



Genome Sequencer System

Application Note No. 6 / March 2007



Comprehensive Analysis of MicroRNAs using the Genome Sequencer System



Comprehensive Analysis of MicroRNAs using the Genome Sequencer System

Fritz Thümmler^{1*}, Ronald Plasterk², Eugene Berezikov², Edwin Cuppen²

Introduction

MicroRNAs (miRNAs) are ~22 nt RNA molecules that are processed from larger hairpin precursors and can regulate gene expression (1). The role of miRNAs in diverse developmental processes and disease is increasingly recognized (2). Several hundred miRNA genes have been identified by sequencing size-fractionated small RNA libraries in human and other vertebrates (3), and computational analyses indicated that there could be substantially more miRNAs in the human genome (4,5).

To discover miRNAs that escaped cloning in previous studies, we used the Genome Sequencer System for deep sequencing small RNA libraries prepared from human fetal brain and specific regions of adult chimpanzee brain (6). Our results show that this approach is suitable for miRNA discovery as well as miRNA expression profiling.

Materials and Methods

Materials

Equipment:

Genome Sequencer Instrument (20 or FLX)
(GS 20: Software Version 1.0.53 or
GS FLX: Software Version 1.1.01)

Reagents from Roche Applied Science:

Sample Preparation:

GS emPCR Kit II (Amplicon A, Paired End);
GS emPCR Kit III (Amplicon B),
FastStart High Fidelity PCR System

Sequencing: GS 20 Sequencing Kit (40x75 or 70x75)*;
GS SR70 Sequencing Kit**
GS LR70 Sequencing Kit**
GS PicoTiterPlate Kit (40x75* or 70*x75**)

* for GS 20

** for GS FLX

Methods

For complete details on the Library preparation procedure and sequencing, please refer to the GS Guide to Amplicon Sequencing, the GS emPCR Kit User's Manual, and the Operator's Manual appropriate for your Genome Sequencer Instrument.

The sequence data is analyzed as raw reads in a FASTA file.

*Corresponding author: Fritz Thümmler, vertis Biotechnologie AG, Germany, Email: thuemmler@vertis-biotech.com

1) vertis Biotechnologie AG, Lise-Meitner-Str. 30, D-85354 Freising, Germany

2) Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

Procedure

Library Construction

To synthesize cDNA from microRNA, we utilized a method developed by Vertis Biotechnologie AG. A critical step in the method is the controlled polyadenylation of the 3'-OH ends of purified miRNA. This very reliable and sensitive method results in cDNA that is ready for sequencing with the Genome Sequencer System.

Tissues used for cDNA synthesis were human fetal brain (mixed composition) and adult chimpanzee brain regions. Tissues were ground under liquid

nitrogen and RNA species smaller than 200 nt were enriched using the *mirVana* miRNA Isolation Kit (Ambion, Austin, Texas, USA). The small RNAs were then separated on a denaturing 12.5% polyacrylamide (PAA) gel. As molecular mass standard, a mixture of oligonucleotides that range in size from 15 to 30 bases was used (Figure 1). The population of small RNA in this size range, including the main fraction of mature miRNA (20-22 nt), was obtained by passive elution of the RNAs from the gel.

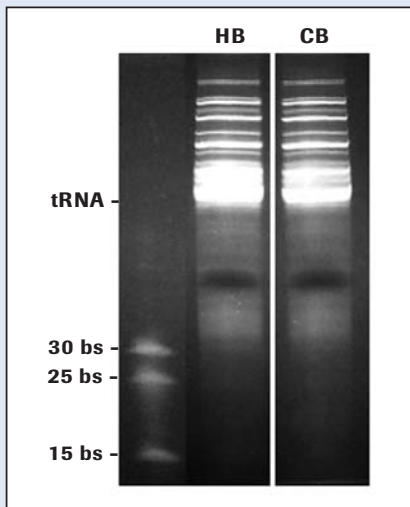


Figure 1: Analysis of small RNAs from human (lane HB) and chimpanzee (lane CB) brain on a denaturing 12.5% polyacrylamide gel.

For cDNA synthesis, the miRNAs were first poly(A)-tailed using poly(A) polymerase followed by ligation of a specific RNA adapter to the 5' phosphate of the miRNAs. First-strand cDNA synthesis was then performed using an oligo(dT)-adapter primer and M-MLV-RNase H- reverse transcriptase. The resulting cDNAs were then PCR-amplified in 22 cycles using Taq polymerase. The 5' RNA adapter, the oligo(dT)-adapter, and the primers used for PCR amplification were designed for amplicon sequencing according to the instructions of the GS Guide to Amplicon Sequencing. The combined length of the flanking sequences is 105 bases. Therefore, PCR products containing miRNA cDNAs of 15-30 bp must have a total length of 120 to 135 bp. PCR products of this size range were obtained by purification on a 6% PAA gel. For both samples, cDNA yields were approximately 200 ng. The gel-purified cDNAs were analyzed on a 4% agarose gel (Figure 2).

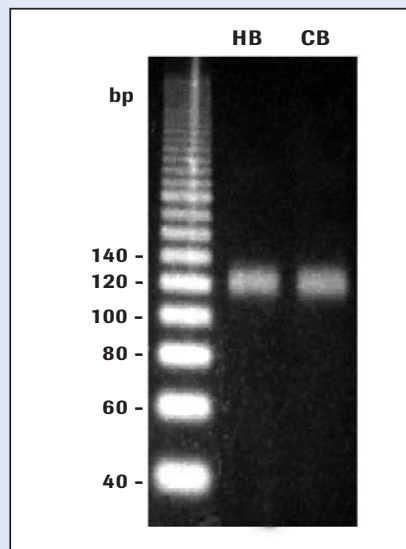


Figure 2: Agarose (4%) gel electrophoresis of PCR-amplified and gel-purified cDNAs (20 ng each) synthesized from human (lane HB) and chimpanzee (lane CB) brain miRNAs.

DNA sequencing

Massively parallel sequencing was performed by 454 Life Sciences Corporation (Branford, CT, USA) using the Genome Sequencer System (one 20 Gb sequence run per library).

Computational analysis of cloned small RNA sequencing reads

Base calling and quality trimming of sequence reads were performed by the Genome Sequencer software. A computational pipeline was developed to further process the sequencing data and to distin-

guish miRNA candidates from other types of small RNAs. In addition, phylogenetic conservation was determined throughout species for which sequenced genomes were available.

Results

More than 87,000 human reads and 140,000 chimpanzee reads were mapped to the respective genome sequences, and annotations for the mapped loci were retrieved from the Ensembl database www.ensembl.org. The distribution of the types of the sequenced elements for the human library is shown in Figure 3. Known miRNAs represent 80 and 60% of the reads in the human and chimpanzee library, respectively. The fraction of reads derived from repeat elements, rRNAs, tRNAs, other noncoding RNA, and repeat-related regions is three times higher in the chimpanzee library (31% versus 11%). Reads that do not correspond to known

miRNAs, other noncoding RNAs, and repeats comprise 9% of the reads in both libraries. Of these, 244 human and 230 chimpanzee regions have the predicted hairpin structure characteristic of miRNAs and are therefore very likely to be novel miRNAs. Many of the novel miRNAs are not conserved beyond primates, indicating their recent origin, and some miRNAs appear species specific, whereas others are expanded in one species through duplication events (6). These data suggest that evolution of miRNAs is an ongoing process and that along with ancient, highly conserved miRNAs, there are a number of emerging miRNAs.

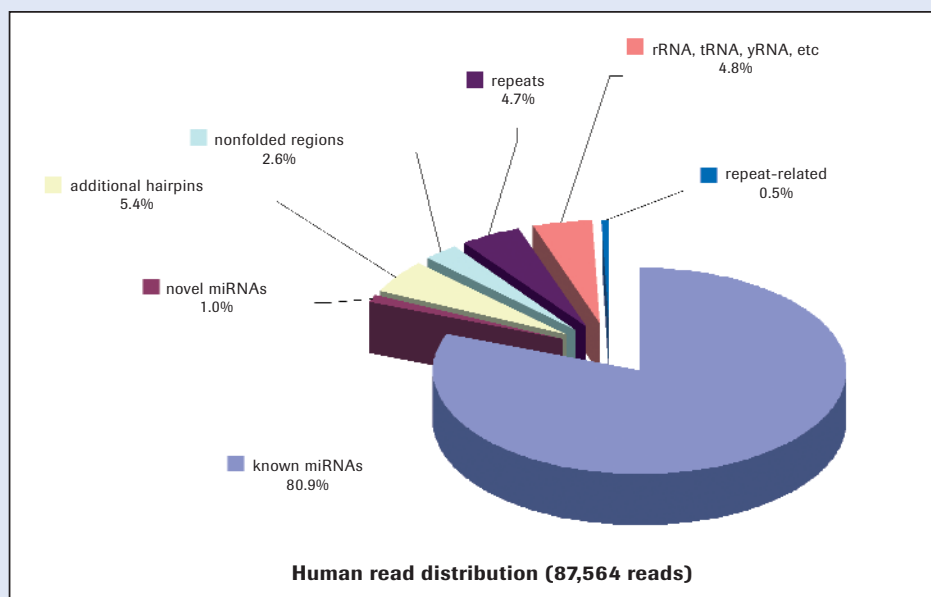


Figure 3: Sequence composition of human fetal brain small RNA libraries.

Conclusion

Novel miRNAs are expressed at low levels, with only a few miRNAs represented by more than one read. Although the novel discovered miRNAs constitute only 1% of the small RNA transcripts in the tissues studied, they more than double the diversity of already known miRNAs. Furthermore, a significant number of novel small RNA molecules were identified that do not belong to the miRNA class and for which the function is currently unknown. Therefore, the Genome Sequencer System appears to be an extremely valuable tool for the exhaustive analysis and rapid experimental discovery of naturally occurring endogenous small RNAs. In addition, deep sequencing results in a quantitative measurement of known small RNAs and can thus be used

for expression profiling of this class of molecules. Although the current study was not set up for a direct expression comparison, we can easily detect expression differences between the human fetal brain sample and the adult chimpanzee brain sample (Figure 4).

In conclusion, deep sequencing of small RNA libraries using the Genome Sequencer System is a powerful approach for the simultaneous expression profiling of known small RNA molecules as well as for the discovery of novel small RNA species. As no prior knowledge of genome or transcriptome sequences is absolutely required, this approach can, in principle, be applied to any species of interest.

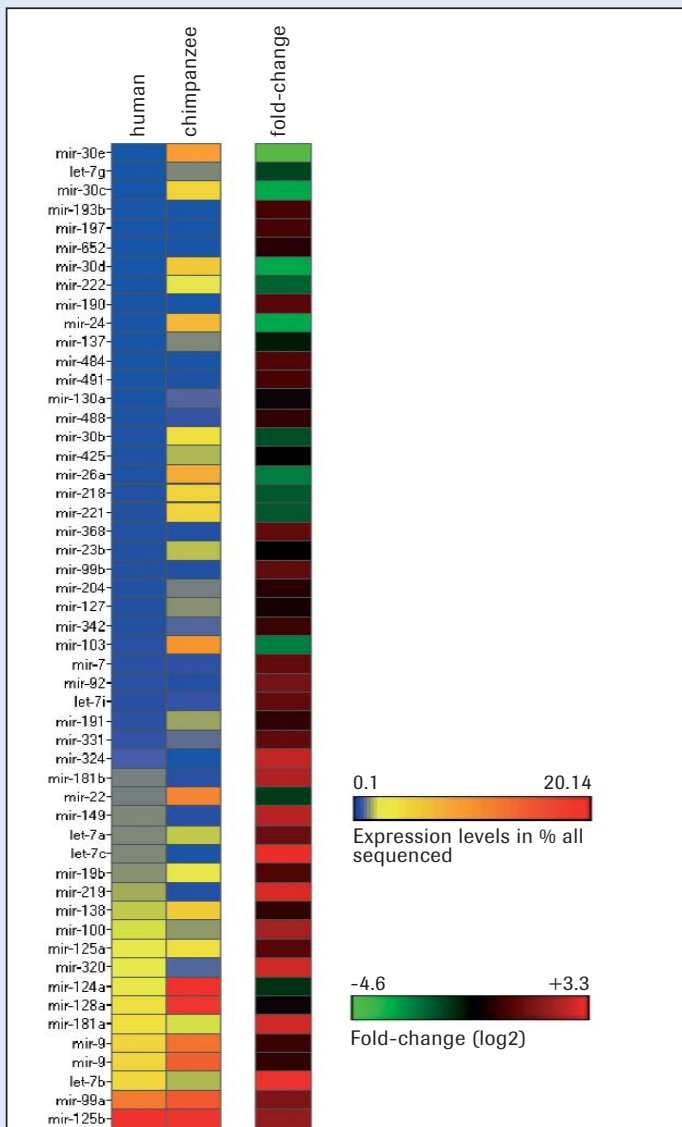


Figure 4: Expression profiling of miRNAs by deep sequencing. Known miRNAs (first column) are sorted by their abundance in the human (second column) and chimpanzee (third column) libraries. A heat map of the expression differences for each miRNA is shown in the fifth column.

References

1. Bartel, D. P. "MicroRNA-Directed Cleavage of HOXB8 mRNA", *Cell* (2004) 116, 281.
2. Alvarez-Garcia, I., Miska, E. A. "MicroRNA functions in animal development and human disease", *Development* (2006), 132, 4653-4662.
3. Griffiths-Jones, S. *et al.*, "miRBase: microRNA sequences, targets and gene nomenclature", *Nucleic Acids Res.* (2006), 34, D140.
4. Berezikov, E. *et al.*, "Phylogenetic Shadowing and Computational Identification of Human microRNA Genes", *Cell* (2005) 120, 1, 21-24.
5. Bentwich, I. *et al.*, "Identification of hundreds of conserved and nonconserved human microRNAs", *Nature Genetics* 2005 Jul;37(7):766-70. Epub 2005 Jun 19.
6. Berezikov, E. *et al.*, "Diversity of MicroRNAs in Human and Chimpanzee Brain", *Nature Genetics* - 38, 1375 - 1377 (2006) Published online: 29 October 2006; doi:10.1038/ng1914.
7. Henderson, I. R. *et al.*, "Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning", *Nature Genetics* 2006 Jun; 38(6):721-5 Epub 2006 May 14.
8. Girard, A. *et al.*, "A germline-specific class of small RNAs binds mammalian Piwi proteins", *Nature* 2006 Jul 13; 442(7099):199-202 Epub 2006 Jun 4.
9. Lau, N. C. *et al.*, "Characterization of the piRNA Complex from Rat Testes", *Science* 2006 Jul 21;313(5785):363-7. Epub 2006 Jun 15.
10. Burnside, J. *et al.*, "Marek's Disease Virus Encodes MicroRNAs That Map to *meq* and the Latency-Associated Transcript", *Journal of Virology*, September 2006, p. 8778-8786, Vol. 80, No. 17.
11. Lu, C. *et al.*, "MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant", *Genome Research*, 2006 October; 16(10): 1276-1288, Published online before print September 5, 2006; DOI: 10.1101/gr.5530106.
12. Qi, Y. *et al.*, "Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation", *Nature*, Published online before print September 24, 2006 doi:10.1038/nature05198.
13. Axtell, M. J. *et al.*, "A Two-Hit Trigger for siRNA Biogenesis in Plants", *Cell* 127, 565-577, November 3, 2006.
14. Berezikov, E. *et al.*, "Diversity of microRNAs in Human and Chimpanzee Brain", *Nature Genetics*, Published Online 29 October 2006; doi:10.1038/ng1914.
15. Pak, J., Fire A. "Distinct Populations of Primary and Secondary Effectors During RNAi in *C. elegans*", *Science*, Published online 23 November 2006.
16. Ruby, J. G. *et al.*, "Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in *C. elegans*", *Cell*, 127, 1193-1207, December 15, 2006.
17. Rajagopalan, R. *et al.*, "A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*", *Genes & Dev.* 20: 3407-3425, December 15, 2006.
18. Johnson, C. *et al.*, "CSRDB: a small RNA integrated database and browser resource for cereals", *Nucleic Acids Research*, Advance Access published December 14, 2006.

NOTICE TO PURCHASER

RESTRICTION ON USE: Purchaser is only authorized to use the Genome Sequencer Instrument with PicoTiterPlate devices supplied by 454 Life Sciences Corporation and in conformity with the procedures contained in the Operator's Manual.

Trademarks

454, 454 LIFE SCIENCES, GENOME SEQUENCER, PICOTITERPLATE and emPCR are trademarks of 454 Life Sciences Corporation, Branford, CT, USA.

Other brands or product names are trademarks of their respective holders.

For more information, visit

www.genome-sequencing.com



Diagnos**t**ics

Roche Diagnostics GmbH
Roche Applied Science
68298 Mannheim
Germany
www.roche-applied-science.com