

AAAS.ORG |

Science Signaling

Enter Search Term

ADVANCED

UNIVERSITY COLLEGE LONDON ALERTS | ACCESS RIGHTS | MY ACCOUNT | SIGN IN

[SAGE KE Home](#) [Issue Archive](#) [Literature & News](#) [Community](#) [Resources](#) [Highlights](#) [About SAGE KE](#)

[Home](#) > [SAGE KE Home](#) > [Literature & News](#) > [Viewpoint](#) > Bengtsson et al., p. vp8

*Sci. Aging Knowl. Environ.*, 19 December 2001

Vol. **2001**, Issue 12, p. vp8

[DOI: 10.1126/sageke.2001.12.vp8]

[To Advertise](#) [Find Products](#)

## VIEWPOINT

### Identifying Differentially Expressed Genes in cDNA Microarray Experiments Authors

Henrik Bengtsson, Brent Calder, I. Saira Mian, Matt Callow, Eddy Rubin, and Terry P. Speed

H. Bengtsson is in the Department of Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Sweden. B. Calder is at The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. I. S. Mian is in the Department of Radiation Biology and Environmental Toxicology, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. M. Callow and E. Rubin are in the Genome Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. T. P. Speed is in the Department of Statistics, University of California, Berkeley, CA, USA, and the Division of Genetics and Bioinformatics, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia. E-mail: [SMian@lbl.gov](mailto:SMian@lbl.gov)

<http://sageke.sciencemag.org/cgi/content/full/sageke;2001/12/vp8>

**Key Words:** cDNA microarrays • transcription profiling • expression profiling • gene chips

Transcription profiling monitors simultaneously the abundance of tens of thousands of transcripts in a biological sample. In biogerontology, such profiling is but one arrow in the quiver required to ascertain how the physiological vigor of an organism declines over time. Currently, expression profiling technology comes in a variety of flavors, including cDNA microarrays, high-density nylon membrane arrays, serial analysis of gene expression, short or long oligonucleotide arrays, fiber optic arrays, and so on (see [Site 1](#)). The exponential rise in the number of profiling-related Medline articles is a testament to its popularity. If this experimental tool is to deliver on its promises, however, a growing list of statistical, computational, technological, and biological issues will need to be addressed. (For incomplete lists, see [Site 2](#) and [Site 3](#).)

This discussion forum focuses on the elemental but critical issue of statistical approaches to identifying differentially expressed genes, given cDNA microarray data (it should be noted that many of the same issues arise with the other technologies). The nature and type of data to be analyzed, as well as a need to distinguish between biologically and statistically significant differential expression, highlight the multiplicity of potential answers to such a deceptively simple question.

The starting premise for the forum is that a suitable scientific question has been formulated, the appropriate study designed, standard operating procedures employed to extract mRNA samples from relevant specimens, the transcription profiling experiments performed, and images of hybridized microarrays acquired. Accounting for technological and/or experimental confounding factors can reduce observation noise (errors associated with the measurement process or technology itself) but has little influence on model noise (variation due to the stochastic nature of gene expression and the underlying biology). It is assumed that strategies for minimizing observation noise and data analysis are independent of the scientific question. That is, aging is sufficiently similar to cancer, development, and other biological processes that no "new" statistical methods need to be invented to analyze transcription profiling data from biogerontology studies. Thus, employing a study of mouse lipid metabolism as the framework within which to structure the discussion forum and to illustrate pertinent issues seems appropriate.

The goal of the lipid study was to identify genes with altered expression in two mouse models with low HDL cholesterol levels compared to inbred control mice. cDNA microarrays containing 5000 mouse expressed sequence tags (ESTs) were used to screen specimens from the livers of apolipoprotein AI-knockout mice (ApoAI treatment group), scavenger receptor BI transgenic mice (SRBI), and control mice (CT). Eight microarray experiments compared ApoAI samples with CT samples, and eight more compared SRBI with CT. Details of the starting data and subsequent statistical analyses performed can be seen [here](#).

To capture the salient features of the aforementioned or another study, the following terminology is proposed and employed. Consider a cDNA microarray slide arrayed with  $N$  genes (here,  $N_{\text{genes}} = 5000$ ). Let  $A, B, \dots$  denote samples of interest (ApoAI and SRBI) and  $R$  a reference sample that may or may not be biologically meaningful (CT). A directed edge between two samples represents a specific labeling scheme with one dye assigned to the arrowhead and the second to the tail. Thus,  $A \rightarrow B$  represents a slide containing sample  $A$  labeled with the tail dye and  $B$  labeled with the head dye (CT  $\rightarrow$  ApoAI, CT  $\rightarrow$  SRBI). By extension,  $A \leftrightarrow B$  represents a dye-swap study; that is, profiling experiments  $A \rightarrow B$  and  $B \rightarrow A$ .

There are many types of experimental replication, all of which are designed to improve the precision of the measured expression values. "Spot repeated measure" refers to the same gene arrayed  $N_{\text{spot}}$  times on a single slide ( $N_{\text{spot}} = 1$ ). "Sample repeated measure" refers to an individual mRNA sample hybridized to  $N_{\text{sample}}$  slides ( $N_{\text{sample}} = 8$ ). "Specimen replication" refers to  $N_{\text{specimen}}$  specimens ( $N_{\text{specimen}} = 16$ ). Using the syntax  $[A(N_{\text{genes}}, N_{\text{spot}}) \rightarrow B]N_{\text{sample}}$ , the mouse lipid study can be summarized as  $[CT(5000, 1) \rightarrow ApoAI]_8$  and  $[CT(5000, 1) \rightarrow SRBI]_8$ .

Identifying differentially expressed genes corresponds to distinct statistical questions, each of which yields its own answer(s). For both within and between treatment groups, the number of experiments analyzed affects the set of genes designated as having altered expression,  $[CT(5000, 1) \rightarrow ApoAI]_{1-8}$  or  $[CT(5000, 1) \rightarrow SRBI]_8$  and  $[CT(5000, 1) \rightarrow ApoAI]_8$  versus  $[CT(5000, 1) \rightarrow SRBI]_8$ . To demonstrate this, results from the application of a specific set of statistical techniques are presented here. The approaches emphasize detecting genes with large changes in expression and overlook biologically important genes that may have small but reproducible changes in expression. The nature of the slides precludes investigation of the role of "spot level replication," whereas the lack of dye-swap experiments prevents evaluation of gene-specific dye effects.

What role does transcription profiling play in providing fundamental and deep insights into the basic biology (how) and evolution (why) of

aging? Paul Klee once wrote, "Art does not reproduce the visible; it makes things visible." Transcription profiling attempts to reproduce alterations in mRNA levels and can generate collections of differentially expressed genes. Making aging visible will require abstracting knowledge from this and multiple other sources of information.

---

December 19, 2001 **Citation:** H. Bengtsson, B. Calder, I. S. Mian, M. Callow, E. Rubin, T. P. Speed, Identifying Differentially Expressed Genes in cDNA Microarray Experiments Authors. *Science's SAGE KE* (19 December 2001), <http://sageke.sciencemag.org/cgi/content/full/sageke;2001/12/vp8>

---

THIS ARTICLE HAS BEEN CITED BY OTHER ARTICLES:

**Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions.**

A. E. Teschendorff and M. Widschwendter (2012)

*Bioinformatics* **28**, 1487-1494

[Abstract »](#) [Full Text »](#) [PDF »](#)

Science of Aging Knowledge Environment. ISSN 1539-6150

[News](#) | [Science Journals](#) | [Careers](#) | [Multimedia](#) | [Collections](#) | [Help](#) | [Site Map](#) | [RSS](#)

[Subscribe](#) | [Feedback](#) | [Privacy / Legal](#) | [About Us](#) | [Advertise With Us](#) | [Contact Us](#)

© 2001 American Association for the Advancement of Science. All Rights Reserved.

AAAS is a partner of [HINARI](#), [AGORA](#), [OARE](#), [PatientInform](#), [CrossRef](#), and [COUNTER](#).