



Diagnostics

454 LIFE SCIENCES

Small RNA analysis using the Genome Sequencer™ FLX system

The Genome Sequencer™ FLX System from 454 Life Sciences™ and Roche Applied Science is a versatile sequencing platform suitable for a wide range of applications, including *de novo* sequencing and assembly of genomic DNA, transcriptome sequencing, metagenomics analysis and amplicon sequencing. The Genome Sequencer FLX is built upon 454 Sequencing™ technology. One application in which the technology has had a considerable impact is small RNA analysis.

Small RNAs, including microRNAs, endogenous siRNAs and Piwi-interacting RNAs, are a diverse collection of molecules with several important biological functions. Genome Sequencer technology is ideally suited for the analysis of small RNA, and several small-RNA studies that use this technology have been published^{1–8}. With the Genome Sequencer FLX System (Fig. 1), each small RNA molecule within a mixture is sequenced individually. The ability to quickly sequence tens to hundreds of thousands of individual small RNA molecules without cloning into bacteria allows quick identification and quantification of the small RNAs in a sample.

Preparing small RNA for sequencing

Preparation of samples for sequencing begins with the generation of double-stranded cDNA from the small RNA. To facilitate the study of small RNAs, these molecules are typically modified with flanking adaptors and converted to cDNA⁹ using standard protocols. The resulting cDNA, which is suitable for analysis, comprises the small RNAs, typically 20–30 bp in length, flanked by common adaptors of ~20 bp. An online protocol for the preparation of the double-stranded cDNA is available from the David Bartel lab at the Whitehead Institute (http://web.wi.mit.edu/bartel/pub/protocols_reagents.htm) and from the Victor Ambros lab at Dartmouth Medical School (http://banjo.dartmouth.edu/lab/MicroRNAs/Ambros_microRNAcloning.htm).

There are two general methods for preparing these double-stranded cDNAs for sequencing with the Genome Sequencer FLX System. In one method, the double-stranded cDNA is used as starting material for the standard Genome Sequencer FLX DNA library preparation protocol. The Genome Sequencer process-specific A and B adaptors (44-mer adaptors) are blunt-end ligated onto the double-stranded

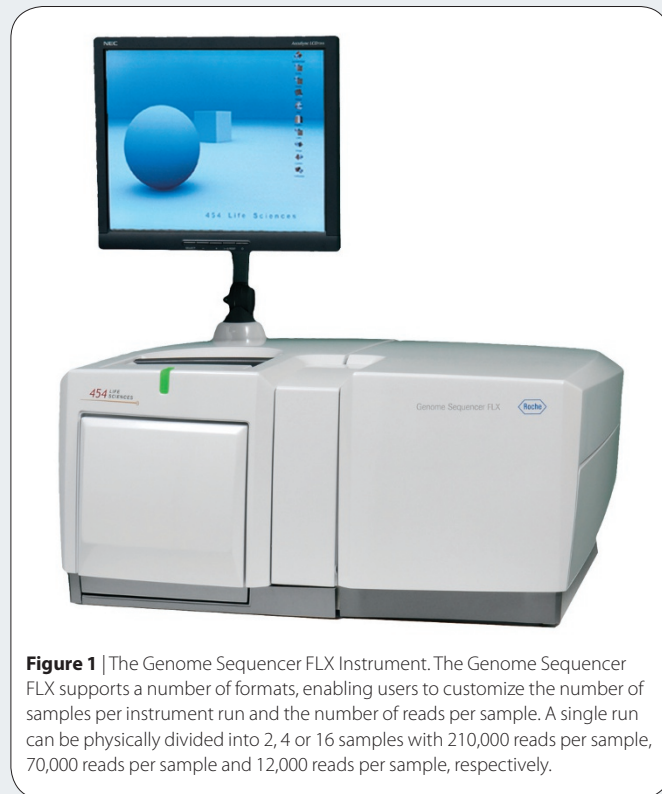


Figure 1 | The Genome Sequencer FLX Instrument. The Genome Sequencer FLX supports a number of formats, enabling users to customize the number of samples per instrument run and the number of reads per sample. A single run can be physically divided into 2, 4 or 16 samples with 210,000 reads per sample, 70,000 reads per sample and 12,000 reads per sample, respectively.

cDNA copy of the small RNA molecules. The resulting double-stranded library is transferred to the emulsion PCR (emPCR™) step for clonal amplification and subsequent sequencing of the individual small RNA molecules (Fig. 2). Sequence data derived from this method have been presented in several publications^{5–8}.

In the second method, Genome Sequencer process-specific A and B adaptors (Fig. 3) are appended onto the double-stranded cDNA copy of the small RNA during a PCR step. In this method, the PCR amplification primers are a fusion primer pair comprising the ~20-bp small RNA adaptor and a 19-bp fixed sequence (Primer A or Primer B) that

Thomas Jarvie¹ & Timothy Harkins²

¹454 Life Sciences, 20 Commercial Street, Branford, Connecticut 06405, USA. ²Roche Diagnostics, Roche Applied Science, 9115 Hague Road, Indianapolis, Indiana 46250, USA. Correspondence should be addressed to T.J. (tjarvie@454.com).

APPLICATION NOTES

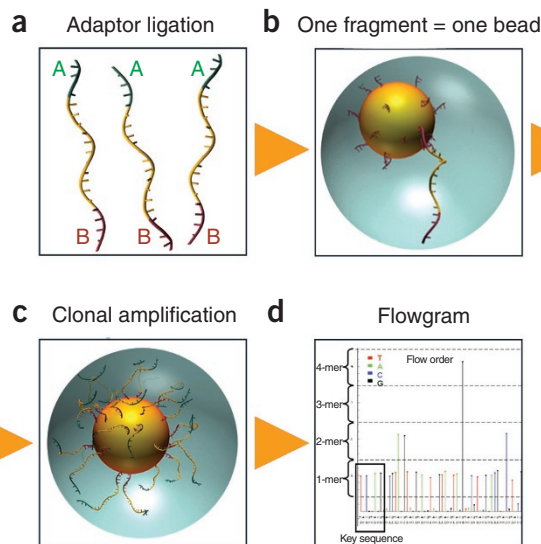


Figure 2 | One fragment = one bead = one read. (a–d) Genome Sequencer process-specific A and B adaptors are appended to each fragment or small RNA (a) to allow binding to DNA capture beads (b). The beads are placed into a water-in-oil emulsion that contains hundreds of thousands of PCR microreactors. Each microreactor contains all reagents necessary for PCR. The entire collection of millions of fragments and beads is amplified in parallel (c), with each fragment amplified without the introduction of competing or contaminating sequences. When sequenced on the Genome Sequencer FLX, each clonally amplified fragment generates its own unique sequence read, represented by a flowgram (d), with over 400,000 reads generated per instrument run.

serve as priming sites during the clonal amplification and the sequencing reaction. The double-stranded library of small RNA molecules is clonally amplified via emPCR and sequenced. Several studies have used this method^{2,4,7}, and detailed protocols are available in these publications.

Preparation of single-stranded tailed templates for sequencing

A variant of the method for appending Genome Sequencer process-specific A and B adaptors onto small RNA by PCR (the second method described above) was developed by Wendy Johnston in the David Bartel lab. This protocol generates single-stranded 19-mer tailed templates that can be used directly in emPCR and sequencing. The protocol, presented below, is a modification of a published method¹⁰. Sequence data derived by this method have been presented in several publications^{3,7}.

Protocol for preparing templates

Prepare PCR templates with one tailed primer and one poly(A)-extended tailed primer having the following sequences: 5' primer, 5'-GCCTCCCTCGCGCCATCAGTATCGTAGGCACCTGAGA-3'; 3' primer, 5'-AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA/iSp18/GCCTTGCCAGCCGCTCAGTATTGATGGTGCCTACAG-3'.

- Prepare a 100- μ l reaction as follows:
 - 0.5 μ l of template

- 0.75 μ l of 100 μ M 5' primer
- 0.75 μ l of 100 μ M 3' primer
- 1 μ l of *Taq* polymerase (5 units/ μ l)
- 10 μ l of 10 \times dNTPs (1 \times : 0.2 mM each dNTP)
- 10 μ l of 10 \times PCR buffer (1 \times : 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin)
- 77 μ l H₂O
- Run 12–16 cycles of PCR (each cycle: 1 min 94 $^{\circ}$ C, 1 min 50 $^{\circ}$ C, 1 min 72 $^{\circ}$ C), according to how quickly the band appears. Use the lowest possible number of cycles.
- Stop PCR with the addition of EDTA to a final concentration of 2 mM.
- Precipitate the DNA: add NaCl to a final concentration of 300 mM, add 2 volumes of ethanol and place reaction mix at –20 $^{\circ}$ C for at least 2 h.

Prepare a 90% formamide, 8% acrylamide gel (1.5 mm thick); use 7.5 inch \times 7.5 inch plates and 20-well comb (each lane is 0.5 cm wide).

- In a beaker, combine 10 ml 10 \times Tris-borate-EDTA buffer (TBE), 400 mg bisacrylamide, 90 ml formamide and 7.6 g acrylamide. Mix using stir plate until dissolved.
- Filter-sterilize, using suction, to remove any particles and oxygen. Store in a light-blocking bottle.
- For a 50-ml gel, use 500 μ l of 10% ammonium persulfate and 150 μ l of TEMED. Mix vigorously. Polymerization is very slow; it is best to allow the gel to polymerize overnight.
- Prepare 1 \times formamide loading dye: 900 μ l of formamide, 100 μ l of 10 \times TBE, 10 μ l of bromophenol blue and xylene cyanol dye solution.

Gel-purify PCR products.

- Centrifuge precipitations for 30 min at full speed.
- Resuspend in 12–15 μ l of 1 \times formamide loading dye.
- For each marker lane, add 10 μ l of 1 \times formamide loading dye to 2 μ l of 1 μ g/ μ l 10-bp ladder in water.
- Pre-run gel at 20 watts for 45 min.
- Heat samples, including ladder, for 10 min at 85 $^{\circ}$ C.
- Load samples on gel, with a ladder next to the sample. Run gel at 20 watts until bromophenol blue runs off the gel.
- Place the gel on plastic wrap in a container. Stain with a solution of 1 \times TBE–ethidium bromide (~20 μ l ethidium bromide (4 mg/ml) per 200 ml of 1 \times TBE). Add only enough so that the solution touches only the gel and plastic wrap, not the dish. Stain for

Primer A sequence	5'-GCCTCCCTCGCGCCATCAG-3'
Primer B sequence	5'-GCCTTGCCAGCCCGCTCAG-3'

Figure 3 | Before emPCR and sequencing, Genome Sequencer process-specific A and B adaptors must be appended to the cDNA copy of the small RNA. Fusion primers of the A and B adaptors and the small-RNA cloning adaptors are used in a round of amplification to append the A and B adaptors to the cDNA copy of the small RNAs.

~ 45 min without shaking (with shaking, ~25 min). (If concerned about cross-contamination, load each sample on a different gel. Otherwise, load multiple samples on the same gel, each with a marker lane, then cut the gel into pieces and stain each piece separately.)

- Visualize staining with UV light. Two bands should be present: one at 130 bp and one at 100 bp. Cut the gel between the two bands (at about 115 bp), approximately 15–20 bp above the 130-bp band, and approximately 15–20 bp below the 100-bp band.
- Elute the two pieces separately in 450 µl of 300 mM NaCl overnight.
- Ethanol-precipitate higher bands as for the DNA template above. Wash pellet in cold 70% ethanol; centrifuge again. Evaporate any remaining formamide (30–60 min) by vacuum concentration. Resuspend pellet in 20 µl of water.

The sample is ready for sequencing.

Summary

Sequencing with the Genome Sequencer FLX System provides a quick, efficient and cost-effective means of sequencing small RNAs. Preparing small RNAs for sequencing is straightforward: the starting point for sample preparation is the standard small RNA-to-cDNA protocol. With multiple sequencing formats available, you can tailor the number of clonal small RNA reads per sample to experimental goals. And, unlike other next-generation sequencing technologies, the read lengths in the system are long enough to allow a complete read-through of the standard cloning adaptors in addition to the small RNA, thereby providing a quality assurance of your sequence results.

Additional information about the Genome Sequencer System is available from Roche Applied Science (<http://www.genome-sequencing.com>). Genome Sequencer, 454, 454 Life Sciences, 454 Sequencing, and emPCR are trademarks of 454 Life Sciences Corporation, Branford, Connecticut, USA. License disclaimer information is available online (<http://www.genome-sequencing.com>).

1. Henderson, I.R. *et al.* Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–725 (2006).
2. Girard, A. *et al.* A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
3. Lau, N.C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367 (2006).
4. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008–1012 (2006).
5. Axtell, M.J. *et al.* A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565–577 (2006).
6. Pak, J. & Fire, A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315**, 241–244 (2007).
7. Ruby, J.G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, **127**, 1193–1207 (2006).
8. Rajagopalan, R. *et al.* A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Dev.* **20**, 3407–3425 (2006).
9. Ambros, V. & Lee, R.C. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol. Biol.* **265**, 131–158 (2004).
10. Williams, K.P. & Bartel, D.P. PCR product with strands of unequal length. *Nucleic Acids Res.* **23**, 4220–4242 (1995).

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.