high-throughput

MDPI

Opinion

High-Throughput Methods to Detect Long Non-Coding RNAs

Shizuka Uchida

Cardiovascular Innovation Institute, University of Louisville, 302 E Muhammad Ali Blvd, Louisville, KY 40202, USA; heart.lncrna@gmail.com; Tel.: +1-502-854-0570; Fax: +1-502-852-7195

Academic Editor: Massimo Negrini
Received: 21 August 2017; Accepted: 29 August 2017; Published: 31 August 2017

Abstract: Increasing evidence suggests that the numbers of long non-coding RNAs (lncRNAs) are more than those of protein-coding genes in various organisms. Although the detection methods for lncRNAs are being increasingly established, there are advantages and disadvantages that exist for each method. In this opinion article, I highlight the differences between microarrays and RNA sequencing (RNA-seq) for the detection of lncRNAs. Compared to RNA-seq, microarrays are limited to the known sequences. However, the detection method as well as data analysis workflow is more established, which makes it easier to analyze the data for bench scientists without extensive knowledge about computer programming. In order to highlight the usage of microarrays over RNA-seq for the detection of lncRNAs, we are organizing a special issue for *High-Throughput* called "Microarrays in Non-Coding RNAs Profiling", which will include the specific usages of microarrays for lncRNAs.

Keywords: long non-coding RNA; non-coding RNA; microarray; RNA-seq

---

It is now firmly recognized that coding parts of protein-coding genes occupy a very minor part of the mammalian genome, although a vast majority of the mammalian genome is transcribed to RNA [1–3]. Currently, those RNAs that do not encode for proteins are collectively called "non-coding RNAs (ncRNAs)", which are further categorized by their lengths rather than their functionalities. Although the functions of small RNAs, including microRNAs (miRNAs), have been elucidated in the past two decades, potential functions of longer ncRNAs are understudied and just recently being investigated in various fields of study. Among longer ncRNAs, those longer than 200 nucleotides are classified as "long non-coding RNAs (lncRNAs)", which were once considered "junk DNA" [4]. As technology advances, more and more lncRNAs are discovered and reported, although their functions remain mostly unknown. Although their functions are unknown, it is speculated that the number of lncRNAs increases as organisms move up in the evolutionary ladder [5]. Thus, understanding lncRNAs might shed light on the complexity of organisms during evolution [6–13]. Also, lncRNAs are involved in a variety of cellular processes [14–20] and their dysregulations are linked to some diseases [21–26].

To detect lncRNAs, there are two most common methods: microarrays and RNA sequencing (RNA-seq) using next generation sequencers. The former is well-established method, which has a rather long history for the detection of protein-coding genes. Given that probes can be designed as long as sequences are known, it is not surprising that microarrays can be designed to detect lncRNAs. More importantly, the previously built microarrays contain probe sequences that match lncRNAs, although they were initially designed for protein-coding genes. This is due partially to the re-annotation of previously thought protein-coding genes as lncRNAs. Recently, we and others re-annotated the probe sequences of microarrays and found that many lncRNAs can be recorded from Affymetrix-based GeneChips as these types of microarrays contain many probe sequences that do not

match exons of protein-coding genes [27,28]. In the case of RNA-seq, unbiased genome-wide screening of lncRNAs is possible. Currently, there are two main methods of generating a sequencing library for RNA-seq experiments. One is to use oligo dT beads to enrich mRNAs with poly A tails. Another is to deplete ribosomal RNAs (rRNAs) from the total RNA to enrich for RNAs that are not rRNA as rRNA constitutes ~80% of total RNA followed by 15% transfer RNAs (tRNAs) and only 5% for all other RNAs, including protein-coding genes and lncRNAs [29]. The former method will result in the identification of protein-coding genes and lncRNAs with poly A tails (~60% of total lncRNAs) [30], while the latter can identify the rest of lncRNAs and a newly emerging class of lncRNAs called "circular RNAs (circRNAs)" [31–34], in addition to those identified in the former method. The presence of circRNAs is only detected with the latter method as circRNAs arise from exons and/or introns that are spliced out, which are devoid of poly-A tails. Compared to microarrays, the data analysis of RNA-seq is still a matter of debate as there are many algorithms available to normalize the data and to quantify the transcripts, for which no one method is superior over others [35–38]. Furthermore, there is an ongoing debate about the quantification of RNA-seq reads (e.g., RPKM (Reads Per Kilobase Million), FPKM (Fragments Per Kilobase Million), and TPM (Transcripts Per Kilobase Million)), which makes the RNA-seq data analysis complicated compared to that of microarrays.

Although RNA-seq offers more comprehensive coverage of whole transcriptomes compared to microarrays, there is a question of how deep the RNA-seq reads should be. It is commonly accepted that at least 10–20 million reads are needed to be comparable with the expression profiling performed via microarrays [39]. However, this is based on the poly-A-based RNA-seq, which cover protein-coding genes and ~60% of lncRNAs but not the remaining lncRNAs without poly-A tails and circRNAs. Thus, it is not as comprehensive coverage of whole transcriptome one would expect. Given such situation, microarrays may perform better for surveying the known sequences. It is absolutely true that unknown (and novel) lncRNAs will be discovered via RNA-seq than microarrays. However, a caution is needed as many of such novel lncRNAs contain many repetitive elements as we recently reported [40]. Furthermore, the mapping rate of sequence reads depends on the version of reference genome, such that the most recent human genome assembly GRCh38/hg38 has much better mapping rate than GRCh37/hg19 [40]. In addition, the computational power needed to analyze RNA-seq data is much more demanding than that of microarrays. Furthermore, the involvement of experienced bioinformaticians is required to analyze RNA-seq data properly, which is not the case for microarray data analysis as this technology is much more mature than RNA-seq and many well-established software and bioinformatics tools are available for microarray data analysis.

One additional point about the difference between microarrays and RNA-seq is that, in most cases, microarrays do not involve PCR-based amplification of sample probes (especially in the case of gene/exon arrays of Affymetrix), whereas RNA-seq requires multiple cycles of PCR amplifications. Given that PCR amplifications might introduce a certain bias, it may skew the data for GC-rich regions. Taken together, although RNA-seq is more powerful on detecting novel lncRNAs as well as splicing and RNA modifications (e.g., A-to-I RNA editing [41,42]), if one is interested in finding out about known signaling pathways and lncRNAs (although not many), it is better off using microarrays than RNA-seq as the protocol and data analysis are more established. In order to highlight the usage of microarrays over RNA-seq for the detection of lncRNAs, we are organizing a special issue for *High-Throughput* called "Microarrays in Non-Coding RNAs Profiling", which will include the specific usages of microarrays for lncRNAs.

Acknowledgments: This manuscript was supported by the V.V. Cooke Foundation (Kentucky, USA); and the startup funding from the Mansbach Family, the Gheens Foundation and other generous supporters at the University of Louisville.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [CrossRef] [PubMed]
2. Uchida, S.; Dimmeler, S. Long noncoding RNAs in cardiovascular diseases. *Circ. Res.* **2015**, *116*, 737–750. [CrossRef] [PubMed]
3. Uchida, S.; Gellert, P.; Braun, T. Deeply dissecting stemness: Making sense to non-coding RNAs in stem cells. *Stem Cell Rev.* **2012**, *8*, 78–86. [CrossRef] [PubMed]
4. Flynn, R.A.; Chang, H.Y. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* **2014**, *14*, 752–761. [CrossRef] [PubMed]
5. Necsulea, A.; Soumillon, M.; Warnefors, M.; Liechti, A.; Daish, T.; Zeller, U.; Baker, J.C.; Grutzner, F.; Kaessmann, H. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **2014**, *505*, 635–640. [CrossRef] [PubMed]
6. Gaiti, F.; Fernandez-Valverde, S.L.; Nakanishi, N.; Calcino, A.D.; Yanai, I.; Tanurdzic, M.; Degnan, B.M. Dynamic and widespread lncRNA expression in a sponge and the origin of animal complexity. *Mol. Biol. Evol.* **2015**, *32*, 2367–2382. [CrossRef] [PubMed]
7. Soumillon, M.; Necsulea, A.; Weier, M.; Brawand, D.; Zhang, X.; Gu, H.; Barthes, P.; Kokkinaki, M.; Nef, S.; Gnirke, A.; et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **2013**, *3*, 2179–2190. [CrossRef] [PubMed]
8. Kutter, C.; Watt, S.; Stefflova, K.; Wilson, M.D.; Goncalves, A.; Ponting, C.P.; Odom, D.T.; Marques, A.C. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **2012**, *8*, e1002841. [CrossRef] [PubMed]
9. Marques, A.C.; Ponting, C.P. Catalogues of mammalian long noncoding RNAs: Modest conservation and incompleteness. *Genome Biol.* **2009**, *10*, R124. [CrossRef] [PubMed]
10. Managadze, D.; Rogozin, I.B.; Chernikova, D.; Shabalina, S.A.; Koonin, E.V. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol. Evol.* **2011**, *3*, 1390–1404. [CrossRef] [PubMed]
11. Quinn, J.J.; Zhang, Q.C.; Georgiev, P.; Ilik, I.A.; Akhtar, A.; Chang, H.Y. Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes Dev.* **2016**, *30*, 191–207. [CrossRef] [PubMed]
12. Zhang, R.; Deng, P.; Jacobson, D.; Li, J.B. Evolutionary analysis reveals regulatory and functional landscape of coding and non-coding RNA editing. *PLoS Genet.* **2017**, *13*, e1006563. [CrossRef] [PubMed]
13. Hezroni, H.; Koppstein, D.; Schwartz, M.G.; Avrutin, A.; Bartel, D.P.; Ulitsky, I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **2015**, *11*, 1110–1122. [CrossRef] [PubMed]
14. Ayoub, N.; Richler, C.; Wahrman, J. Xist RNA is associated with the transcriptionally inactive XY body in mammalian male meiosis. *Chromosoma* **1997**, *106*, 1–10. [CrossRef] [PubMed]
15. Mele, M.; Rinn, J.L. "Cat's cradling" the 3D genome by the act of lncRNA transcription. *Mol. Cell* **2016**, *62*, 657–664. [CrossRef] [PubMed]
16. Melissari, M.T.; Grote, P. Roles for long non-coding RNAs in physiology and disease. *Pflugers Arch.* **2016**, *468*, 945–958. [CrossRef] [PubMed]
17. Sun, X.; Haider Ali, M.S.S.; Moran, M. The role of interactions of long non-coding RNAs and heterogeneous nuclear ribonucleoproteins in regulating cellular functions. *Biochem. J.* **2017**, *474*, 2925–2935. [CrossRef] [PubMed]
18. Bischoff, F.C.; Werner, A.; John, D.; Boeckel, J.N.; Melissari, M.T.; Grote, P.; Glaser, S.F.; Demolli, S.; Uchida, S.; Michalik, K.M.; et al. Identification and functional characterization of hypoxia-induced endoplasmic reticulum stress regulating lncRNA (hyperlnc) in pericytes. *Circ. Res.* **2017**, *121*, 368–375. [CrossRef] [PubMed]
19. Leisegang, M.S.; Fork, C.; Josipovic, I.; Richter, F.M.; Preussner, J.; Hu, J.; Miller, M.J.; Epah, J.; Hofmann, P.; Gunther, S.; et al. Long noncoding RNA mantis facilitates endothelial angiogenic function. *Circulation* **2017**, *136*, 65–79. [CrossRef] [PubMed]

20. Michalik, K.M.; You, X.; Manavski, Y.; Doddaballapur, A.; Zornig, M.; Braun, T.; John, D.; Ponomareva, Y.; Chen, W.; Uchida, S.; et al. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ. Res.* **2014**, *114*, 1389–1397. [CrossRef] [PubMed]
21. Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **2015**, *21*, 1253–1261. [CrossRef] [PubMed]
22. Piccoli, M.T.; Gupta, S.; Viereck, J.; Foinquinos, A.; Samolovac, S.; Kramer, F.; Garg, A.; Remke, J.; Zimmer, K.; Batkai, S.; et al. Inhibition of the cardiac fibroblast-enriched lncRNA Meg3 prevents cardiac fibrosis and diastolic dysfunction. *Circ. Res.* **2017**, *121*, 575–583. [CrossRef] [PubMed]
23. Atanasovska, B.; Rensen, S.S.; van der Sijde, M.R.; Marsman, G.; Kumar, V.; Jonkers, I.; Withoff, S.; Shiri-Sverdlov, R.; Greve, J.W.M.; Faber, K.N.; et al. A liver-specific long noncoding RNA with a role in cell viability is elevated in human nonalcoholic steatohepatitis. *Hepatology* **2017**, *66*, 794–808. [CrossRef] [PubMed]
24. Fava, V.M.; Manry, J.; Cobat, A.; Orlova, M.; Van Thuc, N.; Moraes, M.O.; Sales-Marques, C.; Stefani, M.M.; Latini, A.C.; Belone, A.F.; et al. A genome wide association study identifies a lncRNA as risk factor for pathological inflammatory responses in leprosy. *PLoS Genet.* **2017**, *13*, e1006637. [CrossRef] [PubMed]
25. Greco, S.; Zaccagnini, G.; Fuschi, P.; Voellenkle, C.; Carrara, M.; Sadeghi, I.; Bearzi, C.; Maimone, B.; Castelvecchio, S.; Stellos, K.; et al. Increased BACE1-AS long noncoding RNA and β-amyloid levels in heart failure. *Cardiovasc. Res.* **2017**, *113*, 453–463. [CrossRef] [PubMed]
26. Micheletti, R.; Plaisance, I.; Abraham, B.J.; Sarre, A.; Ting, C.C.; Alexanian, M.; Maric, D.; Maison, D.; Nemir, M.; Young, R.A.; et al. The long noncoding RNA wisper controls cardiac fibrosis and remodeling. *Sci. Transl. Med.* **2017**, *9*. [CrossRef] [PubMed]
27. Gellert, P.; Ponomareva, Y.; Braun, T.; Uchida, S. Noncoder: A web interface for exon array-based detection of long non-coding RNAs. *Nucleic Acids Res.* **2013**, *41*, e20. [CrossRef] [PubMed]
28. Liao, Q.; Xiao, H.; Bu, D.; Xie, C.; Miao, R.; Luo, H.; Zhao, G.; Yu, K.; Zhao, H.; Skogerbo, G.; et al. ncFANS: A web server for functional annotation of long non-coding RNAs. *Nucleic acids Res.* **2011**, *39*, W118–W124. [CrossRef] [PubMed]
29. Lodish, H.; Berk, A.; Zipursky, L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*, 4th ed.; W. H. Freeman: New York, NY, USA, 2000.
30. Cheng, J.; Kapranov, P.; Drenkow, J.; Dike, S.; Brubaker, S.; Patel, S.; Long, J.; Stern, D.; Tammana, H.; Helt, G.; et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **2005**, *308*, 1149–1154. [CrossRef] [PubMed]
31. Jeck, W.R.; Sharpless, N.E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* **2014**, *32*, 453–461. [CrossRef] [PubMed]
32. Jeck, W.R.; Sorrentino, J.A.; Wang, K.; Slevin, M.K.; Burd, C.E.; Liu, J.; Marzluff, W.F.; Sharpless, N.E. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **2013**, *19*, 141–157. [CrossRef] [PubMed]
33. Boeckel, J.N.; Jae, N.; Heumuller, A.W.; Chen, W.; Boon, R.A.; Stellos, K.; Zeiher, A.M.; John, D.; Uchida, S.; Dimmeler, S. Identification and characterization of hypoxia-regulated endothelial circular RNA. *Circ. Res.* **2015**, *117*, 884–890. [CrossRef] [PubMed]
34. Militello, G.; Weirick, T.; John, D.; Doring, C.; Dimmeler, S.; Uchida, S. Screening and validation of lncRNAs and circRNAs as miRNA sponges. *Brief. Bioinform.* **2016**. [CrossRef] [PubMed]
35. Engstrom, P.G.; Steijger, T.; Sipos, B.; Grant, G.R.; Kahles, A.; Ratsch, G.; Goldman, N.; Hubbard, T.J.; Harrow, J.; Guigo, R.; et al. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **2013**, *10*, 1185–1191. [CrossRef] [PubMed]
36. Steijger, T.; Abril, J.F.; Engstrom, P.G.; Kokocinski, F.; Hubbard, T.J.; Guigo, R.; Harrow, J.; Bertone, P. Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* **2013**, *10*, 1177–1184. [CrossRef] [PubMed]
37. Eder, T.; Grebien, F.; Rattei, T. Nvt: A fast and simple tool for the assessment of RNA-seq normalization strategies. *Bioinformatics* **2016**, *32*, 3682–3684. [CrossRef] [PubMed]
38. Evans, C.; Hardin, J.; Stoebel, D.M. Selecting between-sample RNA-seq normalization methods from the perspective of their assumptions. *Brief. Bioinform.* **2017**. [CrossRef]
39. Wang, Y.; Ghaffari, N.; Johnson, C.D.; Braga-Neto, U.M.; Wang, H.; Chen, R.; Zhou, H. Evaluation of the coverage and depth of transcriptome by RNA-seq in chickens. *BMC Bioinform.* **2011**, *12* (Suppl. 10), S5. [CrossRef] [PubMed]

40. Weirick, T.; Militello, G.; Muller, R.; John, D.; Dimmeler, S.; Uchida, S. The identification and characterization of novel transcripts from RNA-seq data. *Brief. Bioinform.* **2016**, *17*, 678–685. [CrossRef] [PubMed]
41. John, D.; Weirick, T.; Dimmeler, S.; Uchida, S. Rnaeditor: Easy detection of RNA editing events and the introduction of editing islands. *Brief. Bioinform.* **2016**. [CrossRef] [PubMed]
42. Stellos, K.; Gatsiou, A.; Stamatelopoulos, K.; Perisic Matic, L.; John, D.; Lunella, F.F.; Jae, N.; Rossbach, O.; Amrhein, C.; Sigala, F.; et al. Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. *Nat. Med.* **2016**, *22*, 1140–1150. [CrossRef] [PubMed]

© 2017 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).