Arraystar LncRNA Microarrays

2019 New Releases: Human V5.0, Mouse V4.0, Rat V3.0

Highlights

- Most sensitive and best technology for lncRNA profiling, superior to RNA-seq
- Comprehensive and robust full-length lncRNA* collection curated from all major latest databases and landmark publications
- Systematic and specialized lncRNA annotation, including genomic context, epigenomic context*, completeness*, subcellular localization**, miRNA recognition site...
- Unambiguous, reliable and accurate detection and quantification of lncRNA transcript isoforms otherwise difficult by RNA-seq
- Simultaneous lncRNA and mRNA profiling on the same array for co-expressional and correlational expression and regulation

Introduction

LncRNAs are a major RNA class in the transcriptome [1]. These noncoding RNAs are transcribed from genomic sites either in association with a protein coding gene nearby or in the intergenic regions as lincRNAs (Fig. 1), with functions in gene expression regulation by multiple mechanisms, either in cis or in trans, at transcriptional or post-transcriptional levels (Fig. 2). LncRNAs are a key player in a wide range of biological systems and diseases. Cutting edge lncRNA science has resolved many long standing mysteries in, for example, chromosomal inactivation, developmental and differentiation programming, and diseases of unknown etiology. In general, lncRNAs exhibit more restricted cell type-specific expression compared to mRNAs, making lncRNAs a class of higher specificity biomarker. With the broadened horizon and modern paradigm of studying gene regulation, the science of gene expression profiling has now gone beyond past mRNA-only to encompass both classes of the coding and non-coding RNAs.

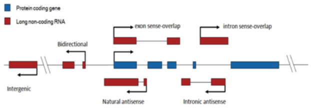

Fig1. LncRNA classification based on genomic contexts with the closest protein coding gene.

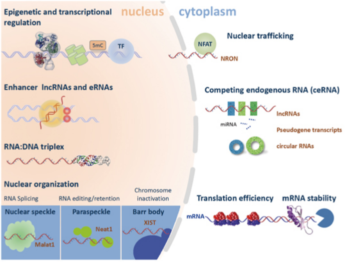

Fig 2. LncRNAs may regulate gene expression by various mechanisms, such as recruiting chromatin modifiers/remodelers to epigenetically regulate gene expression; by enhancer RNAs; by nuclear substructures; by nuclear-cytoplasmic transport; by competing endogenous RNAs via miRNAs; or by mRNA stability and translation; in cis or in trans, at transcription or post-transcriptional levels.

Arraystar is the leader in lncRNA expression profiling technologies, using LncRNA Microarrays as the best performing platform to systematically profile lncRNAs together with mRNAs. To date, these microarrays have been an empowering tool and invaluable resource in lncRNA research touting many high impact publications. To incorporate rapid scientific advances and new data, Arraystar has now released new Human V5.0 and Mouse V4.0, and Rat V3.0 LncRNA Expression Microarrays.

Consolidated, comprehensive, robust, most up-to-date full-length lncRNA contents*

Unlike well-established protein coding genes, publically available lncRNAs are often sparsely annotated, partial in scope and scattered in collection. Large proportions of reported "lncRNAs" tend to be incomplete at 5' or 3' ends. Also, RNA-seq reads are not uniform in covering the 5' and 3' ends. These inaccurate and truncated lncRNA

* Applicable to Human V5.0 ** Applicable to Human V5.0 and Mouse V4.0

annotations can have a profound impact on downstream uses of the data, such as misinterpreting mRNA fragments as lncRNAs, unreliable transcript abundance estimate by FPKM, and misidentification of lncRNA promoter sites [1,2]

Arraystar maintains high quality proprietary transcriptome and lncRNA databases that extensively collect lncRNAs through all major external data sources, knowledge-based mining of scientific publications, and our lncRNA collection pipelines. Especially, we place premium attention on full-length lncRNAs collection. Full length lncRNAs as annotated or experimentally supported in the public databases are compiled with high priority. The lncRNAs in Arraystar proprietary transcriptome databases and published lncRNA studies are carefully assessed by supporting evidence for their sequence completeness: 5' ends by host gene histone marks [3-5],CAGE cluster [6-9] , and DNA hypersensitivity (DHS) [5]data; 3' -ends by poly(A)-position profiling (3P-Seq) [10]. Additionally, lncRNA candidates are evaluated for protein coding potentials by a combination of prediction methods [11-13]. Only the lncRNAs that pass these assessments are curated into the full length lncRNA collections (Fig. 3).

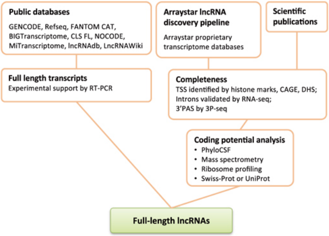
Public databases
GENCODE, Refseq, FANTOM CAT, BIGTranscriptome, CLS FL, NOCODE, MiTranscriptome, lncRNAdb, LncRNAWiki
Arraystar lncRNA discovery pipeline
Arraystar proprietary transcriptome databases
Scientific publications
Full length transcripts
Experimental support by RT-PCR
Completeness
TSS identified by histone marks, CAGE, DHS;
Introns validated by RNA-seq;
3'PAS by 3P-seq
Coding potential analysis
• PhyloCSF
• Mass spectrometry
• Ribosome profiling
• Swiss-Prot or UniProt
Full-length lncRNAs

Fig 3. Comprehensive and robust collection of full-length lncRNAs from all major sources.

For the total of 39,317 lncRNAs on Arraystar Human LncRNA Array V5.0, we further place the lncRNA collection in two tiers: 8,393 Gold Standard LncRNAs and 30,924 Reliable LncRNAs.

The Gold Standard lncRNAs are well annotated and experimentally supported genuine lncRNAs, compared with very large numbers of partial fragments, incomplete UTRs, and less reliable sequences deposited as "lncRNAs" in the public databases. The Gold Standard lncRNAs are complete with annotations of transcription units, transcript isoforms, molecular/functional mechanisms and subcellular localizations.

The Reliable lncRNAs are the comprehensive yet highly reliable lncRNA collection tier, which are the lncRNAs remained from the Gold Standard lncRNA collection. The sequences are consolidated by the transcription unit models (TU). One best transcript is selected as the representative lncRNA from each TU based on the transcript source, length, and other helpful information. 32,667 Reliable LncRNAs were constructed from 308,525 putative lncRNA sequences.

* Applicable to Human V5.0 ** Applicable to Human V5.0 and Mouse V4.0

Towards systematic and functional annotation of LncRNAs

The Arraystar LncRNA microarray package includes systematic and detailed lncRNA annotations, subclassification, and analyses to gain insight into the complex biological functions of the lncRNAs. LncRNAs with reported biological processes or associated with human diseases are researched, annotated and cross referenced. This rich source of information helps to unravel functional roles and molecular mechanisms of the LncRNAs.

Genomic context LncRNAs are systematically classified based on their genomic relationships with the nearest protein coding genes into Intergenic (LincRNA), Intronic, Bidirectional, Sense-overlapping, Antisense and Pseudogene LncRNAs (Fig. 1). These subclasses help dissect various cis- or trans-regulatory functions on the target genes transcriptionally or post-transcriptionally (Fig. 2).

Epigenomic context* lncRNAs can be transcribed in and regulated by a promoter or enhancer region with characteristic promoter or enhancer epigenetic marks [5].Many active promoter and enhancer regions are themselves transcription units, capable of generating functionally active noncoding RNAs for these cis-regulatory DNA elements. The lncRNAs are thus classified into promoter-lncRNAs (p-lncRNA) and enhancer-lncRNAs (e-lncRNA) based on the epigenomic context (e.g. DNase I hypersensitive sites). The p-lncRNAs are further grouped into intergenic and divergent p-lncRNAs based on their genomic context (Fig. 4). p-lncRNAs are often positively correlated with transcription of their protein-coding genes under the same promoters. e-lncRNAs often trap TF proteins to the local sites, modify the local chromatin environment, and organize three-dimensional nuclear topology domains for correct activation of the target gene program.

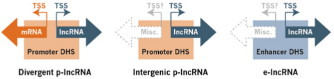

Fig 4. Promoter and enhancer lncRNA categories based on the epigenomic and genomic context. lncRNAs are classified into intergenic p-lncRNA, divergent p-lncRNA, e-lncRNA, and other, based on their TSS and DNase I hypersensitive sites (DHS) in the promoter (marked by H3K4me3), enhancer (marked by H3K4me1, H3K27ac and H3K9ac), or dyadic regulatory (enhancer-promoter alternating states) regions.

Completeness* The sequence completeness of the lncRNA 5' and 3' ends are important for many lncRNA follow-up studies, e.g. the location of the accurate lncRNA transcription start site (TSS), the promoter region, or CRISPR-Cas screen targeting site design. Here, the lncRNA end completeness statuses are annotated as: Complete 5' end, Complete 3' end, and Full length (complete both 5' and 3' end).

Subcellular localization** The molecular functions of lncRNAs are tightly coupled with their subcellular localization [14-16]. For example, lncRNAs localized in the nucleus or chromatin often regulate the gene expression by epigenetic modification and transcription. LncRNAs in the cytoplasm are more likely involved in translation regulation or miRNA sponging such as competing endogenous RNAs (ceRNA)[17-20].

miRNA recognition site Predicted or experimentally identified microRNA sites on the lncRNAs are annotated to indicate potential post-transcriptional regulatory functions in the miRNA regulatory network, such as acting as competing endogenous RNAs (ceRNA).

Highly conserved lncRNAs Certain lncRNA genes harbor ultraconserved regions (UCR) or ultraconserved non-coding elements (UCNE) that do not vary in sequence across species, which imply these sequences being biologically indispensible [21-27]. As many lncRNAs regulate target genes by cis-mechanism, human lncRNAs syntenic to orthologous lncRNAs in other species are also collected even with modest homology, as their genomic context with the neighboring target genes, rather than the sequence conservation, can be more relevant in gene regulation [28].

Tissue specific lncRNAs* The function of a lncRNA can be directly or indirectly related to and indicated by the tissue or cell type in which it is specifically expressed . In Human LncRNA Microarray, 6,059 cell lineage and cancer associated lncRNAs are annotated.

Disease-associated lncRNAs* LncRNAs known to be associated with diseases, such as cataloged in LncRNADisease [29,30], are annotated for clinical and translational investigations.

Coding potential for small peptides* Although most lncRNAs are noncoding, some lncRNAs can contain small open reading frames (smORFs) to encode small peptides [31], as predicted or experimentally detected as cataloged in LncRNAWiki [32].

*Applicable to Human V5.0 ** Applicable to Human V5.0 and Mouse V4.0

LncRNA transcript isoforms

LncRNAs, just like mRNAs, can be alternatively processed as transcript isoforms and have distinct functions. Arraystar LncRNA Microarrays use "transcript-specific" probes that hybridize to the splice junctions or exon sequences that are unique to each transcript isoform from the same gene (Fig. 5). Compared with the microarrays, RNA-seq performs poorly in transcript-specific profiling due to short sequencing reads, low isoform/splice junction read coverage, and inherent computational complexity.

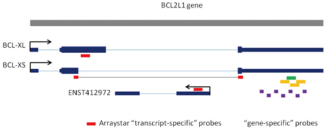

Fig 5. Arraystar LncRNA Microarray transcript-specific probes unambiguously and accurately detect and quantify transcript isoforms BCL-XL, BCL-XS, and ENST412972 having distinct oncogenic functions. The "Gene-specific" probes not designed for lncRNA isoforms cannot make such distinction. The arrows indicate the transcription direction.

Why use arraystar LncRNA microarray over RNA-seq for LncRNA profiling?

LncRNAs often express and function at low abundance, buried in other classes of abundant RNAs (Fig. 6A). There are serious limitations of RNA-seq for lncRNA profiling.

LncRNA quantification. For mere detection of the presence of a lncRNA, a few reproducible sequencing reads should suffice. But for quantification, at least hundreds read counts are required to reliably represent the RNA level [33] (Fig. 6A). LncRNAs are generally ~10X less abundant than mRNA[34]. RNA-seq quantification at these low lncRNA levels is unacceptably poor and not nearly sufficient for differential expression analysis [2,35] (Fig. 6C, 6D). Even if the sequencing coverage is increased to an unaffordably deep coverage (dotted curve, several hundred times the normal RNA-seq coverage at 20 mil), a large proportion (40%) of transcripts can never be reliably quantified [35] (Fig.6B). Additionally, FPKM (Fragments Per Kilobase of transcript per Million mapped reads) calculation in RNA-seq depends on accurate lncRNA transcript model lengths, many of which still lack completeness in lncRNA annotation [1]. In contrast, LncRNA Microarray oligo probes hybridize the target RNA at high affinity, independent of other abundant RNAs. The microarrays are highly sensitive and accurate even for low abundance lncRNAs [37](Fig. 6D).

LncRNA transcript isoforms. LncRNAs often have multiple transcript isoforms and function differently in complex genomic and regulatory relationships with their target mRNA genes. Profiling lncRNAs at transcript-specific level is important. However, RNA-seq coverage for the splice profiles is weak and non-uniform, particularly for non-predominant isoforms[2](Fig. 7). Even at saturating coverage, accurate reconstruction of transcript isoform is inherently challenging due to the missing connectivity information with the short reads in distant exons on the same RNA fragment [2]. These make reconstructing lncRNA transcript isoforms and quantification very difficult [38-41].For LncRNA Microarrays, the transcript-specific array probe design is based on well-established transcript models for each lncRNA isoform, which is unambiguous and highly accurate in isoform detection and quantification (Fig. 5).

LncRNA annotation and analysis. Unlike well established and curated protein coding genes, RNA-seq raw data are still in need of well-resourced and consolidated reference bases for mapping and annotation, which are not readily publically available. Arraystar Microarray lncRNA contents are based on the foundation of high quality proprietary Arraystar lncRNA transcriptome databases that extensively collect lncRNAs through all major public databases and repositories, knowledge-based mining of scientific publications, and our lncRNA collection pipelines. The microarray annotation and analyses are, rich, detailed, and comprehensive, unrivaled by any other profiling platforms.

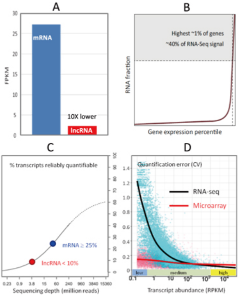

Fig 6. (A) The median lncRNA expression level is approximately 10X lower than that of mRNAs (based on GENCODE data) [34]. (B) Top 1% of the highest expressed genes, such as housekeeping genes, occupy ~40% of RNA-seq signal. Lowly expressed lncRNAs receive very little sequencing coverage [35]. (C) In a typical mRNA-seq depth at 40 million reads, < 10% lncRNAs can be reliably quantified [36]. (D) While quantitative error becomes unacceptably high for RNA-seq when the RNA level is low, microarray continues to perform very well [37].

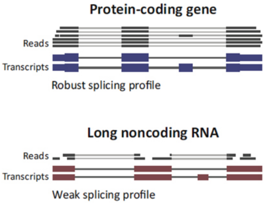

Fig 7. Compared with better expressed mRNAs, lowly expressed lncRNA isoforms cannot be adequately covered by short RNA-seq reads to reconstruct the exon models nor their quantification[2]

Table 1. LncRNA Microarray vs RNA-seq for lncRNA profiling

LncRNA Microarray	RNA-Seq
High sensitivity and quantification accuracy for lncRNAs as low as 1 transcript/cell.	Most lncRNAs at low levels cannot be accurately and reliably quantified.
Natively specific for RNA strandedness for both sense and antisense lncRNAs.	Stranded RNA-sequencing library prep required.
Unambiguous and specific lncRNA isoform detection/quantification.	Poor sensitivity and accuracy for lncRNA isoforms.
Arraystar LncRNA Microarray premium lncRNA collection, annotation and analyses. Entire coding mRNA gene set also included.	Public lncRNA reference databases can be deficient. Systematic lncRNA annotation and analyses are not readily available for the RNA-seq data.

LncRNA Study Roadmap

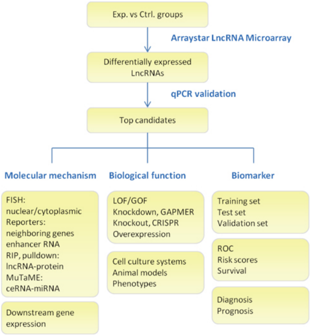

Fig 8. LncRNA research roadmap for studying the identified differentially expressed lncRNAs, for their regulatory molecular mechanisms, biological functions, and biomarker development.

Arraystar LncRNA Array Specifications

	Human V5.0	Mouse V4.0	Rat V3.0
Total number of distinct probes	60,491	60,641	38,352
Probe length	60 nt		
Probe selection region	Specific exon or splice junction along the entire length of the transcript		
Probe specificity	Transcript-specific		
Labeling method	cRNAs are labeled along the entire length without 3' bias, even for degraded RNA at low amount.		
mRNAs	21,174	30,924	27,770
lncRNAs	40,173	37,949	10,582
Gold Standard LncRNAs	8,393		
Reliable LncRNAs	30,924		
mRNA sources	Refseq, UCSC, GENECODE, FANTOM5 CAT	Refseq, Known Gene, GENECODE	Refseq, Ensembl
LncRNA sources	**Databases** (up to2018) FANTOM5 CAT (v1); GENECODE (v29); RefSeq (2018.11); BIGTranscriptome (v1); knownGene (2018.11); LncRNAdb; LncRNAWiki; RNAdb; NRED; CLS FL; NONCODE (v5); MiTranscriptome (v2)	**Databases** (up to 2018) GENE-CODE(VM19); RefSeq; KnownGene; GenBank	**Databases** (up to 2018) Refseq; Ensembl
	Literatures (up to 2018)		
	Arraystar LncRNA collection pipelines (up to 2018)		
Array format	8 × 60K	8 × 60K	4 × 44K

Selected Publications

Since Arraystar launched the first commercial LncRNA Microarray in 2009, over 400 publications citing our superior LncRNA array service were published in top scientific journals.

NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death. Huang D, et al. ***Nature immunology,*** 2018

Long Non-coding RNA GMAN, Upregulated in Gastric Cancer Tissues, is Associated with Metastasis in Patients and Promotes Translation of Ephrin A1 by Competitively Binding GMAN-AS. Zhuo W, et al. ***Gastroenterology,*** 2018

GUARDIN is a p53-responsive long non-coding RNA that is essential for genomic stability. Hu W L, et al. ***Nature Cell Biology,*** 2018

Long noncoding RNA lnc-TSI inhibits renal fibrogenesis by negatively regulating the TGF-β/Smad3 pathway. Wang P, et al. ***Science translational medicine,*** 2018

A 3-LncRNA Risk Scoring System for Prognosis of Adult Acute Myeloid Leukemia. ***Feng Y, et al. Blood,*** 2018

Long non-coding RNA NEXN-AS1 mitigates atherosclerosis by regulating the actin-binding protein NEXN. Hu Y W, et al. ***The Journal of clinical investigation,*** 2018

The LINK-A lncRNA interacts with PtdIns (3, 4, 5) P3 to hyperactivate AKT and confer resistance to AKT inhibitors. Lin A, et al. ***Nature Cell Biology,*** 2017

Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. Yong-xi Song, et al. ***Nature communications,*** 2017

A novel lncRNA GClnc1 promotes gastric carcinogenesis and may act as a modular scaffold of WDR5 and KAT2A complexes to specify the histone modification pattern. Sun T T, et al. Cancer discovery, 2016
Integrative Transcriptome Analyses of Metabolic Responses in Mice Define Pivotal LncRNA Metabolic Regulators. Ling Yang, et al. ***Cell Metabolism,*** 2016

Wnt signalling modulates transcribed-ultraconserved regions in hepatobiliary cancers. ***Carotenuto P, et al. Gut,*** 2016

References

1. Uszczynska-Ratajczak, B., et al. (2018) Nat Rev Genet. 19(9):535-548 [PMID: 29795125]
2. Deveson, I. W., et al. (2017) Trends Genet. 33(7):464-478 [PMID: 28535931]
3. Guttman, M., et al. (2009) Nature. 458(7235):223-7 [PMID: 19182780]
4. Khalil, A. M., et al. (2009) Proc Natl Acad Sci U S A. 106(28):11667-72 [PMID: 19571010]
5. Hon, C. C., et al. (2017) Nature. 543(7644):199-204 [PMID: 28241135]
6. Lagarde, J., et al. (2017) Nat Genet. 49(12):1731-1740 [PMID: 29106417]
7. Marques, A. C., et al. (2013) Genome Biol. 14(11):R131 [PMID: 24289259]
8. Alam, T., et al. (2014) PLoS One. 9(10):e109443 [PMID: 25275320]
9. Mele, M., et al. (2017) Genome Res. 27(1):27-37 [PMID: 27927715]
10. You, B. H., et al. (2017) Genome Res. 27(6):1050-1062 [PMID: 28396519]
11. Wang, L., et al. (2013) Nucleic Acids Res. 41(6):e74 [PMID: 23335781]
12. Kong, L., et al. (2007) Nucleic Acids Res. 35(Web Server issue):W345-9 [PMID: 17631615]
13. Lin, M. F., et al. (2011) Bioinformatics. 27(13):i275-82 [PMID: 21685081]
14. Kaewsapsak, P., et al. (2017) Elife. 6([PMID: 29239719]
15. Mas-Ponte, D., et al. (2017) RNA. 23(7):1080-1087 [PMID: 28386015]
16. Benoit Bouvrette, L. P., et al. (2018) RNA. 24(1):98-113 [PMID: 29079635]
17. Salmena, L., et al. (2011) Cell. 146(3):353-8 [PMID: 21802130]
18. Tay, Y., et al. (2011) Cell. 147(2):344-57 [PMID: 22000013]
19. Chan, J. J. and Tay, Y. (2018) Int J Mol Sci. 19(5[PMID: 29702599]
20. Tay, Y., et al. (2014) Nature. 505(7483):344-52 [PMID: 24429633]
21. Braconi, C., et al. (2011) Proc Natl Acad Sci U S A. 108(2):786-91 [PMID: 21187392]
22. Calin, G. A., et al. (2007) Cancer Cell. 12(3):215-29 [PMID: 17785203]
23. Esteller, M. (2011) Nat Rev Genet. 12(12):861-74 [PMID: 22094949]
24. Lujambio, A., et al. (2010) Oncogene. 29(48):6390-401 [PMID: 20802525]
25. Fabris, L. and Calin, G. A. (2017) Int Rev Cell Mol Biol. 333:159-172 [PMID: 28729024]
26. Bejerano, G., et al. (2004) Science. 304(5675):1321-5 [PMID: 15131266]
27. Dimitrieva, S. and Bucher, P. (2012) Bioinformatics. 28(18):i395-i401 [PMID: 22962458]
28. Ulitsky, I. (2016) Nat Rev Genet. 17(10):601-14 [PMID: 27573374]
29. Bao, Z., et al. (2019) Nucleic Acids Res. 47(D1):D1034-D1037 [PMID: 30285109]
30. Chen, G., et al. (2013) Nucleic Acids Res. 41(Database issue):D983-6 [PMID: 23175614]
31. Couso, J. P. and Patraquim, P. (2017) Nat Rev Mol Cell Biol. 18(9):575-589 [PMID: 28698598]
32. Ma, L., et al. (2015) Nucleic Acids Res. 43(Database issue):D187-92 [PMID: 25399417]
33. Anders, S. and Huber, W. (2010) Genome Biol. 11(10):R106 [PMID: 20979621]
34. Derrien, T., et al. (2012) Genome Res. 22(9):1775-89 [PMID: 22955988]
35. Ira W. Deveson., et al. (2017) Trends in Genetics. 33(7):464-478 [PMID: 28535931]
36. Labaj, P. P., et al. (2011) Bioinformatics. 27(13):i383-91 [PMID: 21685096]
37. Zhang, X., et al. (2010) Endocrinology. 151(3):939-47 [PMID: 20032057]
38. Consortium, S. M.-I. (2014) Nat Biotechnol. 32(9):903-14 [PMID: 25150838]
39. Liu, Y., et al. (2013) PLoS One. 8(6):e66883 [PMID: 23826166]
40. Steijger, T., et al. (2013) Nat Methods. 10(12):1177-84 [PMID: 24185837]
41. Baruzzo, G., et al. (2017) Nat Methods. 14(2):135-139 [PMID: 27941783]