

Serial Analysis of Gene Expression Using ABI PRISM® DNA Sequencers and BigDye™ Terminators

Serial Analysis of Gene Expression (SAGE™) is a sequence-based approach to the identification of differentially expressed genes by comparative analyses. SAGE is a powerful tool, one of the more comprehensive methods available for detailed analysis of large numbers of cellular transcripts, leading to a profile of the expressed genes.¹ SAGE provides an accurate quantitative analysis of the relative levels of genes expressed in a specimen. The ability to count many thousands of genes allows the detection of genes that are expressed at very low levels in a high-throughput manner. A three-step molecular approach, the SAGE process allows simultaneous analyses of sequences derived from various cell populations or tissues.²

Introduction

In 1995, Victor Velculescu first described SAGE, a pioneering approach to gene expression analysis that enabled a detailed characterization of expression patterns without prior knowledge of previously isolated genes. The method was originally developed to investigate genes that might be differentially expressed in colon cancer.⁴

Cancer research remains the focus of the vast majority of SAGE studies. The identification of novel markers for cancer progression and prognosis—as in the case of uterine cancer, in which early stage I detection can be difficult—is critical for successful treatment. The SAGE method can identify potential targets for

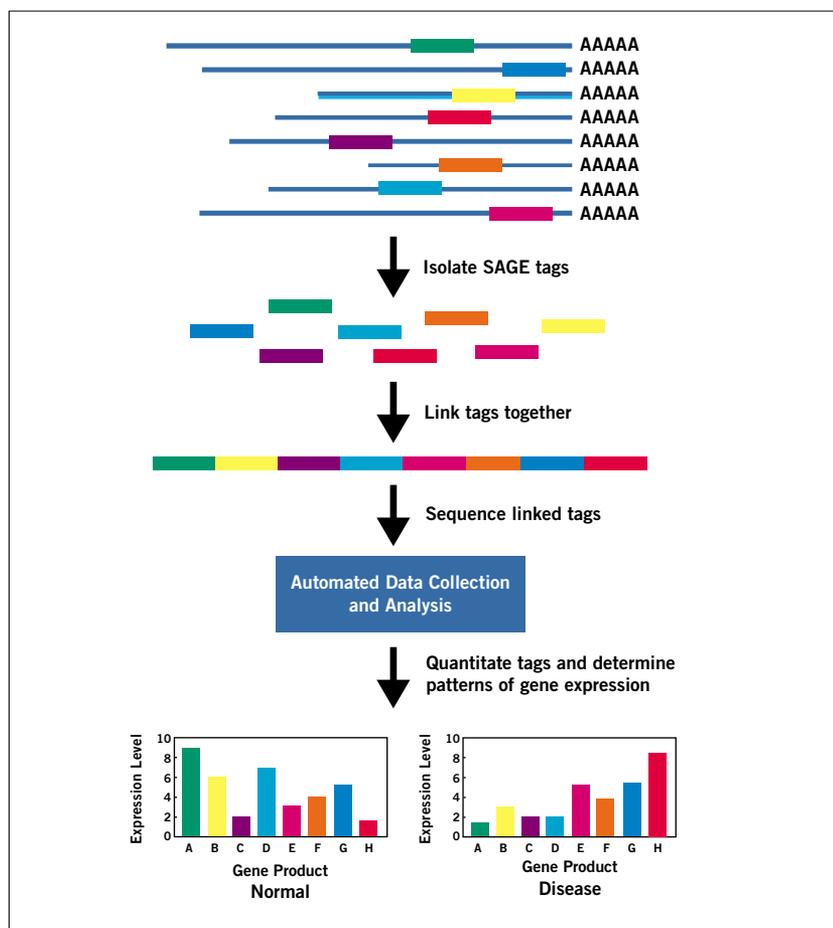


Figure 1. Schematic of SAGE method (Courtesy of sagenet.org)³

therapeutic intervention and find key biochemical pathways. This application is well suited for gene expression analysis of any tissue from any organism.

SAGE is based on the principle that—once isolated from a defined position within a transcript—a sequence tag of 9 or 10 base pairs (bp) provides sufficient information to identify a transcript. These tags can be serially analyzed by ligating or concatenating

them to form longer chains. After they are sequenced, software tools are used to identify and determine abundance. This process involves the following three steps:

1. Biotinylated primers are used to reverse-transcribe mRNA. The resultant cDNA is then digested with a frequent-cutting enzyme (anchoring enzyme AE). Using streptavidin-coated beads, the 3' end of the cDNA is bound and then

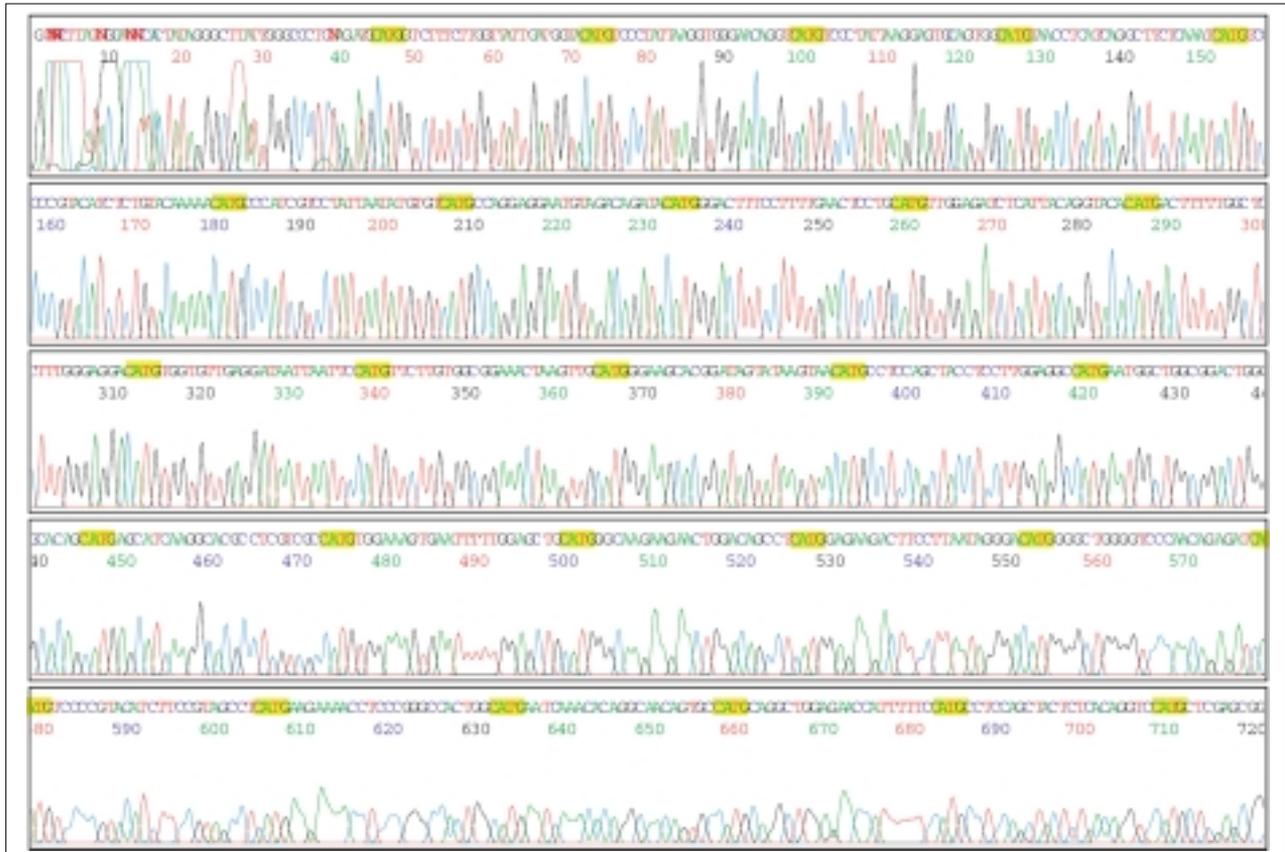


Figure 2. Electropherogram of a typical SAGE sequencing run on the ABI PRISM 3700 DNA Analyzer

Run information: Using POP-5™ polymer on the ABI PRISM 3700 DNA Analyzer, a sequence content of 26 ditags was obtained. Reactions were performed using the ABI PRISM BigDye Terminator v 1.0 chemistry protocol. Cloned inserts were isolated using SeqWright, Inc.'s modified alkaline lysis method. 200 ng of template DNA was sequenced using standard cycling conditions and full volume reactions. Samples were precipitated using isopropanol and resuspended in 20 mL of Hi-Di™ formamide, denatured, and then loaded on the instrument. See the instrument's DNA Sequencing Chemistry Guide for recommended protocols. (DNA was provided by Dr. Jill Powell, The Richard Dimbleby ICRF Department of Cancer Research, United Medical and Dental Schools, London.)

split into two pools, A and B. A primer (A' and B') containing a recognition site for a type IIS restriction enzyme is ligated to the nucleotides in each of these pools.

2. Following digestion with a relevant type IIS tagging enzyme, eluted DNA fragments are ligated and amplified using the A' and B' primers. Following PCR, the anchoring enzyme removes the priming sites A and B. Concatamers are then formed and cloned into a vector. A "ditag" containing the sequence information of two independent cDNA tags is punctuated by the anchoring site.*

3. Cloned inserts are verified by agarose gel and sequenced. SAGE analysis tools then analyze these sequences. Finally, a comparison of the sequence data determines variations in expression patterns.

Applied Biosystems Capillary Electrophoresis Systems and SAGE Technology

Most SAGE investigations require analysis of more than 50,000 tags. Obtaining sequencing information quickly and accurately is vital. Applied Biosystems capillary electrophoresis systems and BigDye™ terminator chemistries offer an easy, hands-free approach to sequencing. ABI PRISM® 3700 and 3100 DNA sequencers accurately and reliably

determine the transcript information necessary to build a profile. ABI PRISM BigDye terminators provide signal uniformity and sensitivity. Together they offer a cost-effective solution to large-scale transcript analysis using SAGE technology.

Signal Detection

Accurate basecalling at long readlengths is important for ditag analysis. BigDye terminators provide the high degree of accuracy that transcript analysis of short 9- or 10-bp sequences demands. This accuracy, coupled with the sensitive detection capacity of the ABI PRISM 3100 or 3700 DNA sequencers, ensures a fast and accurate read. BigDye™ Terminator v 3.0 enables even greater

Instrument	Applications	Capillary Length	Run Time	Read Length	24-Hr Throughput
3100 Analyzer	Rapid sequencing	36 cm	60 min	>500 bases (98.5% accuracy)	24 runs 7680 tags
3100 Analyzer	Standard sequencing	50 cm	2.5 hr	>650 bases (98.5% accuracy)	9 runs 3744 tags
3100 Analyzer	Long read sequencing	80 cm	3.5 hr	>1020 bases (98.5% accuracy)	7 runs 4480 tags
3700 Analyzer	Standard sequencing	50 cm	3 hr	>700 bases (98.5% accuracy)	8 runs 21,504 tags
377 Sequencer	Standard Sequencing	96 lanes/48 cm gel	10 hr	>800 bases	2 runs 4000 tags

Table 1. Sequencing Throughput for the 3100, 3700, and 377 Instruments

signal uniformity for samples that show a tendency toward signal loss in longer fragments.

Complete Automation

Both the 3100 and 3700 DNA sequencers are designed for more than 24 hours of unattended operation. Human intervention is required only to place the 96- or 384-well sample trays on the instrument, and to replenish the polymer supply. All other steps, including capillary filling, sample loading, capillary rinsing, data extraction, and analysis are completely automated.

SAGE Sequence Protocol

In experiments using typical sequencing methods, along with the ABI PRISM® 3100 Genetic Analyzer and 3700 DNA Analyzer for sequencing, Applied Biosystems scientists produced reproducible and accurate ditag sequences.

Samples were prepared from mRNA derived from a cell line, MCF-7 (human breast adenocarcinoma). A sequence content of approximately 26 ditags was obtained using both the standard protocol with POP-5™ polymer on the 3700 DNA Analyzer, and the Rapid Read protocol on the 3100 Genetic Analyzer. For optimal sequence efficiency, the insert length should be sufficient to fully utilize the length-of-read for the instrument platform and BigDye chemistry.

Throughput

Throughput is an important consideration when choosing a platform for sequence analysis. The elimination of gel tracking and other problems that can arise from slab-gel sequencing greatly improves the sequencing capacity of the average research laboratory. Throughput requirements depend on the quantity of tags to be analyzed. ABI PRISM DNA sequencers provide this flexibility by offering medium-to-high throughput.

Capillary lengths of 36 cm, 50 cm, and 80 cm on the 3100 Genetic Analyzer afford flexible run times and read lengths. This system is designed with 16 capillaries, and run times can be determined, depending on insert size. Typically, a rapid run will read tag sequences to 550 bp with 98.5% accuracy. We recommend the 80-cm array for long-read runs for inserts of 1 kb or larger. (See Table 1 for throughput.)

Using standard run conditions, the 50-cm array on the ABI PRISM® 3700 DNA analyzer will provide long reads with POP-5 polymer over a 3-hour run time. Three 24-hour running periods, therefore, will yield almost all the sequence information required to profile one tissue type (See Table 1).

Analysis

When sequencing and basecalling are complete, the investigator needs to assess the presence and frequency of ditags. Various software tools are available to analyze the ditag sequence and determine transcript abundance. When all sequence files have been processed, it is possible to view tag abundances, match tags to reference sequences, and compare tag abundances from multiple projects. Analysis generally involves matching the tag information against the GenBank-UniGene⁵ database and generating a reference list. Project tags are compared with this list to identify matches of known genes and other sequences.

The National Institutes of Health (NIH) has established SAGEmap, a public repository and resource for SAGE data.⁶ Originally developed by researchers at the National Center for Biotechnology Information (NCBI) in Bethesda, MD, SAGEmap was designed to archive SAGE data produced through the Cancer Genome Anatomy Project (CGAP). However, it also accepts submissions of SAGE sequence from any source. The National Cancer Institute (NCI) has chosen to use SAGE exclusively for The Human Tumor Gene Index (hTGI) initiative, which is part of CGAP.

Conclusion

SAGE is a well-recognized method of gene-expression profiling. The information gathered from SAGE analysis enables a clear view of the genes associated with normal, developmental, and disease states. Researchers require an efficient means of accessing the sequence information contained in the concatamer of transcript tags. ABI PRISM, 3100 and 3700 DNA sequencers, coupled with BigDye terminator chemistry, provide a versatile approach to generating ditag sequences, and make the SAGE application readily accessible to disease investigators.

References

1. Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science*. 270:484–487.
2. Kozian, D.H., and Kirschbaum, B.J. 1999. Comparative gene-expression analysis. *Trends in Biotechnology*. 17:73–78.
3. Johns Hopkins University SAGE Web site: <http://www.sagenet.org>. Schematic of SAGE method. April 2001.
4. Dr. K. Kinzler, Dr. B. Vogelstein, Johns Hopkins University, Baltimore, MD.
5. National Institutes of Health Genetic Sequence Database: <http://www.ncbi.nlm.nih.gov/Genbank/> and <http://www.ncbi.nlm.nih.gov/UniGene/index.html>. April 2001.
6. Lash, A.E., Riggins, G.J. et al. 2000. SAGEmap: A public gene expression resource. *Genome Research*. 10:1050–1060.

*Invitrogen Corporation has released I-SAGE™, the first pre-assembled kit for performing SAGE™ technology. This all-inclusive kit provides sufficient pre-tested materials to generate up to six SAGE libraries for subsequent sequencing.

Worldwide Sales Offices

Applied Biosystems vast distribution and service network, composed of highly trained support and applications personnel, reaches into 150 countries on six continents. For international office locations, please call the division headquarters or refer to our Web site at www.appliedbiosystems.com.

Applera Corporation is committed to providing the world's leading technology and information for life scientists. Applera Corporation consists of the Applied Biosystems and Celera Genomics businesses.

Headquarters

850 Lincoln Centre Drive
Foster City, CA 94404 USA
Phone: 650.638.5800
Toll Free: 800.345.5224
Fax: 650.638.5884

For Research Use Only.
Not for use in diagnostic procedures.

AB (Design), Applera, BigDye, Hi-Di, POP, and POP-5 are trademarks, and ABI PRISM and Applied Biosystems are registered trademarks of Applera Corporation or its subsidiaries in the US and certain other countries. AmpliTaq is a registered trademark of Roche Molecular Systems, Inc. I-SAGE is a trademark of Invitrogen Corporation. SAGE is a trademark of Genzyme Molecular Oncology. All other trademarks are properties of their respective owners.

©2001 Applied Biosystems. All rights reserved.
Information subject to change without notice.

Printed in the USA, 7/2001, JPI
Publication 106AP05-01