TRANSCRIPTOME PROFILING TECHNOLOGIES
at the British Columbia Cancer Agency

Khattra J, Chan S, Asano J, Pandoh P, Coughlin S, McDonald H, Girn N, Vatcher G, Schnerch A, Freeman D, Zuyderduyn S, Leung D, Teague K, Jones S, Marra M

British Columbia Cancer Agency
Vancouver, British Columbia, Canada

Canada's Michael Smith
Genome Sciences Centre
www.bcgsc.ca

Objective and Design

The Genome Sciences Centre has established a Gene Expression Laboratory with the purpose of consolidating high throughput transcriptome analysis platforms and related technology development.

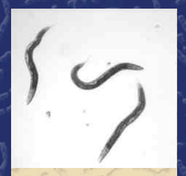
Global gene expression profiling is being applied to research in cancer genomics (V. Ling, C. Eaves), atlas of development in mouse (P. Hoodless, E. Simpson, C. Helgason), biology of *C. elegans* (D. Baillie, D. Moerman, D. Riddle, J. McGhee), a variety of mouse and human embryonic stem cells (K. Humphries, C. Eaves, J. Thomson), transient hypoxia of human tumor cells (R. Durand), and host response to pathogens (R. Brunham, C. Astell).

Our choice and design of experimental platforms addresses the following issues critical to successful gene expression profiling in the wide variety of projects listed above:

1- Isolation and analysis of top quality RNA, including efficient isolation from minute samples.

2- Quantitative global gene expression profiling of both known and novel gene transcripts using Affymetrix *GeneChips* and Serial Analysis of Gene Expression (SAGE).

3- Accurate validation of selected transcripts via quantitative real-time RT-PCR (ABI *Prism* 7900HT Sequence Detection System).



The nematode *C. elegans* glp-4(bn2ts) gonadless mutant used for manual microdissection of guts for SAGE analysis. An adult worm is about 1 mm long (magnified 30x). Photograph courtesy of Dr. James D. McGhee, University of Calgary.

Methods and Platforms

RNA isolation and analysis:

Harvesting and transport of tissue at ambient temperature is enabled by the use of RNA*later* solution (Ambion).

A PowerGen 125 rotor-stator homogenizer (Fisher Scientific) with 5 mm generator is utilized for shearing samples with volumes as low as 30 microlitres. The use of disposable generators ensures no sample cross-contamination.

Total RNA is usually isolated with TRIzol (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate, and phase lock gels (Eppendorf). The latter enables clean and quantitative recovery of the RNA-containing upper aqueous phase.

The removal of any contaminating DNA is performed with Ambion's DNA-*free* reagent, a novel method that does not require subsequent organic extraction, alcohol precipitation, heating, or the addition of EDTA to the DNase-treated RNA sample.

All experiments begin with a rigorous analysis of RNA quality and quantity. A rapid assessment of RNA quality is performed with as little as 200 picograms of total RNA in a 1 microlitre sample volume. This is achieved using the Agilent 2100 Bioanalyzer (Agilent Technologies) and RNA 6000 Pico *LabChip* kit.

RNA LabChip from Agilent Technologies. Miniaturized fluidics and electrokinetics are combined to allow rapid sizing and quantitation of nucleic acids at the picogram level. RNA 6000 Pico LabChips can process eleven samples and the run time is about one minute per sample. Controlling software outputs RNA concentration and ratio of ribosomal RNA bands.

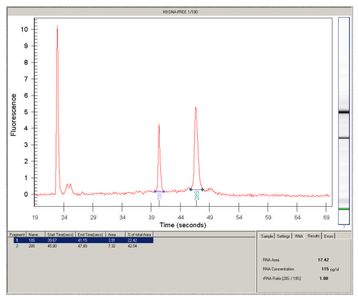


Figure 1. Agilent Bioanalyzer electropheregram for total RNA from a human embryonic stem cell line. Note the two distinct small and large subunit ribosomal RNA peaks at about 40 and 47 seconds, and small RNA species around 25 seconds. The RNA concentration is a mere 115 pg/uL and 28S/18S rRNA ratio is 1.9.

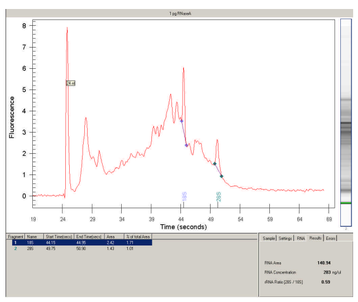


Figure 2. Agilent Bioanalyzer electropheregram of degraded total RNA from some wild-type *C. elegans* embryos. Note the absence of two distinct small and large subunit ribosomal RNA peaks. The rRNA peaks ratio is a low 0.6.

The sharp peak at about 25 seconds is a spiked reference marker.

Serial Analysis of Gene Expression:

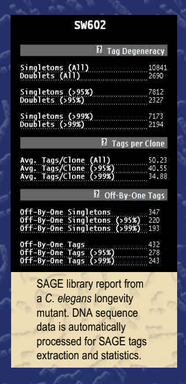
SAGE is unique in its ability to generate quantitative digital datasets of known and novel transcripts. Making a regular library requires 5 micrograms of good total RNA, and takes eleven working days to construct and validate quality of clones. This is followed by DNA sequencing and bioinformatics analysis. We have established a still-growing production group to achieve an estimated throughput of 120 SAGE libraries per year, for the next three years.

We are applying modified SAGE protocols and have constructed our first LongSAGE library of 21 bp cDNA tags from a human embryonic stem cell line. This library is enabling comparison with a 14 bp tag library from the same RNA sample.

Work is underway on constructing libraries from sub-microgram amounts of total RNA by amplification (SAGE-Lite and PCR-SAGE).

We are also poised to make gains on the number of SAGE tags obtained per sequence reaction via our DNA sequencing production group's newly acquired Applied Biosystems 3730*xl* DNA Analyzers, capable of generating 1000 bp read lengths.

Affymetrix *GeneChip* snapshot depicting a processed raw image. Each square cell contains the same 25mer probe and is 18 microns across. About half a million such cells are present on each genome-wide expression microarray such as the human U133A and B set. The chequered edge is a spiked control oligo which allows monitoring quality and providing landmarks for a grid overlay.



SAGE library report from a *C. elegans* longevity mutant. DNA sequence data is automatically processed for SAGE tags extraction and statistics.

Platforms (continued)

Affymetrix GeneChip technology:

Figure 3. Hardware for the Affymetrix platform. The fluidics station (on the left), for washing and staining, can process up to 4 *GeneChips* simultaneously in about 90 minutes. The epifluorescence scanner (middle) has a resolution of 3 microns and scans each chip twice over 10 minutes. Both machines are software-controlled.

High density oligonucleotide *GeneChips* are manufactured by Affymetrix using photolithography and solid-phase combinatorial chemistry. These include various species-specific genome-wide expression microarrays that simultaneously probe for thousands of transcripts. Regular eukaryotic expression profiles are generated with as little as 5 micrograms of good total RNA and takes about 4 working days. RNA amplification (Eberwine aRNA method) can be performed on smaller eukaryotic samples (10-100 nanograms total RNA). Such an amplification has been applied to a human embryonic stem cell line sample for comparison with it's *GeneChip* expression data from 15 micrograms of the same starting total RNA.

We have constructed a web-based interface for submission of sample information and data retrieval.

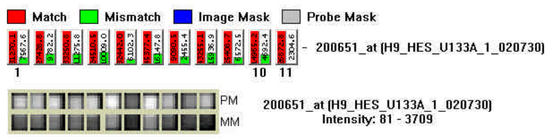


Figure 4. Probeset intensity and image data for GNB2L1 mRNA detected on a human genome-wide U133A expression *GeneChip* from an embryonic stem cell line sample. Non-specific signal from each mismatch probe (shown in green) is subtracted from the corresponding perfect match probe (in red) to generate a true signal value. The data from 11 such probe pairs is averaged to give a signal value and quality call for each transcript.

Results

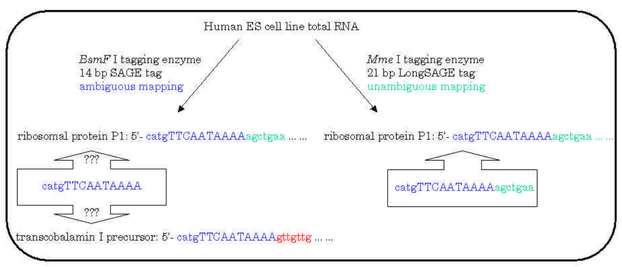
Our resources for RNA work have enabled successful RNA isolation and analysis from samples as minute as a few *C. elegans* guts, a nematode 1 mm in length. Fine instrumentation and fluorescent dye combinations allow for minimal sacrifice of precious samples harvested by tedious manual or laser-capture microdissection and fluorescence-assisted cell sorting.

SAGE:

We have constructed our first comprehensive set of SAGE libraries, which investigate the genetic specification of life span in *C. elegans*. A second set compares expression profiles of human and mouse ES cells. Our third set of libraries are from human tumor cells subject to differing levels of transient hypoxia in vitro. Also constructed is the first of about 200 SAGE libraries for an atlas of gene expression in mouse development.

LongSAGE:

Preliminary analysis of 21 vs. 14 bp cDNA tags derived from the same human ES cell line RNA sample has demonstrated the utility of 21 bp tags in reducing the ambiguity of tag-to-gene mappings.



DNA microarrays:

Genome-wide *GeneChip* expression arrays have been applied to the same human and mouse ES cells' RNA utilized for SAGE library construction. These datasets are providing a comparison of the two platforms.

Preliminary work in the study of host response to pathogens involved *GeneChip* expression profiling of human lymphocytes challenged with inactivated influenza virus. We identified distinct transcriptome signatures of two lymphocyte groups characterized as strongly or weakly responsive to the viral challenge via a measure of cytokines secretion. Furthermore, we observed the characteristic up-regulation of metallothioneins, which was shown previously to occur in influenza-challenged cells in culture (Geiss, G. K. *et al*, J. Virology, vol. 75, 2001). This induction is believed to be the host cells' response to virus-driven oxidative stress. Interestingly, the clinical administration of zinc lozenges to combat the onset of cold-like symptoms has its basis in the metal's ability to induce metallothionein gene expression (Marshall, S., Can. Fam. Phys., vol. 44, 1998).

Reproducibility of expression data by this technology has been achieved with technical replicates for human ES cells' RNA differing by no more than 0.15-0.46 % (59-179 transcripts) for a two-chip assay of 39,000 transcripts.

Results (continued)

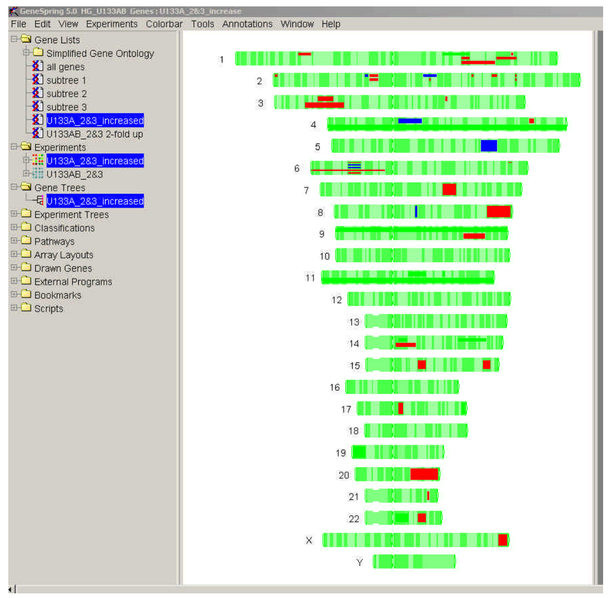


Figure 5. *GeneSpring* 5.0 (Silicon Genetics) chromosomal view showing the location of strongly induced transcripts (in red and blue) from a human lymphocytes sample challenged with inactivated influenza virus. Data is from Affymetrix genome-wide expression *GeneChips* (U133A and U133B).

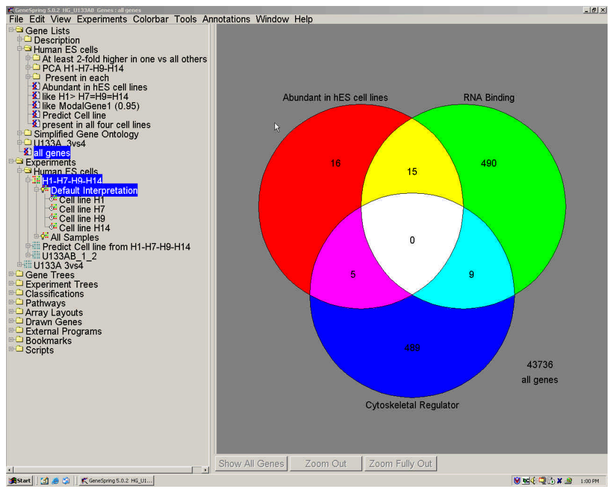


Figure 6. *GeneSpring* 5.0 (Silicon Genetics) Venn diagram showing a distribution of 20 transcripts found to be highly abundant in all four human embryonic stem cell lines profiled with Affymetrix expression *GeneChips*. The RNA Binding and Cytoskeletal Regulator gene lists are derived from a simplified Gene Ontology database structure.

Conclusion

Combining the latest in laboratory instrumentation and select molecular biology techniques has positioned us to routinely perform transcriptome-wide gene expression analyses and is preparing us for tackling minute samples using the same high throughput platforms.

Acknowledgements

We gratefully acknowledge:
- Funding from Genome Canada, Genome BC, National Cancer Institute of Canada, Canadian Institutes of Health Research, US Department of Health and Human Services, and the US National Cancer Institute.
- Laboratory administration support from Letticia Hsiao, Carrie Jang, Steve Messervier, Jill Vardy, Lesa Fehr, and Lynn Hume.
- Projects administration support from Dr. Caroline Astell.
- Computer systems support from Kirk Schoeffel, Bernard Li, and Mark Mayo.