directly linked based on large-scale similarity searches retrieved from HomoloGene (National Center for Biotechnology Information) and GeneDB (Sanger Centre)[4,5]. GermOnline can be searched by author name, gene name, keyword, expression profile and phenotype[6] (**Fig. 1a**). The locus report pages provide curated information (**Fig. 1b**) and links to relevant external data sources. Likewise, major databases, such as Swiss-Prot, Saccharomyces Genome Database and GeneDB, provide links to our database. GermOnline is especially useful for genome biologists, who often need to process and interpret data on large numbers of genes for which extensive literature is available.

GermOnline Release 2.0 contains information on 684 genes from *Saccharomyces cerevisiae* involved in meiosis (**Fig. 1c**), spore formation and germination, as well as about 30 prototype contributions from other species. Microarray expression data covering the yeast cell cycle[7], yeast sporulation[8,9] (**Fig. 1d**) and spermatogenesis in the rat (U.S. *et al.*, unpublished data) are provided, as well as external links to relevant studies using *S. cerevisiae*, *Schizosaccharomyces pombe* and *Caenorhabditis elegans*. Release 2.0 is accessible at http://www.germonline.org. Detailed descriptions of how to retrieve and contribute information, as well as the database model and specifications, will be published elsewhere[10]. The approach described is applicable to a wide variety of conserved biological processes studied in different species, including *Homo sapiens*.

ACKNOWLEDGMENTS

We thank B. Braun for discussions, B. Masdoua and V. Cassen for contributing to programme development and M. Cherry and M. Yamamoto for hosting the database on local servers. We acknowledge funding for Workshops from the SNF, CNRS, INSERM and the US National Science Foundation. The initial stages of this project were supported by an EU grant awarded to N.L. Continuing support is provided by a grant from the Swiss Institute of Bioinformatics awarded to M.P. Software and database developers are C.W., R.B., C.S.d.M., L.H., R.K. and U.S.

M Primig[1], C Wiederkehr[1], R Basavaraj[1], C Sarrauste de Menthière[2], L Hermida[1], R Koch[1], U Schlecht[1], H G Dickinson[3], M Fellous[4], J A Grootegoed[5], R S Hawley[6], B Jégou[7], B Maro[8], A Nicolas[9], T Orr-Weaver[10], T Schedl[11], A Villeneuve[12], D J Wolgemuth[13], M Yamamoto[14], D Zickler[15], N Lamb[2] & R E Esposito[16]

[1]Biozentrum & Swiss Institute of Bioinformatics, Basel, Switzerland. [2]Institut de Génétique Humaine, Montpellier, France. [3]Oxford University, UK. [4]Hopital Cochin, Paris, France. [5]Erasmus MC, Rotterdam, The Netherlands. [6]The Stowers Institute, Kansas City, Missouri, USA. [7]The University of Rennes I, Bretagne, France. [8]Université Paris VI, France. [9]Institut Curie, Paris, France. [10]The Whitehead Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [11]Washington University, St. Louis, Missouri, USA. [12]Stanford University, Palo Alto, California, USA. [13]Columbia University, New York, New York, USA. [14]The University of Tokyo, Japan. [15]Université Paris XI, Orsay, France. [16]The University of Chicago, Chicago, Illinois, USA. Correspondence should be addressed to M.P. (michael.primig@unibas.ch).

1. Nasmyth, K. *Annu. Rev. Genet.* **35**, 673–745 (2001).
2. Baxevanis, A.D. *Nucleic Acids Res.* **31**, 1–12 (2003).
3. Yeh, I., Karp, P.D., Noy, N.F. & Altman, R.B. *Bioinformatics* **19**, 241–248 (2003).
4. Wheeler, D.L. *et al. Nucleic Acids Res.* **31**, 28–33 (2003).
5. Wood, V. *et al. Nature* **415**, 871–880 (2002).
6. Schlecht, U. & Primig, M. *Reproduction* **125**, 447–456 (2003).
7. Cho, R.J. *et al. Mol. Cell* **2**, 65–73 (1998).
8. Primig, M. *et al. Nat. Genet.* **26**, 415–423 (2000).
9. Williams, R.M. *et al. Proc. Natl. Acad. Sci. USA* **99**, 13431–13436 (2002).
10. Wiederkehr, C. *et al. Nucleic Acids Res.* (in the press).

Chipping away at the chip bias: RNA degradation in microarray analysis

To the editor:

Measurement of gene expression is based on the assumption that an analyzed RNA sample closely represents the amount of transcripts *in vivo*. Transcripts show stability differences of up to two orders of magnitude *in vivo*[1], raising the possibility that partial degradation during cell lysis could cause a variable extent of bias in quantification of different transcripts. One of the most effective tools for characterizing RNA integrity is capillary electrophoresis, in which RNA degradation is indicated by an altered 28S/18S rRNA signal ratio[2]. In the software of the commonly used system (Bioanalyzer 2100, Agilent), quantification of 18S and 28S rRNA is compromised by the fact that this calculation is based on area measurements that are heavily dependent on definitions of start and end points of peaks (**Fig. 1a**). Even accurate determination of this ratio is not sufficient to detect degradation efficiently (**Fig. 1b**). We developed a mathematical model that results in an objective number for quantitative characterization of RNA degradation. Aside from three prominent peaks (small RNAs, 18S and 28S rRNA), a chromatogram of the size distribution of cellular RNAs shows a broad range of molecular weights with much weaker signals. With increasing degradation, heights of 18S and 28S peaks gradually decrease and additional 'degradation peak signals' appear in a molecular weight range between small RNAs and the 18S peak (**Fig. 1b**). The ratio of the average degradation peak signal to the 18S peak signal multiplied by 100 will hereafter be referred to as the degradation factor. This analysis has been tested on 19 tissues of seven organisms, and it is a reproducible parameter for degradation of mammalian RNA (**Supplementary Table 1** online). As an example, 12 repeated measurements of the same sample yielded an average degradation factor of 27.14 with a standard deviation of 1.06. Degradometer software for calculation of the degradation factor can be downloaded from http://www.dnaarrays.org.

If one RNA sample was intact and the other was degraded during isolation, up to three-quarters of the differential gene expression measured was due solely to differences in RNA integrity between two samples (**Fig. 1d**). **Supplementary Figure 1** online shows changes in mRNA levels caused by alteration of RNA integrity. This effect was independent of the algorithm applied to raw data analysis (**Supplementary Tables 2**, **3** and **4** online).

For *GAPD* and *ACTB*, two transcripts for which signal intensities from 3′ and 5′ portions are frequently measured in microarray analysis, there is a positive correlation between the 3′/5′ ratio and the degradation factor of samples (**Fig. 1c**). This correlation is tissue-dependent (**Supplementary Fig. 2** online).The smaller the difference in degradation factors between samples, the more closely the measured expression differences reflect biological differences (**Fig. 1d**).

Aside from general RNase activity by members of the RNase A family[3], RNase L, an enzyme activated in apoptotic

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

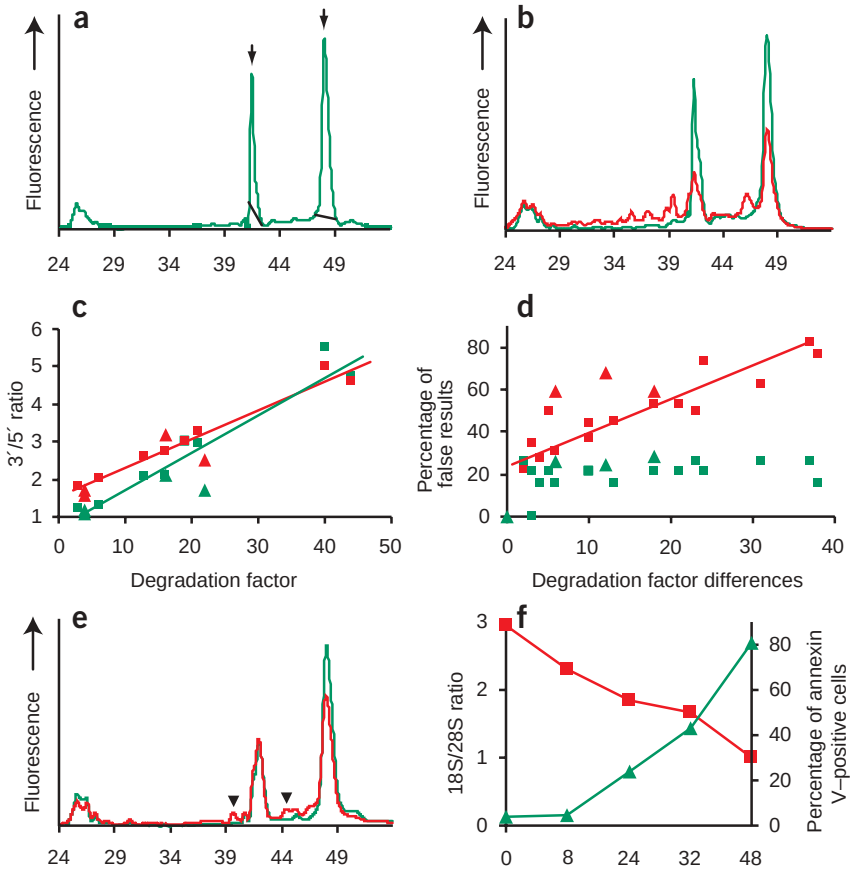

Figure 1 Quantification of RNA integrity and consequences of degradation for expression analysis. (**a**) Chromatograms of RNA size distribution. Area measurements for rRNA calculation can be misleading owing to incorrect definition of the area baseline. Shown is an example of automatic baseline detection by the Bioanalyzer software. (**b**) Green lines show intact RNA (degradation factor 3.9 and rRNA ratio 1.8), and red lines show degraded RNA (degradation factor 22 and rRNA ratio 1.8). (**c**) The relationship between degradation factor and 3′/5′ ratios of *GAPD* (red) and *ACTB* (green), calculated for RNA degraded after purification (squares) and RNA degraded in its cellular environment (triangles). For technical details see **Supplementary Methods** online. The linear correlations are $R^2 = 0.91$ for *GAPD* and $R^2 = 0.89$ for *ACTB*. (**d**) False positive and negative rates of differential gene expression due to degradation. Comparison of expression profiles from two different intact RNA samples is considered the true differential expression pattern (0% false results). Red represents false positive results and green represents false negative results of differential gene expression measurements, caused by differential degradation of the two samples. Results are shown for RNA degraded after purification (squares) and RNA degraded in its cellular environment (triangles). False negative results show no obvious correlation with differences in degradation factor, whereas false positive results show a linear correlation ($R^2 = 0.63$). (**e**) Effects of apoptosis on RNA size distribution. RNA of NB4 cells, mock-treated (green) or treated for 24 h with 5 mM valproic acid to induce apoptosis[7] (red). Fluorescence intensity of 28S rRNA is low relative to that of 18S rRNA. Two additional peaks (arrowheads) appear during apoptosis. (**f**) The relationship between the 28S/18S rRNA ratio and the percentage of annexin V–positive cells during apoptosis. The *x* axis shows duration of treatment with valproic acid.

© 2003 Nature Publishing Group http://www.nature.com/naturegenetics

pathways[4], can alter the size distribution of total RNA. During apoptosis, 28S rRNA is cleaved more quickly than 18S rRNA (**Fig. 1e** and ref. 4). The Degradometer software output file contains the ratio of 28S/18S rRNA signal heights. This ratio allows detection of apoptosis even before cells become positive in annexin-V staining (**Fig. 1f**). Certain treatments of biological samples *ex vivo* can induce apoptosis[5], causing bias in the results of gene expression measurement. When we induced apoptosis in cell culture, gene expression profiles changed significantly (**Supplementary Fig. 1** online).

Reproducibility and reliability of results obtained by microarray technology currently are important goals of the field[6]. Minimum Information About a Microarray Experiment (MIAME) should allow standardization and therefore reproducibility. Inclusion of RNA integrity parameters, such as the degradation factor and the 28S/18S signal heights, in MIAME would help characterize the biasing effects of RNA degradation and apoptosis on microarray results.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank H. Klump, S. M. Tanner and K. Becknell for their critical reading of the manuscript. This work is supported in part by a grant from the US National Cancer Institute.

Herbert Auer[1], Sandya Lyianarachchi[1], David Newsom[1], Marko I Klisovic[2], uido Marcucci[2] & Karl Kornacker[3]

[1]Division of Human Cancer Genetics, [2]Department of Internal Medicine and [3]Division of Sensory Biophysics, The Ohio State University, Columbus, Ohio 43210, USA. Correspondence should be addressed to H.A. (auer-2@medctr.osu.edu).

1. Berger, S.L. & Cooper, H.L. *Proc. Natl. Acad. Sci. USA* **72**, 3873–3877 (1975).
2. Van de Goor, T.A. *PharmaGenomics* **3**, 16–18 (2003).
3. Rosenberg, H.F. & Dyer, K. *Nucleic Acids Res.* **24**, 3507–3513 (1996).
4. Nadano, D. & Sato, T.A. *J. Biol. Chem.* **275**, 13967–13973 (2000).
5. Grasl-Kraupp, B. *et al. Hepatology* **25**, 906–912 (1997).
6. Brazma, A. *et al. Nat. Genet.* **29**, 365–371 (2001).
7. Göttlicher, M. *et al. EMBO J.* **20**, 6969–6978 (2001).