MEDICAL SCIENCE MONITOR BASIC RESEARCH

MOLECULAR BIOLOGY

eISSN 2325-4416
© Med Sci Monit Basic Res, 2014; 20: 138-141
DOI: 10.12659/MSMBR.892101

Received: 2014.07.27
Accepted: 2014.07.27
Published: 2014.08.23

Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

EF 1 Kirk J Mantione
F 1 Richard M. Kream
F 2,3 Hana Kuzelova
F 2 Radek Ptacek
F 2 Jiri Raboch
E 1 Joshua M. Samuel
E 1 George B. Stefano

1 Neuroscience Research Institute, State University of New York, College at Old Westbury, Old Westbury, NY, U.S.A.
2 Center for Molecular and Cognitive Neuroscience, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic
3 Department of Biology and Medical Genetics, 2nd Faculty of Medicine, Charles University in Prague, Prague, Czech Republic

Corresponding Author: Kirk J. Mantione, e-mail: kmantione@sunynri.org
Source of support: KJM, RMK, JMS and GBS, in part, have been supported by Mitogenetics, Corp. (Sioux Falls, South Dakota)

Understanding the control of gene expression is critical for our understanding of the relationship between genotype and phenotype. The need for reliable assessment of transcript abundance in biological samples has driven scientists to develop novel technologies such as DNA microarray and RNA-Seq to meet this demand. This review focuses on comparing the two most useful methods for whole transcriptome gene expression profiling. Microarrays are reliable and more cost effective than RNA-Seq for gene expression profiling in model organisms. RNA-Seq will eventually be used more routinely than microarray, but right now the techniques can be complementary to each other. Microarrays will not become obsolete but might be relegated to only a few uses. RNA-Seq clearly has a bright future in bioinformatic data collection.

MeSH Keywords: Gene Expression Profiling • High-Throughput Nucleotide Sequencing • Microarray Analysis

Full-text PDF: http://www.basic.medscimonit.com/abstract/index/idArt/892101

1286 — 1 25

Indexed in: [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

Background

Understanding the control of gene expression is critical for our understanding of the relationship between genotype and phenotype. The need for reliable assessment of transcript abundance in biological samples has driven scientists to develop novel technologies such as DNA microarray to meet this demand. Microarrays employ nucleic acid probes, typically 60-mers, covalently bound to glass slides. Fluorescently labeled target sequences are then hybridized to the probes and scanned. The images are then converted to signal intensities and the data is processed using software specific to the application of the array. For more quantitatively accurate measurement and to obtain absolute transcript abundance, RNA-Seq has become the favored technique [1,2]. RNA-Seq sequences labeled cDNA in parallel and multiple times, sometimes several million times over. The technique requires fragmenting RNA prior to reverse transcription and labeling with adapter sequences. The sequenced fragmented transcripts are typically 50–500 bp. The read sequences are then counted and assembled into full length transcripts. This review will compare two of the most useful gene expression profiling methods namely, gene expression microarrays and RNA-Seq. Much of the recent literature comparing these techniques has been focused on the Affymetrix microarray platform with the Illumina RNA-Seq method [3–6]. The utility and reproducibility of these methods for gene expression profiling are both well documented [3,7].

Discussion

The difference between the capabilities of each method becomes apparent once the target sequences go beyond known genomic sequences. Hybridization-based techniques like microarray rely on and are limited to the transcripts bound to the array slides. Microarrays are only as good as the bioinformatic data available for the model organism's genome and transcriptome. RNA-Seq also detects annotated transcripts but also will detect novel sequences and splice variants [8]. RNA-Seq can use data from the same experiment to characterize exon junctions, detect non-coding RNA [9], detect single nucleotide polymorphisms, and detect fusion genes [10]. Furthermore, existing data sets can be re-evaluated as new sequences are annotated [11]. Microarrays can detect single nucleotide polymorphisms, map exon junctions, and detect fusion genes but only with arrays designed for those purposes. Annotation of non-coding RNA needs to be completed and included on specialized chips before it can be accurately distinguished by microarray. Non-coding RNA (small RNA) detection by RNA-Seq requires some modifications in sample preparation procedures to exclude larger RNA sequences prior to cDNA generation. Finally, unlike RNA-Seq, microarray chips need to be updated to contain the most up to date sequence information.

The utility of RNA-Seq for other bioinformatic studies besides gene expression profiling far exceeds that of a microarray. RNA-Seq is useful to distinguish host from parasite transcripts, study symbioses, and examine transcripts from non-model organisms, including bacteria [8,12,13]. Monitoring temporal changes in transcript abundance of planktonic bacteria would be nearly impossible without RNA-Seq [14]. Researchers recently have started to examine epigenetic processes using RNA-Seq [15]. The study of epigenetics will certain benefit from the continued use of RNA-Seq in basic research.

RNA-Seq can achieve higher resolution of differentially expressed genes and has a much lower limit of detection than a standard whole genome microarray [6]. In fact, due to the digital nature of RNA-Seq, there is an unlimited dynamic range of detection. Arrays must be customized to have a greater probe density for resolving or detecting low abundance transcripts. The RNA-Seq method to determine differentially expressed genes does have an inherent bias towards longer transcripts [16,17]. The sample processing method involves fragmenting transcripts. The longer the transcript the more fragments available for sequencing. Microarrays do not have this length bias and expression levels are proportional to the degree of hybridization to probes. The only bias that exists in microarray hybridization would be due to the differences in the GC content of the probes used. In addition, biases in both methods exist for higher abundance transcripts and underscore the need for validation of results. Typically, validation of differentially expressed genes can be achieved by quantitative PCR or proteomic methods [4].

Sample handling methods for both techniques start with isolation of total RNA followed by production of cDNA by reverse transcription (Figure 1). RNA-Seq methods require fragmentation and attaching specific sequence linkers to the RNA prior to cDNA production (Figure 1, right side). The adapter-ligated sequences are then ready for reading on the appropriate analyzer. Methods between RNA-Seq platforms differ and RNA labeling methods and preparation of RNA prior to reverse transcription could differ even when using the same platform [2,18]. Some RNA-Seq methods need as little as 10pg of RNA to start whereas microarrays can start with as little as 200 ng of total RNA. A typical gene expression profiling method would use about the same amount of RNA in both methods, generally about 1ug is sufficient. The source of starting tissue can be fresh tissue or frozen tissue and both methods can be adapted to work with formalin fixed paraffin embedded material if necessary [19]. The length of time (4–6 h) to prepare a labeled cDNA or cRNA suitable for microarray or RNA-Seq is about the same. The microarray hybridization step occurs after the labeled cRNA is prepared and purified (Figure1). Hybridization takes 17h and the array slide is washed for a few minutes before the array is scanned and analyzed (Figure 1).

Indexed in: [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS] [Index Copernicus]

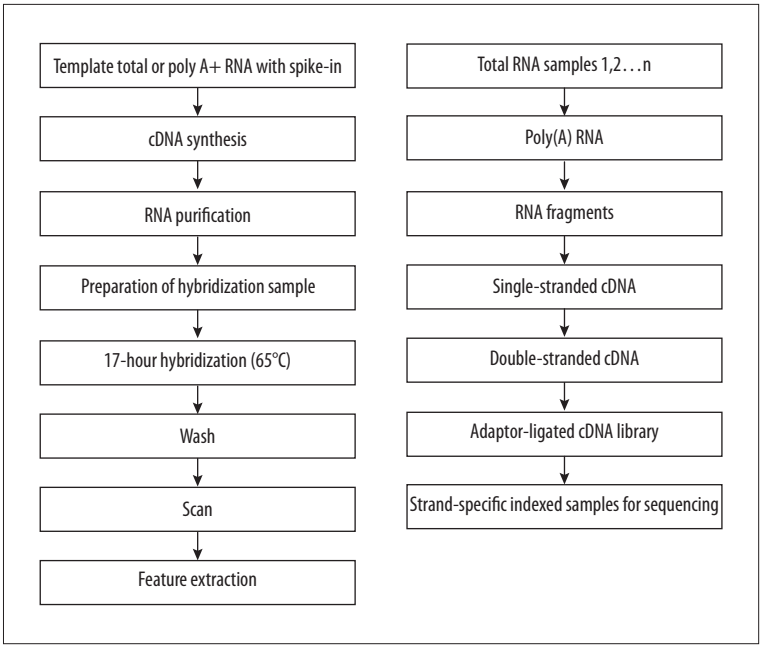

Figure 1. Workflow of sample preparation for Agilent array processing (left) and workflow for strand specific RNA-Seq sample preparation for Illumina platform (right). Adapted from Agilent product package inserts.

Statistical tests for each method require evaluating the null hypothesis that a gene is not differentially expressed between two treatment groups or disease states after calculating *p* values [3] using a t test for microarrays and a Fisher Exact Test for RNA-Seq. Microarrays are capable of detecting a 2 fold change with great reliability while RNA-Seq has far greater resolution and can accurately measure a 1.25 fold change. Over the last decade, data analysis for microarray has become easier for the average user. The software available is user friendly and many software packages are free of charge. The protocols are also more universally applicable and comparable across all platforms. For RNA-Seq, there are many data analysis methods available but not one standard protocol [20]. Analysis of RNA-Seq data also requires extensive experience and the bioinformatics skills necessary to process the data files. The data analysis techniques not only differ in the type of software used to initially reduce the data sets [20] but also for each use of RNA-Seq [21,22]. The size of an average raw data file from an Agilent microarray is 0.7 MB while the normal size of uncompressed RNA-Seq raw file is approximately 5GB. Data sharing in RNA-Seq becomes extremely difficult and the cost to store data is also greater. Overall, it costs about $300 per sample for microarray and up to $1000 per sample for RNA-Seq.

Conclusions

The complicated nature of RNA-Seq data analysis will certainly be mitigated as advances in software and newer techniques are invented. The sequencing technologies are rapidly advancing from second generation techniques to third generation and beyond. The cost of RNA-Seq will certainly drop over time. As of today, however, microarrays are reliable and more cost effective than RNA-Seq for gene expression profiling in model organisms. Our laboratory routinely uses microarray for gene expression profiling in human cells [23–25]. We also find novel, useful, and non-obvious information from examining the pattern of gene expression across large numbers of samples that can be quickly and easily generated in a highly reproducible manner via gene expression microarray. For clinical applications, microarrays have been used for a longer period of time and will probably have regulatory approvals for diagnostic use prior to RNA-Seq obtaining approvals. RNA-Seq will eventually be used more routinely than microarray, but right now the techniques can be complementary to each other. Microarrays will not become obsolete but might be relegated to only a few uses. RNA-Seq clearly has a bright future in bioinformatic data collection.

References:

1. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 2009; 10: 57–63
2. Marguerat S, Bahler J: RNA-seq: from technology to biology. Cell Mol Life Sci, 2010; 67: 569–79
3. Marioni JC, Mason CE, Mane SM et al: RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res, 2008; 18: 1509–17
4. Fu X, Fu N, Guo S et al: Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics, 2009; 10: 161
5. Xu X, Zhang Y, Williams J et al: Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. BMC Bioinformatics, 2013; 14(Suppl.9): S1

Indexed in: [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS] [Index Copernicus]

6. Zhao S, Fung-Leung WP, Bittner A et al: Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One, 2014; 9: e78644
7. Hockley SL, Mathijs K, Staal YC et al: Interlaboratory and Interplatform Comparison of Microarray Gene Expression Analysis of HepG2 Cells Exposed to Benzo(a)pyrene. OMICS, 2009; 13: 115–25
8. Howard BE, Hu Q, Babaoglu AC et al: High-throughput RNA sequencing of pseudomonas-infected Arabidopsis reveals hidden transcriptome complexity and novel splice variants. PLoS One, 2013; 8: e74183
9. Arnvig KB, Comas I, Thomson NR et al: Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. PLoS Pathog, 2011; 7: e1002342
10. Maher CA, Kumar-Sinha C, Cao X et al: Transcriptome sequencing to detect gene fusions in cancer. Nature, 2009; 458: 97–101
11. Roberts A, Schaeffer L, Pachter L: Updating RNA-Seq analyses after re-annotation. Bioinformatics, 2013; 29: 1631–37
12. Croucher NJ, Thomson NR: Studying bacterial transcriptomes using RNA-seq. Curr Opin Microbiol, 2010; 13: 619–24
13. Perkins TT, Kingsley RA, Fookes MC et al: A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. PLoS Genet, 2009; 5: e1000569
14. Ottesen EA, Young CR, Eppley JM et al: Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci USA, 2013; 110: E488–97
15. Liew LC, Singh MB, Bhalla PL: An RNA-seq transcriptome analysis of histone modifiers and RNA silencing genes in soybean during floral initiation process. PLoS One, 2013; 8: e77502
16. Oshlack A, Wakefield MJ: Transcript length bias in RNA-seq data confounds systems biology. Biol Direct, 2009; 4: 14
17. Young MD, Wakefield MJ, Smyth GK, Oshlack A: Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol, 2010; 11: R14
18. Croucher NJ, Fookes MC, Perkins TT et al: A simple method for directional transcriptome sequencing using Illumina technology. Nucleic Acids Res, 2009; 37: e148
19. Zhao W, He X, Hoadley KA et al: Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics, 2014; 15: 419
20. Trapnell C, Roberts A, Goff L et al: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc, 2012; 7: 562–78
21. Drewe P, Stegle O, Hartmann L et al: Accurate detection of differential RNA processing. Nucleic Acids Res, 2013; 41: 5189–98
22. Zhang Z, Wang W: RNA-Skim: a rapid method for RNA-Seq quantification at transcript level. Bioinformatics, 2014; 30: i283–92
23. Stefano GB, Burrill JD, Labur S et al: Regulation of various genes in human leukocytes acutely exposed to morphine: Expression microarray analysis. Med Sci Monit, 2005; 11(5): MS35–42
24. Kim C, Cadet P: Environmental Toxin 4-Nonylphenol and Autoimmune Diseases: Using DNA Microarray to Examine Genetic Markers of Cytokine Expression. Arch Med Sci, 2010; 6: 321–27
25. Kim A, Jung BH, Cadet P: A novel pathway by which the environmental toxin 4-Nonylphenol may promote an inflammatory response in inflammatory bowel disease. Med Sci Monit Basic Res, 2014; 20: 47–54

Indexed in: [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS] [Index Copernicus]