A novel approach for miRNA profiling using massively parallel, ligation-based sequencing

Catalin Barbacioru, Jian Gu, Scott Kuersten, Melissa Barker, Bob Setterquist, Roland Wicki and Raymond R. Samaha
Applied Biosystems, 850 Lincoln Center Dr., Foster City, CA 94404

ABSTRACT

Using a newly developed library protocol which requires low sample input and results in sequence ready samples in less than a day, we explored the expression profiles of small non-coding RNAs in two normal tissues, using our newly released ligation-based two-base encoding sequencing platform. Fold changes generated from these tissues were compared to those of 210 TaqMan® miRNA assays. Significant correlation levels were observed confirming the applicability of this approach for small RNAs expression profiling. Moreover, more than 3000 potentially novel miRNAs or non-coding RNAs were discovered. These potential novel small RNAs are currently being further validated.

INTRODUCTION

The importance of miRNAs in many cellular as well as developmental processes is now well established. Though presently there are fewer than 800 known miRNAs in the Sanger[1] database, it is likely that many more still remain undiscovered. Unfortunately, hybridization based platforms require a priori knowledge of the miRNA sequences and are therefore unsuited for hypothesis free discovery type studies. In here we describe a new robust method for hypothesis-neutral, whole genome analysis of expression patterns of small non coding RNAs in general and miRNAs in particular. This new approach coupled with the SOLiD™ System provides a high throughput method for digital gene expression that enables the discovery of novel RNAs as well as profiling their expression levels, without the probe bias of microarrays. Because of the SOLiD™ System's throughput which is greater than 100M reads per slide, it is particularly suited for the analysis of gene expression being able to deliver the dynamic range required to detect genes expressed at very low levels and to accurately measure fold changes at the same low expression levels.

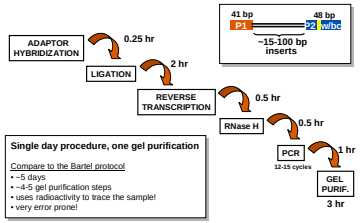


Fig.1: SOLiD™ RNA Expression System Workflow

RESULTS

Two RNA samples (Lung and Placenta) were processed using various conditions to isolate the small RNA fraction (Fig. 2), each of the fifteen resulting samples was processed via the SOLiD™ RNA Expression System and assigned a specific bar code sequence. The resulting libraries were pooled and sequenced using the SOLiD™ System.
The reads from each of the pooled libraries were then separated based on their bar code sequence and mapped against the miRNA sequences available in the Sanger miRNA database. The number of unique reads per miRNA was determined and used as the expression level for that particular miRNA.
The bar codes are designed in such a way that even a 2 base error in the sequence of the bar code still allows the correct assignment.
The even distribution of detected miRNAs in each library is an indicator of the success of the bar coding approach in allowing multiplexing of several samples followed by the ability to segregate the data resulting from each individual library.

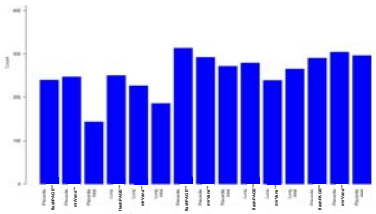


Fig. 2: 15 libraries testing various conditions were separately made, pooled and then sequenced. Consistent numbers of miRNAs were found across libraries.

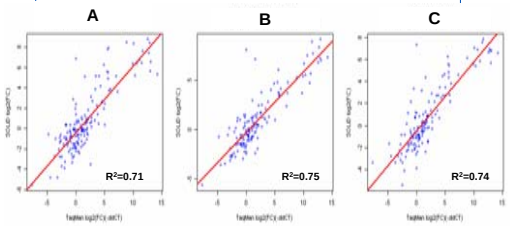
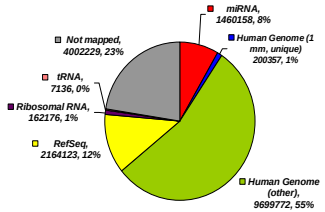


Fig. 3: Fold change correlations with TaqMan® miRNA assays

Furthermore, fold change correlations (placenta vs lung), for three of the libraries (Fig.2; A, B and C), with 210 TaqMan® miRNA assays data were determined. The high level of corelation with TaqMan® assays and the the degree of reproducibility of the correlations between the three libraries is a strong indication of our ability to mix several bar coded samples and extract expression data for each of the pooled samples.



	Reads
Total	17,700,000
miRNA	1,460,158 (8.25%)
tRNA	7,136 (0.5%)
rRNA	162,176 (1%)
RefSeq	2,164,123 (12%)
Genome	9,699,772 (56%)

Fig. 4: Mapping statistics

The reads that did not map to the Sanger database were also mapped sequentially to several other references including tRNAs, rRNAs, RefSeqs and the genome. The mapping statistics are displayed in Fig. 4. The reads that mapped uniquely to the human genome were clustered. 3000 clusters were uniquely mapped, contain at least 2 reads and are of a size<70 bases. These are potentially novel small RNAs/miRNAs and are being further validated

CONCLUSIONS

The SOLiD™ Small RNA Expression Kit provides a streamlined workflow that greatly reduces the time, cost, and experimental variability associated with library preparation. Researchers may now generate small RNA libraries in a single day with a simple, easy to use protocol. The SOLiD™ Small RNA Expression Kit, a LEGenD™ Technology, converts total RNA into a library suitable for emulsion PCR in 5 easy steps with a single purification. This simplified protocol results in a reduction in experimental variability and improves the detection of biologically relevant changes.

REFERENCES

1. miRBase: microRNA sequences, targets and gene nomenclature.
Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ.
NAR, 2006, 34, Database Issue, D140-D144