Digital Gene Expression: Small RNA Discovery and Analysis

Digital Gene Expression: Small RNA Discovery and Analysis is an Illumina Genome Analyzer application that enables the discovery and profiling of all forms of small non-coding RNA. This powerful discovery application, driven by Illumina Sequencing technology allows researchers to investigate small RNA from any organism without prior sequence or secondary structure information.

DGE: Small RNA Discovery and Analysis Highlights

- **Universal Discovery Platform:** No prior sequence or secondary structure information required
- **Customizable Size Selection:** Any small RNA between 17–35 nucleotides can be investigated
- **Wide Dynamic Range:** Up to four million small RNAs can be queried in each flow cell channel
- **Superior Data:** Robust small RNA sequencing and quantification

Figure 1: Flow Cell

Eight samples with up to four million single molecule clusters per channel can be run in parallel on a single flow cell.

Introduction

Recently discovered families of small non-coding RNA have yielded new insights for understanding gene regulation at the transcriptional and post-transcriptional level. Numerous studies suggest that small RNA-mediated silencing may play an important role in development and disease, and imply that small RNA can serve as valuable biomarkers for diagnostic approaches [1–3]. Although thousands of small RNA have been identified over the last ten years, the challenge remains to fully identify all small nuclear RNA and determine their individual functions. Historically, the methods employed for characterizing small RNA have been computational prediction, qPCR, and microarray hybridization. These methods focus primarily on microRNA quantification and are limited to studying small RNA with previous sequence information or secondary stem loop structure. To address these constraints, Illumina has developed Digital Gene Expression (DGE): Small RNA Discovery and Analysis, a robust application that supports researchers interested in novel small RNA discovery and broad species studies4. Driven by Illumina Sequencing technology, DGE: Small RNA Discovery and Analysis is a completely open system that enables small RNA discovery of any size, in any organism, without prior sequence or secondary structure information. With the ability to analyze up to four million small RNAs simultaneously in one sample, DGE: Small RNA Discovery and Analysis offers researchers the broadest and deepest profiling solution for small RNA currently available.

Unlike the relative expression profiles microarray hybridization technology generates, DGE records the numerical frequency of a sequence in the library population, thereby eliminating background signals. Additionally, DGE data can be recorded and annotated using current genome information, and can be easily re-annotated as genome databases evolve.

Small RNA Discovery and Analysis Through Sequencing

Powered by Illumina's proprietary sequencing by synthesis technology, the Illumina Genome Analyzer can be used for many different biological discovery applications such as whole-genome resequencing, DNA methylation status, protein-DNA interactions, whole-genome expression profiling, and small RNA discovery and analysis. Each application differs only in its respective sample preparation protocol and downstream data analysis. For DGE: Small RNA Discovery and Analysis, Illumina scientists have designed a sample preparation protocol that enables sequencing of any small RNA. Through this simple workflow, small RNA libraries are modified to create templates amenable to cluster generation and sequencing by synthesis—the novel pathways at the heart of Illumina Sequencing technology.

The sample preparation protocol offers researchers the flexibility to customize the length of the small RNA they wish to investigate,

enabling size-focused or broad size range investigation of different classes of small RNA on one universal platform. After the small RNAs of choice are isolated based solely on nucleotide length, defined adapters are attached to both ends of the RNA, and cDNA copies are created. The small RNA libraries are loaded onto the fully automated Cluster Station where they bind to complementary adapter oligos grafted onto a proprietary flow cell substrate. The Cluster Station isothermally amplifies these cDNA constructs to create clonal clusters of ~1000 copies each. The resulting high-density array of template clusters on the flow cell is directly sequenced by the fully automated Illumina Genome Analyzer. Using a powerful sequencing by synthesis approach, four proprietary, fluorescently labeled, reversibly terminated nucleotides are used to sequence the millions of clusters base by base in parallel. This combination of unbiased RNA sample preparation and robust sequencing gives researchers unprecedented access to the entire small RNA spectrum.

Ultra-deep Sequencing and Coverage

With an accuracy rate of greater than 99.6% per cycle, the Genome Analyzer sequences and bins data for up to four million clusters per individual flow cell channel. If researchers desire a greater depth of coverage, a single library can be run on multiple channels on the flow cell.

Confident Discovery

With its unparalleled depth, sensitivity, and specificity, Illumina's Small RNA Discovery and Analysis protocol empowers confident identification and characterization of novel small RNA (Figures 2 and 3). DGE: Small RNA Discovery and Analysis is an optimal platform for understanding the role and function of small RNA in cellular processes.

Data Analysis

DGE: Small RNA Discovery and Analysis generates data using open architecture software, allowing researchers to tailor the Illumina Genome Analyzer data analysis software to address their specific needs. Investigators are able to run image analysis, base calling, and standard filtering to generate a list of sequences and counts. It is also possible to annotate small RNA with genomic information and analyze differential gene expression. Sequence data can be easily aligned to publicly available microRNA databases for annotation of known small RNA. Sequences remaining in the discovery fraction can be studied in greater detail through comparisons against other non-coding RNA databases or by alignment to the human genome.

Figure 2: Example of DGE: Small RNA Discovery and Analysis Data

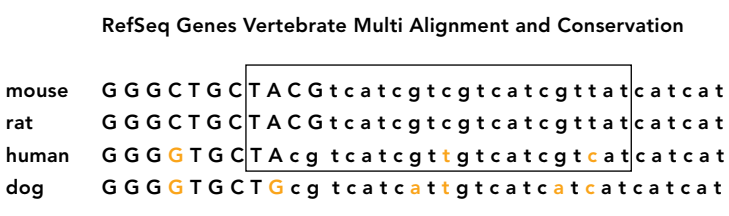
SMALL RNA NAME	CONTENT
hsa-let-7b	76,576
hsa-let-7f	66,100
hsa-let-7a	64,003
hsa-MIR-RG-82	1,931
has-miR-320	1,647
hsa-miR-584	564
hsa-miR-30a-3p	550

Illumina scientists analyzed a small RNA library isolated from human brain total RNA sample. Over 200 small RNA found in the Sanger miRNA database were observed. A portion of the annotated data is shown.

Summary

The Digital Gene Expression Small RNA Discovery and Analysis system can query the entire small non-coding RNA population in any organism, surpassing current discovery methods that can only assay microRNA with previously described sequence or secondary structure information. The Genome Analyzer, in tandem with Illumina Sequencing technology, empowers flexible experimental design with customizable small RNA size selection and offers the deepest detection capabilities for rare small RNA discovery currently available. Like all Illumina RNA Analysis products, Digital Gene Expression Small RNA Discovery and Analysis delivers industry-leading levels of accuracy, efficiency, and affordability.

Figure 3: Detection of Undefined and Unannotated Small RNA Using DGE: Small RNA Discovery and Analysis



Illumina scientists analyzed a small RNA library isolated from human brain total RNA sample. Over 200 small RNA found in the Sanger miRNA database were observed. A portion of the annotated data is shown.

REFERENCES

1. Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, et al. (2007) Requirement of bic/microRNA-155 for normal immune function. Science 316(5824): 530.
2. Campbell TN and Choy FY (2005) RNA interference:past present and future. Curr Issues Mol Biol 7(1):1-6
3. Marsit CJ, Eddy K, Kelsey KT (2006) MicroRNA responses to cellular stress. Cancer Res 66(22):10843-10848.
4. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC et al. (2005) Elucidation of the Small RNA component of the transcriptome. Science 309: 1567-1569

Additional Information

To learn more about Illumina's RNA Analysis Solutions and other Genome Analyzer applications, visit www.illumina.com or contact us at the address below.

Figure 4: Illumina Genome Analyzer

The Genome Analyzer supports a broad range of applications including Small RNA Discovery and Analysis.

Ordering Information

Catalog No.	Product	Description
FC-102-1009	DGE–Small RNA Sample Prep Kit (1)	Contains reagents for preparing eight total RNA samples (eight samples can be loaded on one flow cell).
FC-102-1010	DGE–Small RNA Sample Prep Kit (5)	Contains reagents for preparing 40 total RNA samples (40 samples can be loaded on five flow cells).
FC-103-1008	DGE–Small RNA Cluster Creation Kit (1)	Contains one flow cell, one amplification manifold, and one hybridization manifold for processing up to eight samples.
FC-103-1009	DGE–Small RNA Cluster Creation Kit (10)	Contains ten flow cells, ten amplification manifolds, and ten hybridization manifolds for processing up to 80 samples.
FC-104-1002	26-Cycle Illumina Sequencing Kit (1)	Contains reagents for generating 25 base pair sequences for eight small RNA samples (one flow cell).
FC-104-1003	36-Cycle Illumina Sequencing Kit (1)	Contains reagents for generating 35 base pair sequences for eight small RNA samples (one flow cell).
SY-301-2001	Illumina Cluster Station	Includes the Illumina Cluster Station, computer, software, installation, training, and 1-year warranty.
SY-301-1001	Illumina Genome Analyzer	Includes the Illumina Genome Analyzer, computer, software, installation, training, and 1-year warranty.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, and HiScan are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 770-2007-005 Current as of January 25, 2008

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA