Arraystar Small RNA Microarrays

miRNA, pre-miRNA, tRNA, tsRNA, snoRNA

The new way of sensitive, accurate, simultaneous profiling of major small RNA classes

Highlights

- Simultaneously profile the major small RNA classes: miRNA, pre-miRNA, tRNA, tsRNA, and snoRNA.
- Raise the bar of small RNA profiling to high sensitivity, specificity and accuracy by direct end-labeling and smart probe design.
- Direct and simplified procedures to overcome biases from RNA modifications, RNA fold hindrance, reverse transcription blocks, PCR amplifications, and analysis inaccuracy in small RNA-seq.
- Required RNA sample amounts starting as low as 100 ng, opening up many research opportunities.
- Tolerant for RNA samples at lower qualities: e.g. degraded RNAs, serum/plasma/biofluid RNAs, FFPE RNAs.

Arraystar Small RNA Array combines direct end-labeling and smart probe design microarray technologies to simultaneously detect and quantify small RNAs including miRNAs, pre-miRNAs, tsRNAs, tRNAs and snoRNAs on the same array, providing vital expressional information to study the regulatory functions and biomarker potentials of the small RNAs.

- Biomarker potentials

Small RNAs such as microRNAs have been popularly explored as biomarkers. tRNA and tsRNA populations are now emerging as new classes of biomarkers with greater potentials, owing to their many desired characteristics. The high stability and abundance of tRNA and tsRNA in body fluids (Fig. 1)[1-6] , the involvement in pathological processes, the demonstrated differential expression in solid tumors and hematological malignancies, and their power to discriminate cancer patients from healthy controls open the prospect for development of tRNAs and tsRNA-based biomarker tests. For example, the tRF profiles have been shown to discriminate triple-negative, triple positive breast cancer cells from the normal controls in unsupervised clustering [7] (Fig. 2). The ratio of tsRNAs has also been demonstrated as a good indicator of cancer progression-free survival and a candidate prognostic marker [4].

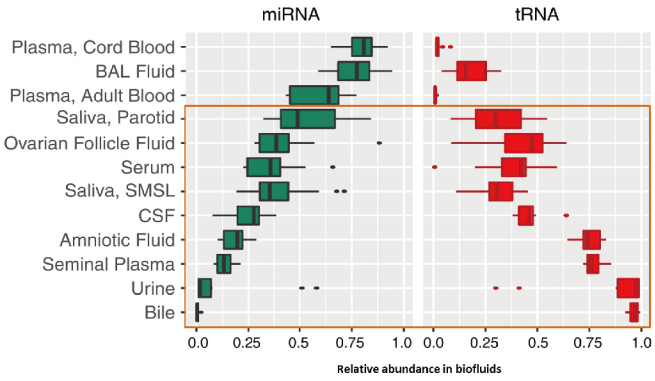

Figure 1. Relative proportions of miRNA vs tRNA in biofluids, where many biofluids have tRNA contents much higher than miRNA [1,2].

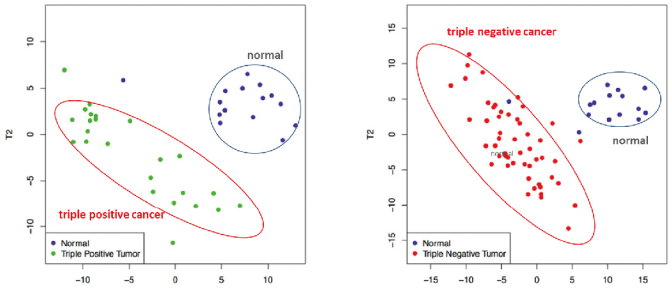

Figure 2. tRF profiles discriminate triple- positive (Left), triple-negative breast cancer cells (Right) from the normal controls in unsupervised clustering[7].

Arraystar Small RNA Array, with the low requirements of RNA amount and quality, opens up opportunities for tRNA/tsRNA biomarker research projects where the samples are rare or of limited supply.

The Challenges of Small RNA Profiling by Sequencing

Small RNA-seq has been used for profiling microRNAs and other small RNAs. However, the use of small RNA-seq data to quantify the relative abundances of small RNAs has been inconsistent with qPCR and Northern Blot results. For example, the level of miR-143 was measured 40 times higher than miR-145 by small RNA-seq, yet their expression levels were equivalent as measured by qPCR or Northern Blot [8-10]. The inaccuracy by small RNA-seq is primarily due to the biases in small RNA-seq library preparation and data analysis [11, 12], which can lead to compromised or misleading results, in some cases by as much as 10,000-fold off the true abundance [13-18].

- *Interference from RNA modifications*

Small RNA modifications (e.g. m1A, m3C and m1G) block reverse transcription and abort cDNA copying, causing sequence biases toward the priming end and failures in building unskewed full length representation in the library. For example, due to the presence of m1A modification in the TUC loop region, small RNA-seq often identifies 18-nt 3' tsRNAs but misses the more predominant 22-nt isoforms as detected by Northern Blot.

- *Library amplification biases*

To produce sufficient cDNA amount for sequencing instrument loading, small RNA-seq requires PCR amplification of the library [19]. However, RNA G/C content, sequence contexts, secondary structures, RNA lengths and priming, and reaction conditions can lead to biases and distortions in the PCR products. That is, when different templates in the transcriptome are amplified together in the PCR reaction, preferential or refractory amplifications for the templates invariably lead to the loss of faithful representation of RNA abundances [20]. Therefore, quantification by small RNA-seq, which requires multiple PCR amplification rounds, is never absolute and necessitates the use of an orthogonal method for validation.

- *Biases caused by TPM in small RNA-seq data analysis*

Reads Per Million reads (TPM) is commonly used in small RNA-seq to represent the relative abundance of a small RNA transcript simply normalized to 1 million small RNAs being sequenced.

However, as illustrated in (Fig. 3)[11], a change in one RNA abundance level can adjust the TPM values for all other RNAs, even though the actual absolute abundance levels of the other RNAs are not changed. Therefore, TPM is dependent on the composition of the RNA population in a sample. A few very highly expressed genes can skew the distribution of TPM expression values. In fact, small RNA repertoires do change substantially under many experimental conditions or across datasets in different studies, compromising TPM to compare small RNA levels between/among samples [21,22].

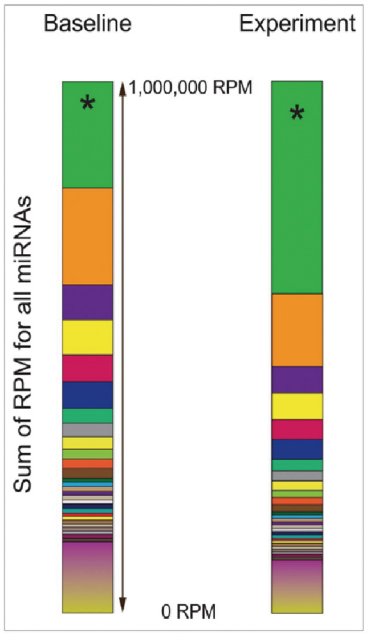

Figure 3. In this simulated situation of miRNA-seq results, except for the *RNA, all other miRNAs between Baseline and Experiment conditions are unchanged. However, the increase of only one *RNA abundance by 2-fold under Experiment condition will simultaneously depress the TPMs for all other miRNAs, even though their actual absolute expression levels are not changed [11].

- *High demand of sample amount*

For tRNA and tsRNA sequencing, 5 ug or even 100 ug total RNA is required for target RNA isolation and pretreatments prior to library construction [24], which precludes research projects with limited sample amount availability.

The Solutions of Small RNA Profiling

Arraystar Small RNA Array, combining direct end-labeling and smart probe design microarray technologies, is designed as a practical and effective solution to overcome these challenges to profile the full spectrum of small RNAs at high sensitivity and accuracy yet at much less input RNA amounts.

- *Raising the bar of small RNA profiling for high sensitivity, specificity and accuracy*

By end-direct labeling, the small RNAs are ligated with pCp-Cy3 onto the 3' -ends by T4 RNA ligase, and one RNA molecule is labeled with one Cy3 label. This method eliminates biases from cDNA synthesis by reverse transcription due to RNA modification interference and RNA folding hindrance as in small RNA-seq; avoids distortions from PCR amplification cycles as in required small RNA-seq library amplification; uses DMSO to reduce the RNA structure and sequence context differences among small RNAs. All these help to preserve the fidelity of native small RNA levels and achieve the unbiased high quantification accuracy better than RNA-seq or even qPCR.

The smart probe design incorporates 5' -hairpin structure and normalized sequence targeting region to specifically distinguish small RNAs with only 1~2 nucleotide differences. Moreover, the high affinity probe hybridization ensures very high sensitivity even for small RNAs at low abundance.

- *Low RNA sample amount requirements*

Arraystar Small RNA Microarray requires as little as 100 ng total RNA, which is magnitudes lower than what small RNA-seq requires. As its direct end labeling chemistry does not require RNA pretreatments that often cause RNA loss, the microarray significantly reduces the demand for input RNA amounts especially for heavily modified RNA biotypes (e.g. tRNA and tsRNA). The low sample amount requirement opens up opportunities for research projects where the samples are rare or of limited supply.

- *Tolerant for RNA samples at lower qualities*

The direct end-labeling is relatively insensitive to nucleotide damage in the substrate RNA sequence as it does not rely on cDNA copying by reverse transcription. Furthermore, whereas the microarray probes are unaffected by unrelated sequence presence, RNA fragments from the abundant rRNAs in degraded RNA samples can contaminate small RNAs in the size range, depressing small RNA coverage in small RNA-seq.

For these reasons, Small RNA Array is particularly advantageous for preserved or chemically treated samples or degraded samples. e.g. serum/plasma/biofluid/FFPE RNAs.

- *Simultaneous profiling of multiple small RNA classes*

Profiling different small RNA classes by sequencing requires separate sequencing methods and experiments: miRNA-seq, tRNA-seq, tsRNA-seq, and regular RNA-seq for longer snoRNA and small RNA precursors. Arraystar Small RNA microarrays use unified labeling chemistry to hybridize to the probes in one array for all major small RNA classes.

	Total probes	Mouse
Total probes	14,707	14,895
miRNAs	2,627 (1,318 5-p-miRNAs; 1,309 3-p-miRNAs)	1,949(966 5-p-miRNAs; 983 3-p-miRNAs)
tsRNAs	4,254	1,809
pre-miRNAs	1,745	1,122
mature tRNAs	346	270
snoRNAs	955	1,323
Small RNA sources	**miRNA:** miRBase **tsRNA:** tRFdb, MINTbase, GtRNADb **pre-miRNA:** miRBase **tRNA:** GtRNADb, ENSEMBL **snoRNA:** ENSEMBL Scientific publications	
Array Format	8 x 15K	

Rich Small RNA Analyses and Annotations

For each small RNA class, the data analyses include profiling measurement values, statistical computations, informative annotations, and publication quality graphics.

- *Differential expression analysis (tsRNA and tRNA as examples)*

tsRNA					
tsRNA_type	tsRNA-sequence	tsRNA-length	tsRNA-precursor	Level	Molecular mechanism
3'tiRNA	ATTCAAAGGTTCCGGGTTCG AGTCCCGGCGGAGTCGCCA	39	tRNA-Arg-TCT-1	Potential	Cytotoxicity to neurons
3'tiRNA	ATGCCGAGGTTGTGAGTTCA AGCCTCACCTGGAGCACCA	39	tRNA-Ile-TAT-3	Potential	Cytotoxicity to neurons

tsRNA_type: tsRNA type (tRF-5, tRF-3, tRF-1, 5-Leader, 5-tiRNA, 3-tiRNA, and i-tRF).
tsRNA-sequence: tsRNA sequence.
tsRNA-length: tsRNA length.
tsRNA-precursor: Symbol for the tsRNA precursor.
Level: Confidence level for the tsRNA
· Functional - Documented with characterized biological functions or disease association;
· Reliable - Recorded in tRFdb or reported by literatures, but without further studies;
· Potential - Predicted by Arraystar based on RNA fragment lengths and cleavage positions in the tRNA.
Mechanism: The molecular mechanism of tsRNA.

tRNA					
tRNA-Sequence	Gene name	GenomeLocus	tRNA promoter locus	pre-tRNA locus	tRNA neighbouring gene
GGGGGTATAGC TCAG... ...	tRNA-Ala-AGC-1-1	chr6:2879596 3-28796035:-	chr6:28795963-28796135:-	chr6:28795952-28796135:-	XXbac-BPG308K3.5
GGGGAATTAGC TCAA... ...	tRNA-Ala-AGC-10-1	chr6:2668725 6-26687329:+	chr6:26687156-26687329:+	chr6:26687156-26687342:+	RP11-457M11.7

Sequence: Sequence of the tRNA isodecoder.
Gene name: Gene name of the isodecoder tRNA.
GenomeLocus: Genome locus of the tRNA isodecoder.
tRNA promoter Locus: Genome locus of the tRNA isodecoder promoter. tRNA promoter - tRNA promoters which include a tRNA gene plus 100 base pairs of upstream sequence. (PMC6108506).
pre-tRNA locus: Genome locus of the tRNA isodecoder precursor. pre-tRNA - precursor tRNA which include a tRNA gene plus 100 base pairs of upstream sequence and a 3'trailer.
tRNA neighboring gene: The nearest gene name of the tRNA isodecoder.

- *Hierarchical clustering*

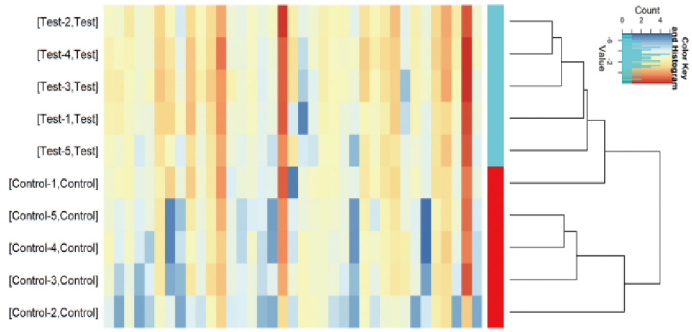

Figure 4. Hierarchical clustering heatmap of differentially expressed miRNAs.

References

[1] Godoy P.M. et al. (2018) Cell Rep [PMID: 30380423]
[2] Schageman J. et al. (2013) Biomed Res Int [PMID: 24205503]
[3] Quek C. et al. (2017) RNA Biol [PMID: 28005467]
[4] Olvedy M. et al. (2016) Oncotarget [PMID: 27015120]
[5] Dhahbi J.M. et al. (2013) BMC Genomics 14:298 [PMID: 23638709]
[6] Schageman J. et al. (2013) Biomed Res Int 2013:253957 [PMID: 24205503]
[7] Telonis A.G. et al. (2015) Oncotarget 6(28):24797-822 [PMID: 26325506]
[8] Chivukula, R. R., et al. (2014) Cell [PMID: 24855947]
[9] Kent, O. A., et al. (2010) Genes Dev [PMID: 21159816]
[10] Akao, Y., et al. (2007) Cancer Sci [PMID: 17892514]
[11] Witwer, K. W. and Halushka, M. K. (2016) RNA Biol [PMID: 27645402]
[12] Raabe, C. A., et al. (2014) Nucleic Acids Res [PMID: 24198247]
[13] Fuchs, R. T., et al. (2015) PLoS One [PMID: 25942392]
[14] Sorefan, K., et al. (2012) Silence [PMID: 22647250]
[15] Hafner, M., et al. (2011) RNA [PMID: 21775473]
[16] McCormick, K. P., et al. (2011) Silence [PMID: 21356093]
[17] Tian, G., et al. (2010) BMC Biotechnol [PMID: 20815927]
[18] Linsen, S. E., et al. (2009) Nat Methods [PMID: 19564845]
[19 Kapranov, P., et al. (2012) Methods Mol Biol [PMID: 22144202]
[20] Dabney, J. and Meyer, M. (2012) Biotechniques [PMID: 22313406]
[21] Zhao, Y., et al. (2021) J Transl Med [PMID: 34158060]
[22] Zhao, S., et al. (2020) RNA [PMID: 32284352]
[23] Chen Y., e. a. (2021)https://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf [PMID]
[24] Shi, J., et al. (2021) Nat Cell Biol [PMID: 33927371]