysis of *reference* non-miRNA sncRNAs highlighted piR-43137, chr6.tRNA59-IleAAT, and piR-33543 as the candidate sncRNAs for plasma exosome, stool, and urine samples, respectively (Figure 4C, 4D).

To further investigate the *reference* sncRNAs identified, an integrative analysis of public resources was performed (Supplementary Table 4A, 4B). Considering the top 10 reference miRNAs and sncRNAs characterized by the low ratio between MAD and median expression (Figure 4A, 4C), their expression was compared with RNA-Seq data from specimens collected in five independent studies and two databases publicly available.

All the 10 most stably expressed miRNAs in plasma exosomes were also detected (average reads >20) in exosome data (analysed individuals,  $n = 40$ ) from [15], plasma samples ( $n = 55$ ) from [6], and venous blood samples data ( $n = 3$ ) from [22]. Six out of 10 stably expressed urine miRNAs were detected in urine RNA ( $n = 4$  and  $n = 55$  analysed by [22] and [6], respectively). Interestingly, six of the 10 top miRNAs were also detected in samples from kidney ( $n = 11$ ) or bladder ( $n = 2$ ) small RNA-Seq data from DASHR database. The expression of the top 10 miRNAs in stool samples was not confirmed in stool data ( $n = 2$ ) from [22], but two miRNAs were detected in colon samples ( $n = 8$ ) by [23].

Among the reference non-miRNA sncRNAs, piR-62011 was detected as abundant in our plasma exosome data as well as plasma, serum and whole blood data from DASHR. Chr6.tRNA152-ValCAC was detected in our urine set and in small RNA-Seq data from DASHR



**Figure 4:** (A) Bar plot showing the top 10 miRNAs characterized by the lower ratio between the MAD and the median expression levels in plasma exosome, stool, or urine samples. (B) Box plot showing the log10 number of normalized reads supporting miRNAs characterized by the lower ratio between the MAD and the median expression level in plasma exosomes stool, or urine samples. (C) Bar plot showing the top 10 non-miRNA sncRNA species characterized by the lower ratio between the MAD and the median expression levels in plasma exosome, stool, or urine samples. (D) Box plot showing the log10 number of normalized reads supporting non-miRNA sncRNA species characterized by the lower ratio between the MAD and the median expression level in plasma exosomes stool, or urine samples.

kidney tissues [21]. All the others reference non-miRNA sncRNAs were generally associated with low expression in most of the datasets analysed.

## DISCUSSION

The study of the expression patterns of different sncRNAs in a wide spectrum of tissues, along with investigations into the functions of these molecules, is yielding novel insights in the fast-growing field of non-coding RNAs in the normal cell biology and pathogenesis. miRNAs have been extensively studied in the extracellular space but little is still known about the presence of other sncRNAs [15]. As diagnostic and therapeutic procedures move from biopsies in the direction to less invasive methodologies, sncRNAs analysed in different biospecimens represent attractive candidates as biomarkers for complex diseases [12].

In the present study, we investigated expression patterns of sncRNAs in different human biospecimens that could be easily and minimally invasively collected also in the context of screening programs. The data presented hereby were obtained from healthy subjects representing, on average, the steady state in normal conditions of the human organism.

The first analysis was focused on miRNA expression distribution across different investigated specimens. Globally, setting up an arbitrary threshold of median 20 reads, almost 400 miRNAs (out of an average of 1,046 unique miRNAs identified across specimens with at least one read) were detected, with many of them specific to one or few specimen types. A large set of miRNAs was expressed only in plasma exosomes ( $n = 155$ ) while less miRNAs were private of stool or urine and only one of cervical scrapes. Plasma exosomes also shared several miRNAs with other specimens (particularly urine with 109 expressed miRNAs in common). Interestingly, considering the whole set of highly expressed miRNAs, it was possible to accurately group samples of the same biological type independently from the others. This aspect is important in search of specific biomarkers representing an altered status of a tissue in relation to a disease [24]. Conversely, eleven miRNAs presented a similar pattern of expression among all specimens. The most commonly investigated resulted miR-320a whose downregulation is associated with different diseases including cancer [25–30]. The relevance of an ubiquitous high expression of this miRNA related to a healthy status is supported by our findings as well. miR-589-5p, miR-636, and miR-4792 have been also described previously in other studies. miR-589-5p resulted a good inhibitor of MAP3K8 and suppressor of CD90+ cancer stem cells in hepatocellular carcinoma [31]. On the other hand, miR-636 was proposed as a good biomarker for several diseases in a large set of tissues and biofluids such as diabetic kidney disease [32], colorectal cancer [33], and pancreatic cancer [34]. Finally, miR-4792

was found dysregulated in oral submucous fibrosis [35], in nasopharyngeal carcinoma tissues [36] and in uterine leiomyoma [37]. Surprisingly, the rest of the commonly expressed miRNAs were not studied in detail before. Besides being found dysregulated in many studies in relation to different diseases, those miRNAs commonly expressed across different types of samples could be taken into consideration as multi-specimen markers. We have compared our results to those of available datasets on same specimens or anatomically-related tissues [6, 15, 16, 21, 22, 38]. The total number of reads obtained and the proportion of the detected sncRNA species is comparable to other studies previously published with the exception of the study of Yeri and colleagues that included YRNAs [6, 15, 38]. For instance, the high expression of the above mentioned miR-320a and miR-589-5p were also observed in all other datasets.

Notably, in our study, we could compare the co-expression of sncRNAs in plasma exosomes/urine or plasma exosomes/stool collected from the same subjects. Again, in the search of specific markers related to disease, it is important to have an overview on the similarities/differences across different biotypes at an individual level. Apparently, except for very few miRNAs mostly detected in urine/plasma, we could not observe any significant relationship between the expression of same sncRNAs in different biospecimens. This aspect is very important, in the sense that a multi-specimen miRNA panel may be more relevant for accurately describing a disease status, providing different miRNA behaviours across tissues. Similar findings were reported by us in a study on miRNA expression levels in both stool and whole plasma of healthy subjects with different dietary habits. Despite similar associations were observed between miRNA and diet (vegans, vegetarian vs omnivorous) or lifestyle habits, miRNA expression levels were not related between the two different specimens [39].

Since isomiRs have emerged as widely expressed in normal and cancer tissues [40, 41], we further investigated whether they were also detectable in the analysed specimens. As reported in Supplementary Table 2C–2D, many isomiRs were predicted in our datasets particularly in plasma and urine samples. Interestingly, among the 11 miRNAs commonly expressed in all specimens, eight were associated to an isomiR predicted in only one or two types of them.

miRNA profiling by NGS in different specimens in relation to healthy status and pathological conditions is becoming more and more frequent, especially in whole plasma [15]. Less explored is the field of other non-miRNA sncRNAs, although RNA sequencing potentialities, new annotation tools available and an increasing number of studies demonstrating their role in the normal physiology of the organism are appearing [42]. These ‘new’ small RNAs may play an important role in RNA silencing, micro-guarding and cancer [43]. In our study, we have confirmed that small RNA-Seq provide a



huge number of reads not mapping to the miRNome in all type of samples analysed, particularly in stool. However, there is still not a consensus on how to comprehensively analyse these RNA molecules. In the present study, we focused on RNA species with a size between 30 to 70 nucleotides, due to the characteristics of the libraries prep kit employed, specific for small RNA sequencing. Considering these criteria, we have obtained a potential list of thousands of RNAs (>30,000) which we have used to filter the remapped reads after their annotation (from DASHR and GENCODE databases). Despite several different sncRNAs identifiable with our thresholds (misc\_RNA, Mt\_tRNA, piRNA, rRNA, snoRNA, snRNA, sRNA, tRNA), we have mainly identified piRNAs and tRNAs. In urine, we observed the largest number of “private” sncRNAs other than miRNAs ( $n = 127$ ). Cervical scrapes had the less abundant number of these species and none of them was private. In total, 112 sncRNAs resulted expressed in all the biospecimens. Again, plasma exosome and urine samples shared many molecules in common ( $n = 150$ ). Interestingly, as for miRNAs, also for the other sncRNAs, several molecules were characteristics of a single specimen while others were in common. Each body fluid appears to have clear differences in extracellular RNA expression profiles. For example, there appears to be a high proportion of piRNAs in urine samples, when compared with other RNA biotypes. This is quite similar to what observed by Yeri *et al.* [6] which observed an overrepresentation in urine of piRNAs and tRNAs. piRNAs hold great promise as potential biomarkers, owing to their sncRNA features such as small size, stability in biofluids and archival materials, and the variety of detection methods. Moreover, considering there are 10–25 times more piRNA species (20,000–50,000) than miRNAs, the impact of their deregulation is likely at least as relevant. Additionally, piRNA expression patterns have been shown to be deregulated in a variety of cancer types [44–46]. Recently, the study of tRNAs and their role in the regulation of gene expression is revealing new interesting aspects in molecular biology. tRNA-derived small RNAs, named tRNA halves (tiRNAs) and tRNA fragments (tRFs), have been reported to be abundant and their dysregulation to be associated with cancer [43]. Interestingly, we have not identified snoRNAs and other sncRNAs as reported in other studies [6, 47]. Better sncRNA tissue atlases that include more comprehensive profiles of the small RNA species will be necessary for better comparisons.

Expression patterns of miRNAs have been extensively studied but there is still controversy on the best endogenous control(s) to employ as reference in studies by RT-qPCR or microarray, especially when analysing biofluids [24]. An overview of the expression levels of sncRNAs in a large set of biofluids/biospecimen could provide a good base for the research of endogenous controls to be used in case-control studies when searching for sncRNAs as biomarkers of disease [24]. We propose miR-

142-5p, miR-655-5p, and miR-196a-1-5p as miRNAs with a high and stable expression in plasma exosome, stool, and urine respectively, while piR-43137, chr6.tRNA59-IleAAT, and piR-33543 as the candidate references among other sncRNAs in plasma exosomes, stool, and urine respectively. miR-142-5p has been found dysregulated in plasma but not in exosomes [47–49] although it has been demonstrated that in rats the activation of the acute stress response modifies its profile in plasma exosomes [50]. miR-655-5p and miR-196a-1-5p have never been studied in stool and urine, except for miR-196a reported to be altered in focal segmental glomerulosclerosis [51]. Considering the top 10 *reference* miRNAs detected in plasma exosomes, stool, or urine sample group, we observed a general coherence between the specificity of isomiR and reference miRNA expression. The only exceptions were two 5' variants of miR-204. However, these variants were detected by imposing two and three 5' mismatches on a 14- and 15 nt sub-sequence of miR-204, respectively. The read alignment against such small sequences makes the read assignment less reliable reinforcing the hypothesis that a deeper sequencing depth is required to characterize properly the expression of these miRNA variants.

The biological samples used in the present work are very attractive for the research of non-invasive biomarkers. Blood plasma and urine belong to the group of easily accessible body fluids, and they are among the most frequently used diagnostic material for the development of surrogate cancer biomarkers [52, 53]. From the first work reporting the presence in plasma of miRNAs by Lawrie and colleagues [54], a growing number of studies have evaluated their expression in relation to a wide range of diseases and focused on the biology and features of circulating miRNAs [55]. Circulating miRNAs are considered as a tool employed in the horizontal gene transfer between cells within the tumor or between tumor and host cells: this is a strong biological rationale to use them as a new class of cancer biomarkers. miRNAs and other sncRNAs can be released by the cell by passive leakage into circulation. However, these molecular species can be released in a more active way from the cells by secretion of shedding microvesicles or exosomes containing free sncRNAs or in the form of ribonucleoprotein complexes [56]. Bladder cells are in direct contact with urine making this body fluid an ideal source for the detection of cancer biomarkers. Urine is collected noninvasively, and the procedure is relatively fast and cost-efficient compared with other clinical samples. In addition, sampling can be repeated at different times, and this makes urine an attractive candidate as a screening test for urogenital cancers that needs constant monitoring [53]. Stool has been extensively used as a potential substrate for developing non-invasive molecular screening tests for gastrointestinal diseases including colorectal cancer and for microbiome analyses. There is a rationale for determination of noncoding RNAs expression levels in stool which includes the observations that colonocytes are continuously

shed into the faecal stream, with a periodicity of exfoliation roughly every 3–4 days. In addition, sncRNAs are extremely stable, enabling accurate and reproducible detection in the stool without need of special stabilization or logistical requirements. Conventional stool-based screening tests present several limitations including low sensitivity and specificity for advanced adenoma and pre-cancerous lesions. No optimal method has been established yet based on faecal DNA- and mRNA-based testing [57]. The role of diet and other lifestyle factors on miRNA and other sncRNA expression profiles in relation to disease risk is still scarcely explored [58]. Dietary components have been implicated in many pathways involved in diseases, including apoptosis, cell-cycle control, inflammation, and angiogenesis. Those pathways are also regulated by different RNAs [59]. Interestingly, recent discoveries point to a role of faecal miRNAs also introduced by the diet on shaping the human microbiota [60]. Cervical exfoliated cells are widely used in cervical cancer screening, both for HPV testing and Pap test. Recently, their use has been extended to miRNA analyses [61]. These few studies show that the potential application of miRNA detection in cervical exfoliated cells deserves further exploration, also as an additional option for triage of HPV-positive women in population-based screening.

We acknowledge that the present study has some limitations but also several strengths. Among the latter, we can consider the large number of samples sequenced, especially for plasma-exosomes, and the possibility to analyse different biospecimens of the same subjects to understand different/similar patterns according to tissue of origin. To our knowledge, we report the largest description of sncRNA data from plasma-derived exosome, as well as the first investigation of this kind on cervical scrape and stool samples by NGS in healthy subjects. Importantly, the outcomes of our study derive from samples analysed with the same protocols by the same operators and analysed by the same pipeline from raw sequencing data to final results. Other studies usually combine different datasets from different studies.

Among the limitations of our study, we can list that the library preparation is optimized for miRNAs while we have also adapted it for detecting a group of other sncRNAs. Additionally, we could not control analyses considering known potential confounders (age, gender) since not all the samples were provided with this information. Finally, some of the samples were investigated only in subjects of one gender only (i.e., urine in males only).

Small RNA-Seq holds promise for exhaustively analyse miRNAs and other sncRNAs in many different types of specimens, as we demonstrated in our study. These RNA molecules are currently investigated for their potential use as diagnostic/prognostic tools. The high resistance to degradation of sncRNAs makes these molecules particularly attractive for researchers that constantly cope with a wide range of incubations and storage conditions, as well as different origins of samples [62]. However, an

optimization and standardization of both the biological and computational procedures to investigate sncRNA expression levels are necessary. Combining molecular aspects with bioinformatics and an epidemiological approach should provide stronger markers to be investigated specifically in particular biospecimens.

## MATERIALS AND METHODS

### Study participants

All samples included in the study were collected from healthy donors participating to different studies running in our laboratories who donated their blood (for plasma extraction), stool, and /or urine for research purposes [63, 64]. For cervical scrapes, samples were collected in the context of a national screening programme (New Technologies for Cervical Cancer screening (NTCC) study, [65]). All subjects provided written informed consent according to the Helsinki declaration. The design of the study was approved by the local Ethics Committees.

### Stool samples (study 1)

In a hospital-based study for colorectal cancer diagnosis, subjects resulting negative to colonoscopy and to any inflammatory disease were included in the present study. For the same individuals, we have collected also plasma samples ( $n = 39$ ). Naturally evacuated stool samples were collected in special tubes with RNA stabilizing solution, returned at the time of performing colonoscopy and stored at  $-80^{\circ}\text{C}$  until RNA extraction.

### Urine (study 2)

The study population included men recruited between the years 2008–2012 in the Turin Bladder Cancer Study (TBCS) who donated an aliquot of blood and urine. A full description of controls is available in Pardini *et al.* [66]. For almost all subjects, we have collected also plasma samples ( $n = 46$ ).

Urine samples from each participant were collected in the morning, stored at  $4^{\circ}\text{C}$  until the processing consisting of centrifugation at 3,000g for 10 min. The urine supernatant aliquots were then transferred in tubes and stored at  $-80^{\circ}\text{C}$  until use.

### Exosome isolation from plasma

In addition to the subjects described above for whom plasma samples were available (Study 1 and Study 2), we have included also 40 plasma samples collected from healthy blood donors for a Leukaemia study (Study 3).

For all subjects, human plasma samples were obtained from 5–8 ml of blood centrifuged for 10 min at 1000 rpm. Plasma aliquots (about 200–300  $\mu\text{l}$  each) were then stored at  $-80^{\circ}\text{C}$  until use. Exosomes were isolated

from 200 µl of plasma using the ExoQuick exosome precipitation solution (System Biosciences, Mountain View, CA, USA) according to the manufacturer's instructions with minor modifications. Briefly, the plasma was mixed with 50.4 µl of ExoQuick solution and refrigerated at 4°C overnight (at least 12 h). The mixture was then further centrifuged at 1500 g for 30 min. The exosome pellet was dissolved in 200 µl of nuclease free water; RNA was extracted immediately from the solution.

### Cervical scrapes

The study is nested in a large Italian multi-centre randomised controlled trial recruiting women in population-based screening programs that actively invite women aged 25–64 years (NTCC Study, [65]). NTCC recruitment was conducted between 2002 and 2004. In the present study, only samples from HPV negative women were included. Cervical scrape samples have been collected and stored in Specimen Transport Medium (STM), or RNA-later at –80°C until RNA extraction.

### RNA extraction and quality control

Total RNA from plasma exosomes was extracted with the miRNeasy plasma/serum mini kit (Qiagen) using the QiaCube extractor (Qiagen). RNA from stool was extracted using the Stool Total RNA Purification Kit (Norgen Biotek Corp). Total RNA from urine was extracted with Urine microRNA Purification kit (Norgen biotek corp), following the manufacturer's standard protocol.

RNA from cervical scrape was extracted from samples stored in STM or RNA-later, using the miRCURY™ RNA Isolation Kit - Cell & Plant (Exiqon) following manufacturer's protocol.

RNA quality and quantity was verified according to MIQE guidelines (<http://miqe.gene-quantification.info/>). For all samples, RNA concentration was quantified by Qubit® 2.0 Fluorometer with Qubit® microRNA Assay Kit (Invitrogen).

### Library preparation for small RNA-Seq

Small RNA transcripts were converted into barcoded cDNA libraries. Library preparation was performed with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs Inc., USA). For each library, 6 µL of RNA (min 35 ng) were used in all the experimental procedures as starting material. Each library was prepared with a unique indexed primer so that the libraries could all be pooled into one sequencing lane. Multiplex adaptor ligations, reverse transcription primer hybridization, reverse transcription reaction and PCR amplification were performed according to the protocol for library preparation (Protocol E7330, New England BioLabs Inc., USA). After PCR amplification, the cDNA constructs were purified with the QIAquick

PCR Purification Kit (Qiagen, Germany) following the modifications suggested by the NEBNext Multiplex Small RNA Library Prep Protocol and loaded on the Bioanalyzer 2100 (Agilent, Germany) using the DNA High Sensitivity Kit (Agilent, Germany) according to the manufacturer's protocol. Libraries were pooled together (24plex) and further purified with a gel size selection.

A concluding Bioanalyzer 2100 run with the High Sensitivity DNA Kit (Agilent Technologies, Germany) that allows the analysis of DNA libraries regarding size, purity and concentration completed the workflow of library preparation. The obtained sequence libraries were subjected to the Illumina sequencing pipeline, passing through clonal cluster generation on a single-read flow cell (Illumina Inc., USA) by bridge amplification on the cBot (TruSeq SR Cluster Kit v3-cBOT-HS, Illumina Inc., USA) and 50 cycles sequencing-by-synthesis on the HiSeq 2000 (Illumina Inc., USA) (in collaboration with EMBL, Heidelberg, Germany).

### Computational analyses (additional information in Supplementary Material)

#### Analysis of miRNAs

miRNA data analysis was performed following the optimized workflow proposed in [19]. The obtained FASTQ files from small RNA-seq were quality-checked using FastQC software.

Reads shorter than 14 nucleotides were discarded from the analysis; the remaining reads were clipped from the adapter sequences using Cutadapt software (<http://journal.embnet.org/index.php/embnetjournal/article/view/200>). The trimmed reads were mapped against the precursor miRNA sequences downloaded from miRBase (Release 21) by the Shrimp algorithm. A matrix of integer values called counting matrix was created.

Since plasma datasets were generated in independent studies and presented a large variability, a SVA [67] was performed to correct the read counts. IsomiR analysis was performed using isomiRID algorithm [68] in default settings. A maximum of three mismatches between reads and reference miRNA sequences was considered for the analysis.

#### Analysis of other sncRNAs

The set of small RNA-Seq reads not aligned by SHRiMP over miRNA sequences were aligned against human genomic sequence hg38 (GRCh38) using Bowtie2 v2.2.7 in default settings [69]. Reads alignment files were used to quantify the expression of ncRNA annotations from Gencode v24 [70] and DASHR database [21]. The annotations with median reads greater than 20 were selected. Then, read counts were normalized by computing the library size factor [71]. The SVA [67] was performed to correct the read counts of plasma studies.

## Bioinformatic tools and data integration

The list and the expression levels of sncRNAs identified in the different specimen types were compared using Venn diagrams and *heatmap.2* R functions. PCA analysis was performed using *prcomp* R function and *autoplot* function from *ggfortify* R package. The contribution of each sncRNA expression level to the classification of specimen type was evaluated using Weka 3.6.12 [72]. miRNA functional enrichment analysis was performed using EnrichR web tool [73] on the list of validated miRNA targets annotated in miRWalk 2.0 database [74].

The set of sncRNAs identified in this study was compared with public lists sncRNAs detected in specimens and tissues from healthy individuals as reported in supplementary materials of target publications and databases.

## Abbreviations

isomiR: isoform of miRNAs; miRNA: microRNA; piRNAs: Piwi-interacting RNA; tRNAs: transfer RNA; RT-qPCR: Quantitative reverse transcription PCR; sncRNA: small non-coding RNA; snoRNAs: small nucleolar RNAs; NGS: Next-Generation Sequencing; RNA-Seq: RNA-Sequencing; HPV: Human Papilloma Virus; DASHR: Database of Small Human non-coding RNAs; PCA: principal component analysis; mRNA: messenger RNA; MAD: Median Absolute Deviation; tiRNAs: tRNA halves; tRFs: tRNA fragments; TBCS: Turin Bladder Cancer Study; NTCC: New Technologies for CC screening; STM: Specimen Transport Medium; GRCh38: Genome Reference Consortium Human Build 38; SVA: Surrogate Variable Analysis.

## Author contributions

Conception and design of the study: BP, AN, GF, FC; Acquisition, analysis and interpretation of data: BP, AN, GF, FC, MA, FR, RAC, ST, GG; Draft of the manuscript: BP, GF, FC, ST, AN; Critical revision of the manuscript: MA, FR, GG, GR, MAI, NK, GM, PV, RAC.

## ACKNOWLEDGMENTS

The authors are very thankful to all volunteers that participated to the present study. We wish to thank EMBL Genomics Core Facilities (EMBL, <http://www.genecore.embl.de>) for sequencing of our libraries.

## FUNDING

This work was supported by the Italian Institute for Genomic Medicine (IIGM) and Compagnia di San Paolo Torino, Italy (to B.P., A.N., P.V. and G.M.), by Fondazione

Umberto Veronesi “Post-doctoral fellowship Year 2014, 2015, 2016 and 2017” (B.P. and S.T. recipients), by the Italian Association on Cancer Research (IG2013 N.14119 to G.R. and A.N.), by Lega Italiana per La Lotta contro i Tumori (to F.C. and A.N.) and by the Grant Agency of the Czech Republic (17-16857S to A.N.).

## CONFLICTS OF INTEREST

The authors declare no competing interests.

## REFERENCES

1. Cech TR, Steitz JA. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*. 2014; 157:77–94. <https://doi.org/10.1016/j.cell.2014.03.008>.
2. Bracken CP, Scott HS, Goodall GJ. A network-biology perspective of microRNA function and dysfunction in cancer. *Nat Rev Genet*. 2016; 17:719–32. <https://doi.org/10.1038/nrg.2016.134>.
3. Cheng Y, Tan N, Yang J, Liu X, Cao X, He P, Dong X, Qin S, Zhang C. A translational study of circulating cell-free microRNA-1 in acute myocardial infarction. *Clin Sci (Lond)*. 2010; 119:87–95. <https://doi.org/10.1042/CS20090645>.
4. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio MV, Visone R, Sever NI, Fabbri M, Iuliano R, Palumbo T, Pichiorri F, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*. 2005; 353:1793–801. <https://doi.org/10.1056/NEJMoa050995>.
5. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011; 12:861–74. <https://doi.org/10.1038/nrg3074>.
6. Yeri A, Courtright A, Reiman R, Carlson E, Beecroft T, Janss A, Siniard A, Richholt R, Balak C, Rozowsky J, Kitchen R, Hutchins E, Winarta J, et al. Total Extracellular Small RNA Profiles from Plasma, Saliva, and Urine of Healthy Subjects. *Sci Rep*. 2017; 7:44061. <https://doi.org/10.1038/srep44061>.
7. Buschmann D, Haberberger A, Kirchner B, Spornraft M, Riedmaier I, Schelling G, Pfaffl MW. Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow. *Nucleic Acids Res*. 2016; 44:5995–6018. <https://doi.org/10.1093/nar/gkw545>.
8. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, Galas DJ, Wang K. The microRNA spectrum in 12 body fluids. *Clin Chem*. 2010; 56:1733–41. <https://doi.org/10.1373/clinchem.2010.147405>.
9. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet*. 2012; 13:358–69. <https://doi.org/10.1038/nrg3198>.
10. Accerbi M, Schmidt SA, De Paoli E, Park S, Jeong DH, Green PJ. Methods for isolation of total RNA to recover miRNAs and other small RNAs from diverse

- species. *Methods Mol Biol.* 2010; 592:31–50. [https://doi.org/10.1007/978-1-60327-005-2\\_3](https://doi.org/10.1007/978-1-60327-005-2_3).
11. Lopez JP, Diallo A, Cruceanu C, Fiori LM, Laboissiere S, Guillet I, Fontaine J, Ragoussis J, Benes V, Turecki G, Ernst C. Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med Genomics.* 2015; 8:35. <https://doi.org/10.1186/s12920-015-0109-x>.
  12. Witwer KW, Halushka MK. Toward the promise of microRNAs - Enhancing reproducibility and rigor in microRNA research. *RNA Biol.* 2016; 13:1103–16. <https://doi.org/10.1080/15476286.2016.1236172>.
  13. Margue C, Reinsbach S, Philippidou D, Beaume N, Walters C, Schneider JG, Nashan D, Behrmann I, Kreis S. Comparison of a healthy miRNome with melanoma patient miRNomes: are microRNAs suitable serum biomarkers for cancer? *Oncotarget.* 2015; 6:12110–27. <https://doi.org/10.18632/oncotarget.3661>.
  14. Yuan T, Huang X, Woodcock M, Du M, Dittmar R, Wang Y, Tsai S, Kohli M, Boardman L, Patel T, Wang L. Plasma extracellular RNA profiles in healthy and cancer patients. *Sci Rep.* 2016; 6:19413. <https://doi.org/10.1038/srep19413>.
  15. Freedman JE, Gerstein M, Mick E, Rozowsky J, Levy D, Kitchen R, Das S, Shah R, Danielson K, Beaulieu L, Navarro FC, Wang Y, Galeev TR, et al. Diverse human extracellular RNAs are widely detected in human plasma. *Nat Commun.* 2016; 7:11106. <https://doi.org/10.1038/ncomms11106>.
  16. Ben-Dov IZ, Whalen VM, Goilav B, Max KE, Tuschl T. Cell and Microvesicle Urine microRNA Deep Sequencing Profiles from Healthy Individuals: Observations with Potential Impact on Biomarker Studies. *PLoS One.* 2016; 11:e0147249. <https://doi.org/10.1371/journal.pone.0147249>.
  17. Fehlmann T, Ludwig N, Backes C, Meese E, Keller A. Distribution of microRNA biomarker candidates in solid tissues and body fluids. *RNA Biol.* 2016; 13:1084–88. <https://doi.org/10.1080/15476286.2016.1234658>.
  18. Bahn JH, Zhang Q, Li F, Chan TM, Lin X, Kim Y, Wong DT, Xiao X. The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin Chem.* 2015; 61:221–30. <https://doi.org/10.1373/clinchem.2014.230433>.
  19. Cordero F, Beccuti M, Arigoni M, Donatelli S, Calogero RA. Optimizing a massive parallel sequencing workflow for quantitative miRNA expression analysis. *PLoS One.* 2012; 7:e31630. <https://doi.org/10.1371/journal.pone.0031630>.
  20. Wright JC, Mudge J, Weisser H, Barzine MP, Gonzalez JM, Brazma A, Choudhary JS, Harrow J. Improving GENCODE reference gene annotation using a high-stringency proteogenomics workflow. *Nat Commun.* 2016; 7:11778. <https://doi.org/10.1038/ncomms11778>.
  21. Leung YY, Kuksa PP, Amlie-Wolf A, Valladares O, Ungar LH, Kannan S, Gregory BD, Wang LS. DASHR: database of small human noncoding RNAs. *Nucleic Acids Res.* 2016; 44:D216–22. <https://doi.org/10.1093/nar/gkv1188>.
  22. Seashols-Williams S, Lewis C, Calloway C, Peace N, Harrison A, Hayes-Nash C, Fleming S, Wu Q, Zehner ZE. High-throughput miRNA sequencing and identification of biomarkers for forensically relevant biological fluids. *Electrophoresis.* 2016; 37:2780–88. <https://doi.org/10.1002/elps.201600258>.
  23. Hamfjord J, Stangeland AM, Hughes T, Skrede ML, Tveit KM, Ikdahl T, Kure EH. Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. *PLoS One.* 2012; 7:e34150. <https://doi.org/10.1371/journal.pone.0034150>.
  24. Hardwick SA, Deveson IW, Mercer TR. Reference standards for next-generation sequencing. *Nat Rev Genet.* 2017; 18:473–84. <https://doi.org/10.1038/nrg.2017.44>.
  25. Xie F, Yuan Y, Xie L, Ran P, Xiang X, Huang Q, Qi G, Guo X, Xiao C, Zheng S. miRNA-320a inhibits tumor proliferation and invasion by targeting c-Myc in human hepatocellular carcinoma. *Onco Targets Ther.* 2017; 10:885–94. <https://doi.org/10.2147/OTT.S122992>.
  26. Lv Q, Hu JX, Li YJ, Xie N, Song DD, Zhao W, Yan YF, Li BS, Wang PY, Xie SY. MiR-320a effectively suppresses lung adenocarcinoma cell proliferation and metastasis by regulating STAT3 signals. *Cancer Biol Ther.* 2017; 18:142–51. <https://doi.org/10.1080/15384047.2017.1281497>.
  27. Yu J, Wang L, Yang H, Ding D, Zhang L, Wang J, Chen Q, Zou Q, Jin Y, Liu X. Rab14 Suppression Mediated by MiR-320a Inhibits Cell Proliferation, Migration and Invasion in Breast Cancer. *J Cancer.* 2016; 7:2317–26. <https://doi.org/10.7150/jca.15737>.
  28. Cordes F, Brückner M, Lenz P, Veltman K, Glauben R, Siegmund B, Hengst K, Schmidt MA, Cichon C, Bettenworth D. MicroRNA-320a Strengthens Intestinal Barrier Function and Follows the Course of Experimental Colitis. *Inflamm Bowel Dis.* 2016; 22:2341–55. <https://doi.org/10.1097/MIB.0000000000000917>.
  29. Tadano T, Kakuta Y, Hamada S, Shimodaira Y, Kuroha M, Kawakami Y, Kimura T, Shiga H, Endo K, Masamune A, Takahashi S, Kinouchi Y, Shimosegawa T. MicroRNA-320 family is downregulated in colorectal adenoma and affects tumor proliferation by targeting CDK6. *World J Gastrointest Oncol.* 2016; 8:532–42. <https://doi.org/10.4251/wjgo.v8.i7.532>.
  30. Xishan Z, Ziying L, Jing D, Gang L. MicroRNA-320a acts as a tumor suppressor by targeting BCR/ABL oncogene in chronic myeloid leukemia. *Sci Rep.* 2015; 5:12460. <https://doi.org/10.1038/srep12460>.
  31. Zhang X, Jiang P, Shuai L, Chen K, Li Z, Zhang Y, Jiang Y, Li X. miR-589-5p inhibits MAP3K8 and suppresses CD90+ cancer stem cells in hepatocellular carcinoma. *J Exp Clin Cancer Res.* 2016; 35:176.
  32. Eissa S, Matboli M, Aboushahba R, Bekhet MM, Soliman Y. Urinary exosomal microRNA panel unravels novel biomarkers for diagnosis of type 2 diabetic kidney disease.

- J Diabetes Complications. 2016; 30:1585–92. <https://doi.org/10.1016/j.jdiacomp.2016.07.012>.
33. Slattery ML, Herrick JS, Mullany LE, Wolff E, Hoffman MD, Pellatt DF, Stevens JR, Wolff RK. Colorectal tumor molecular phenotype and miRNA: expression profiles and prognosis. *Mod Pathol*. 2016; 29:915–27.
  34. Schultz NA, Dehlendorff C, Jensen BV, Bjerregaard JK, Nielsen KR, Bojesen SE, Calatayud D, Nielsen SE, Yilmaz M, Holländer NH, Andersen KK, Johansen JS. MicroRNA biomarkers in whole blood for detection of pancreatic cancer. *JAMA*. 2014; 311:392–404. <https://doi.org/10.1001/jama.2013.284664>.
  35. Chickooree D, Zhu K, Ram V, Wu HJ, He ZJ, Zhang S. A preliminary microarray assay of the miRNA expression signatures in buccal mucosa of oral submucous fibrosis patients. *J Oral Pathol Med*. 2016; 45:691–697. <https://doi.org/10.1111/jop.12431>.
  36. Li Y, Chen X. miR-4792 inhibits epithelial-mesenchymal transition and invasion in nasopharyngeal carcinoma by targeting FOXC1. *Biochem Biophys Res Commun*. 2015; 468:863–69. <https://doi.org/10.1016/j.bbrc.2015.11.045>.
  37. Georgieva B, Milev I, Minkov I, Dimitrova I, Bradford AP, Baev V. Characterization of the uterine leiomyoma microRNAome by deep sequencing. *Genomics*. 2012; 99:275–81. <https://doi.org/10.1016/j.ygeno.2012.03.003>.
  38. Ainsztein AM, Brooks PJ, Dugan VG, Ganguly A, Guo M, Howcroft TK, Kelley CA, Kuo LS, Labosky PA, Lenzi R, McKie GA, Mohla S, Procaccini D, et al, and The NIH Extracellular RNA Communication Consortium. The NIH Extracellular RNA Communication Consortium. *J Extracell Vesicles*. 2015; 4:27493. <https://doi.org/10.3402/jev.v4.27493>.
  39. Tarallo S, Pardini B, Mancuso G, Rosa F, Di Gaetano C, Rosina F, Vineis P, Naccarati A. MicroRNA expression in relation to different dietary habits: a comparison in stool and plasma samples. *Mutagenesis*. 2014; 29:385–91. <https://doi.org/10.1093/mutage/geu028>.
  40. Telonis AG, Magee R, Loher P, Chervoneva I, Londin E, Rigoutsos I. Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res*. 2017; 45:2973–85. <https://doi.org/10.1093/nar/gkx082>.
  41. McCall MN, Kim MS, Adil M, Patil AH, Lu Y, Mitchell CJ, Leal-Rojas P, Xu J, Kumar M, Dawson VL, Dawson TM, Baras AS, Rosenberg AZ, et al. Toward the human cellular microRNAome. *Genome Res*. 2017; 27:1769–81. <https://doi.org/10.1101/gr.222067.117>.
  42. Ng KW, Anderson C, Marshall EA, Minatel BC, Enfield KS, Saprunoff HL, Lam WL, Martinez VD. Piwi-interacting RNAs in cancer: emerging functions and clinical utility. *Mol Cancer*. 2016; 15:5. <https://doi.org/10.1186/s12943-016-0491-9>.
  43. Green D, Fraser WD, Dalmay T. Transfer RNA-derived small RNAs in the cancer transcriptome. *Pflugers Arch*. 2016; 468:1041–47. <https://doi.org/10.1007/s00424-016-1822-9>.
  44. Cui L, Lou Y, Zhang X, Zhou H, Deng H, Song H, Yu X, Xiao B, Wang W, Guo J. Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers. *Clin Biochem*. 2011; 44:1050–57. <https://doi.org/10.1016/j.clinbiochem.2011.06.004>.
  45. Zhang H, Ren Y, Xu H, Pang D, Duan C, Liu C. The expression of stem cell protein Piwil2 and piR-932 in breast cancer. *Surg Oncol*. 2013; 22:217–23. <https://doi.org/10.1016/j.suronc.2013.07.001>.
  46. Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, Zhong D, Ma L, Tong N, Qin C, Yin C, Zhang Z, Wang M. Identification of novel piRNAs in bladder cancer. *Cancer Lett*. 2015; 356:561–67. <https://doi.org/10.1016/j.canlet.2014.10.004>.
  47. Brenu EW, Ashton KJ, Batovska J, Staines DR, Marshall-Gradisnik SM. High-throughput sequencing of plasma microRNA in chronic fatigue syndrome/myalgic encephalomyelitis. *PLoS One*. 2014; 9:e102783. <https://doi.org/10.1371/journal.pone.0102783>.
  48. Sørensen SS, Nygaard AB, Christensen T. miRNA expression profiles in cerebrospinal fluid and blood of patients with Alzheimer's disease and other types of dementia - an exploratory study. *Transl Neurodegener*. 2016; 5:6. <https://doi.org/10.1186/s40035-016-0053-5>.
  49. Huang S, Deng Q, Feng J, Zhang X, Dai X, Li L, Yang B, Wu T, Cheng J. Polycyclic Aromatic Hydrocarbons-Associated MicroRNAs and Heart Rate Variability in Coke Oven Workers. *J Occup Environ Med*. 2016; 58:e24–31. <https://doi.org/10.1097/JOM.0000000000000564>.
  50. Beninson LA, Brown PN, Loughridge AB, Saludes JP, Maslanik T, Hills AK, Woodworth T, Craig W, Yin H, Fleshner M. Acute stressor exposure modifies plasma exosome-associated heat shock protein 72 (Hsp72) and microRNA (miR-142-5p and miR-203). *PLoS One*. 2014; 9:e108748. <https://doi.org/10.1371/journal.pone.0108748>.
  51. Zhang W, Zhang C, Chen H, Li L, Tu Y, Liu C, Shi S, Zen K, Liu Z. Evaluation of microRNAs miR-196a, miR-30a-5P, and miR-490 as biomarkers of disease activity among patients with FSGS. *Clin J Am Soc Nephrol*. 2014; 9:1545–52. <https://doi.org/10.2215/CJN.11561113>.
  52. Gaedcke J, Grade M, Camps J, Sokilde R, Kaczkowski B, Schetter AJ, Difilippantonio MJ, Harris CC, Ghadimi BM, Moller S, Beissbarth T, Ried T, Litman T. The rectal cancer microRNAome—microRNA expression in rectal cancer and matched normal mucosa. *Clin Cancer Res*. 2012; 18:4919–30.
  53. Matullo G, Naccarati A, Pardini B. MicroRNA expression profiling in bladder cancer: the challenge of next-generation sequencing in tissues and biofluids. *Int J Cancer*. 2016; 138:2334–45. <https://doi.org/10.1002/ijc.29895>.
  54. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, Banham AH, Pezzella F, Boulwood J, Wainscoat JS, Hatton CS, Harris AL. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma.

- Br J Haematol. 2008; 141:672–75. <https://doi.org/10.1111/j.1365-2141.2008.07077.x>.
55. Vodicka P, Pardini B, Vymetalkova V, Naccarati A. Polymorphisms in Non-coding RNA Genes and Their Targets Sites as Risk Factors of Sporadic Colorectal Cancer. *Adv Exp Med Biol*. 2016; 937:123–49. [https://doi.org/10.1007/978-3-319-42059-2\\_7](https://doi.org/10.1007/978-3-319-42059-2_7).
  56. Redova M, Sana J, Slaby O. Circulating miRNAs as new blood-based biomarkers for solid cancers. *Future Oncol*. 2013; 9:387–402. <https://doi.org/10.2217/fon.12.192>.
  57. Di Lena M, Travaglio E, Altomare DF. New strategies for colorectal cancer screening. *World J Gastroenterol*. 2013; 19:1855–60. <https://doi.org/10.3748/wjg.v19.i12.1855>.
  58. Vilaprinyo E, Forné C, Carles M, Sala M, Pla R, Castells X, Domingo L, Rue M, and Interval Cancer (INCA) Study Group. Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PLoS One*. 2014; 9:e86858. <https://doi.org/10.1371/journal.pone.0086858>.
  59. Ross SA, Davis CD. MicroRNA, nutrition, and cancer prevention. *Adv Nutr*. 2011; 2:472–85. <https://doi.org/10.3945/an.111.001206>.
  60. Liu S, da Cunha AP, Rezende RM, Cialic R, Wei Z, Bry L, Comstock LE, Gandhi R, Weiner HL. The Host Shapes the Gut Microbiota via Fecal MicroRNA. *Cell Host Microbe*. 2016; 19:32–43. <https://doi.org/10.1016/j.chom.2015.12.005>.
  61. He Y, Lin J, Ding Y, Liu G, Luo Y, Huang M, Xu C, Kim TK, Etheridge A, Lin M, Kong D, Wang K. A systematic study on dysregulated microRNAs in cervical cancer development. *Int J Cancer*. 2016; 138:1312–27. <https://doi.org/10.1002/ijc.29618>.
  62. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA*. 2008; 105:10513–18. <https://doi.org/10.1073/pnas.0804549105>.
  63. Critelli R, Fasanelli F, Oderda M, Polidoro S, Assumma MB, Viberti C, Preto M, Gontero P, Cucchiarale G, Lurkin I, Zwarthoff EC, Vineis P, Sacerdote C, et al. Detection of multiple mutations in urinary exfoliated cells from male bladder cancer patients at diagnosis and during follow-up. *Oncotarget*. 2016; 7:67435–48. <https://doi.org/10.18632/oncotarget.11883>.
  64. Russo A, Modica F, Guarrera S, Fiorito G, Pardini B, Viberti C, Allione A, Critelli R, Bosio A, Casetta G, Cucchiarale G, Destefanis P, Gontero P, et al. Shorter leukocyte telomere length is independently associated with poor survival in patients with bladder cancer. *Cancer Epidemiol Biomarkers Prev*. 2014; 23:2439–46. <https://doi.org/10.1158/1055-9965.EPI-14-0228>.
  65. Bergeron C, Giorgi-Rossi P, Cas F, Schiboni ML, Ghiringhello B, Dalla Palma P, Minucci D, Rosso S, Zorzi M, Naldoni C, Segnan N, Confortini M, Ronco G. Informed cytology for triaging HPV-positive women: substudy nested in the NTCC randomized controlled trial. *J Natl Cancer Inst*. 2015; 107:107. <https://doi.org/10.1093/jnci/dju423>.
  66. Pardini B, Viberti C, Naccarati A, Allione A, Oderda M, Critelli R, Preto M, Zijno A, Cucchiarale G, Gontero P, Vineis P, Sacerdote C, Matullo G. Increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of bladder cancer. *Br J Cancer*. 2017; 116:202–10. <https://doi.org/10.1038/bjc.2016.411>.
  67. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28:882–83. <https://doi.org/10.1093/bioinformatics/bts034>.
  68. de Oliveira LF, Christoff AP, Margis R. isomiRID: a framework to identify microRNA isoforms. *Bioinformatics*. 2013; 29:2521–23. <https://doi.org/10.1093/bioinformatics/btt424>.
  69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–59. <https://doi.org/10.1038/nmeth.1923>.
  70. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012; 22:1760–74. <https://doi.org/10.1101/gr.135350.111>.
  71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
  72. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. *Bioinformatics*. 2004; 20:2479–81. <https://doi.org/10.1093/bioinformatics/bth261>.
  73. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016; 44:W90–7. <https://doi.org/10.1093/nar/gkw377>.
  74. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat Methods*. 2015; 12:697. <https://doi.org/10.1038/nmeth.3485>.