Normalization of array hybridization experiments in differential gene expression analysis

Bodo Eickhoff*, Bernhard Korn[1], Matthias Schick[1], Annemarie Poustka[1] and Juergen van der Bosch

Division of Experimental Immunopharmacology, Research Center Borstel, Parkallee 22, D-23845 Borstel, Germany and [1]Division of Molecular Genome Analysis, Resource Center (RZPD), Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany

Received July 21, 1999; Revised and Accepted September 29, 1999

ABSTRACT

For detecting and confirming differentially expressed genes it is necessary to have a trustworthy reference. So called 'housekeeping genes' are frequently used for this purpose as internal standard. However, if the influence of new experimental conditions is to be analyzed it is not safe to assume a priori that the expression of these genes is not affected. Therefore two synthetic poly(A)-RNAs were generated by PCR and *in vitro* transcription. They were used as external standards for normalization of northern blots and cDNA arrays where non-regulated genes as internal reference were not available.

INTRODUCTION

One method for analyzing differential gene expression is complex hybridization. For this purpose commercially manufactured gene arrays become increasingly available. These are arrays of bacterial cDNA clones with known sequences [e.g. the GDA filters from Genome Systems or high density clone filters of the Resource Center (RZPD)] or cDNA arrays consisting of PCR-generated probes (e.g. the Atlas Human cDNA Array from Clontech or PCR high density filters of the RZPD). These membranes are hybridized with a radioactive (usually ^{32}P- or ^{33}P-labeled) probe generated by reverse transcription of poly(A)$^+$ RNA or total RNA from the cells or tissues to be compared. After washing non-specific probe away, the arrays are subjected to phosphorimaging and data analysis.

A major problem in evaluating the results is normalizing the two filters. In order to identify differentially expressed genes, the two filters cannot usually be compared directly. Instead, they need to be normalized in order to compensate for differences due to varying efficiencies of RNA preparation, reverse transcription, probe purification, hybridization, filter quality, etc. For normalization, housekeeping genes are frequently used as an internal reference. On both clone filters and cDNA arrays there are several such genes available. However, comparing unknown states of gene expression, constancy of housekeeping gene expression cannot be assumed a priori. Furthermore, there are several reports in the literature describing housekeeping genes to be regulated (1,2).

In order to circumvent these drawbacks of housekeeping genes as internal standards, two synthetic RNAs were synthesized as external references to be added to the cell extract prior to RNA preparation. Important criteria to be met by the normalization standards were: (i) being an RNA molecule, (ii) presence of an oligo(A) tail for selection with oligo-dT cellulose or beads and as primer binding site for reverse transcription, (iii) a sequence not related to human sequences, (iv) the presence of hybridization targets for the standards on the arrays, and (v) an easy way of synthesis. Consequently, we developed two synthetic RNA molecules, one derived from the ampicillin resistance (amp) and one from the kanamycin resistance (kan) gene, for use as normalization standards in differential gene expression analysis.

MATERIALS AND METHODS

Synthesis of standard-RNAs

791 bp of the amp gene and 1045 bp of the kan gene were amplified by polymerase chain reaction (PCR) incorporating a T7 promoter via the 5′ primer and an oligo(A)$_{30}$ tail via the 3′ primer (Fig. 1) (3). For amplifying the amp gene 10 ng pBluescript KS+ as template DNA was submitted to the following temperature program: 96°C, 2 min; 30 cycles of 96°C for 20 s, 60°C for 30 s, 72°C for 60 s; final extension of 72°C, 2 min. Each reaction contained 1 U Taq polymerase (MBI), 2.5 µl 10× reaction buffer, 2.5 µl 25 mM $MgCl_2$, 0.5 µl 10 mM dNTPs, 30 pmol each primer and nuclease-free water to a final volume of 25 µl. Primer sequences: 5′ampT7: GACTAATACGACTCACTATAGGTCCATAGTTGCCTGACTCCC; 3′amppA: (T)$_{30}$TTCCGTGTCGCCCTTATTC.

For amplifying the kan gene, 10 ng of pCRII vector (Invitrogen) was used as template with the amplification protocol given above. Primer sequences: 5′kanT7: GACTAATACGACTCACTATAGGCGGAACACGTAGAAAGCCAG; 3′kanpA: (T)$_{30}$CGATACCGTAAAGCACGAGG.

The resulting PCR products were purified with the Qiaquick PCR Purification Kit (Qiagen) and then transcribed to RNA using a T7 Transcription Kit (MBI Fermentas). Subsequently, the template DNA was removed by DNase I digestion (message clean kit, Genehunter Corporation) for 30 min at 37°C. After phenol/chloroform extraction, RNA was ethanol-precipitated

*To whom correspondence should be addressed. Tel: +49 4537 188458; Fax: +49 4537 188404; Email: beickhof@fz-borstel.de

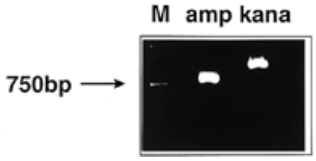

Figure 1. 1% Agarose gel electrophoresis of ampicillin-resistance amplicon and kanamycin-resistance amplicon (ethidium bromide stained). Lane M, DNA size marker (1 kb ladder, MBI); lanes 2 and 3, 10 µl of each indicated PCR product.

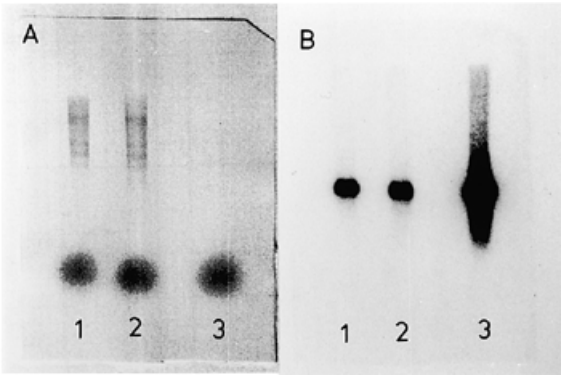

Figure 2. (**A**) Methylene blue staining of northern blot. Lane 1, 2 µg mRNA #1; lane 2, 2 µg mRNA #2; lane 3, 1.6 ng *in vitro* transcribed amp-RNA. The spots present at the bottom of all three lanes derive from the loading dye. (**B**) The same blot after hybridization with a ^{32}P-labeled single-strand antisense probe to amp-RNA after 48 h exposure of Kodak Biomax MS X-ray film with intensifying screen. Probe preparation was as follows: 3 µg amp-RNA, 1 µl M-MuLV RT (MBI), 3 µl 5× reaction buffer, 60 pmol 3′ primer, 80 µCi [α-^{32}P]dATP and 2 µl dNTPs (5 mM each dCTP, dGTP, dTTP, 50 µM dATP) were incubated at 37°C for 1 h, unincorporated nucleotides removed with a G-25 spun column (Pharmacia) and the probe hybridized to the membrane for 2 h at 68°C in ExpressHyb solution (Clontech). Washing was performed as described by the manufacturer.

and quantified using spectrophotometric absorbance at 260 nm (4). Aliquots of each standard RNA were shock-frozen in liquid nitrogen and stored at –70°C. They could hence be used singly as well as in combination. Absence of DNA was confirmed via PCR using the synthesis-primers for each gene. Incorporation of the oligo(A) tail at the 3′ end of the nascent RNA was confirmed by oligo-dT selection with Dynabeads (Dynal) of the [α-^{32}P]UTP labeled *in vitro* transcript. For RNA isolation ~1 ng of each standard RNA was added to cell lysate obtained from 10^4 human lung adenocarcinoma cells.

Complex hybridization onto cDNA array

The probe for complex hybridization was synthesized using 5 µg lung adenocarcinoma total RNA, 1.5 µl Superscript II RNase H– reverse transcriptase (Life Technologies), 2 µM oligo(dT)$_{18}$VN primer, 0.5 mM each dCTP, dGTP, dTTP and 5 µM dATP, 100 µCi [α-^{32}P]dATP (3000 Ci/mmol, ICN) and 333 µM DTT in 1× First Strand Buffer (Life Technologies) for 90 min at 37°C. Unincorporated nucleotides were removed using a G-25 spun column (Pharmacia) and the probe was heat-denatured for 3 min at 99°C. The filter was prehybridized in 10 ml ExpressHyb (Clontech) solution with 20 µg yeast tRNA at 68°C for 30 min. Hybridization was performed overnight at 68°C. Washing was done at 68°C, twice for 15 min with 2× SSC/1% SDS and twice for 20 min with 0.1× SSC/0.1% SDS.

RESULTS

Two RNA molecules, one derived from the amp gene and one from the kan gene, were synthesized by PCR and *in vitro* transcription for use as normalization standards. These sequences are easily available from plasmid vectors and do not hybridize to any human sequence as confirmed by BLAST (5) search and northern blots (data not shown).

In order to control probe preparation *ab initio*, the standards were added at the time of cell lysis within the lysis buffer (6). An aliquot of standard RNA was thawed on ice and 5 µg were mixed with freshly lysed cells (5×10^7). RNeasy Midi kit (Qiagen) was used for total RNA isolation. Enrichment of poly(A)$^+$ RNA was performed twice with oligo-dT Dynabeads (Dynal) and followed by DNase I digestion. Aliquots of 2 µg each of the two samples to be compared were separated on a 1% formaldehyde–agarose gel (7) and blotted overnight onto a positively charged Nylon membrane (Nytran, Schleicher & Schuell) using neutral transfer buffer and the Turboblotter system (Schleicher & Schuell). The membrane was UV-cross-linked, stained with methylene blue (7) and photographed (Fig. 2A). Next, the membrane was hybridized with an amp-probe to check the quality, integrity and amounts of mRNA samples prepared. As shown in Figure 2B, the two parallel processed samples contained equal amounts of intact standard amp RNA of identical length. Hybridization with a kan-probe gave corresponding results (data not shown). After these quality-checks, the samples were used for complex hybridization.

To meet the needs of external reference normalization in complex hybridization, the gene-specific and standard-specific hybridization targets must be spatially separated on the filter. For this purpose, cDNA arrays have been designed and manufactured that contain 150 different PCR products generated from clones containing genes for cytokines and cell surface proteins. Furthermore, kan PCR products have been spotted onto the filter as guide dots and for normalization purposes. As shown in Figure 3, the singly spotted, regularly aligned kan-spots can easily be distinguished from gene-specific signals which appear as double spots. The presence of kan-dots on every part of the filter allows locally different background and filter-inhomogeneity to be considered in the normalization process. For the experiment shown, lung adenocarcinoma mRNA was used. The resulting signals mainly derived from cell-surface proteins. For the 144 guide dots present on each filter, the spot-to-spot signal intensity variability (calculated as the standard deviation) was determined to be ~25% of the mean value. This was the case for all filters analyzed so far. Thus, normalization was performed by using the mean signal intensity of the guide dots of each filter. As there are also housekeeping genes present on the filter, it is possible to determine to what extent their expression is affected by

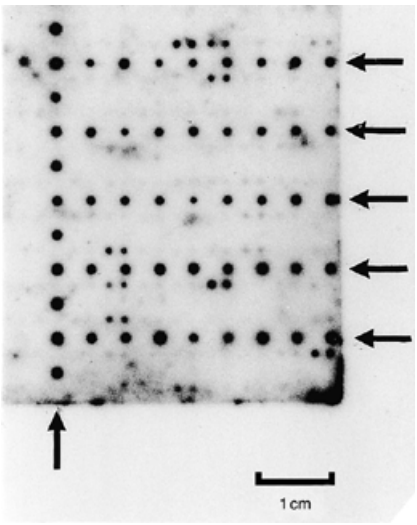

Figure 3. A PCR generated cDNA array (Resource Center, German Genome Project, Heidelberg, Germany) containing 150 human genes for cytokines and cell surface proteins as well as kanamycin-resistance amplicon guide dots (lines indicated by arrows). Five nanograms of PCR product (5 ng/µl) was spotted for each guide dot, which covered an average area of 0.125 mm^2. For each of the other 150 amplicons 1 µl (10 × 100 nl) of a solution containing 5–50 ng/µl DNA was spotted. (18 h exposure of Kodak Biomax MS X-ray film with intensifying screen.)

experimental conditions. For lung adenocarcinoma cells, GAPDH expression, as verified by northern blot, proved to be stable with respect to the experimental conditions investigated, whereas β-actin expression did not.

DISCUSSION

The external reference system described cannot be used with bacterial clone filters (such as GDA filters from Genome Systems), because the target DNA on these filters are cDNA vector and insert. As most plasmid vectors contain amp and/or kan genes they can hybridize with amp or kan probes. So additive signals from standard and mRNA will result and standard signals may unpredictably affect mRNA signals. Due to different colony growth, plasmid copy number and membrane processing, the spot-to-spot variability on the filters has been determined to be very high (data not shown).

For the manufactured PCR cDNA arrays shown in Figure 3 the spot-to-spot variability has been shown to be only ~25% of the mean guide dot signal intensity. This variability presumably derives from a sum of experimental variances, e.g. filter inhomogeneities caused by filter processing, fluctuations of hybridization- and washing-temperature (especially during solution changes) and even phosphorimaging (e.g., counting errors due to screen inhomogeneities). Since for the 150 target genes the spotted amounts vary by a factor of 10 and due to different hybridization kinetics for each sequence, the ratio of signal intensities does not reflect the relative expression levels of the different genes in one sample. How the DNA concentration spotted onto the filter influences the resulting signal will be determined in future experiments. It will probably be a non-linear relation because at >5 ng DNA per spot the DNA won't be saturated under standard conditions, only ~3% of the DNA in each spot hybridizes to the probe.

In summary, two *in vitro* synthesized oligo(A)-RNAs were developed as external normalization reference for use with PCR-generated cDNA arrays and northern blots in differential gene expression analysis. They can be used with several RNA isolation methods: RNeasy Mini-, RNeasy Midi-, Oligotex direct mRNA-Purification kit (Qiagen), Dynabeads oligo-dT (Dynal) and guanidiniumisothiocyanate/phenol/chloroform (7) have all been tested. Furthermore, they allow judgement of quality, quantity and integrity of total RNA and poly(A)$^+$ RNA when monitored with northern blot or RT–PCR. In addition they could be used as external standards/controls for subtractive hybridization techniques or Taqman-real time quantitative RT–PCR. Using various standard RNAs with different sequence compositions at defined ratios in the same sample, their ratio may serve to check hybridization and enzymatic reactions like reverse transcription or RT–PCR for any bias due to sequence (e.g., GC-content), amount, length or higher order structure of mRNA.

High density arrays from the RZPD in future will be provided with kan guide dots replacing ink dots to enable normalization with the presented reference system.

ACKNOWLEDGEMENTS

The idea of using guide dots on filters was contributed by Hans Lehrach (MPI Molecular Genetics, Berlin-Dahlem, Germany) and is gratefully acknowledged. Furthermore, we would like to thank S. Rüller and P. Scheinert for helpful discussions. This article is based on a doctoral study by Bodo Eickhoff in the Faculty of Biology, University of Hamburg.

REFERENCES

1. Savonet,V., Maenhaut,C., Miot,F. and Pirson,I. (1997) *Anal. Biochem.*, **247**, 165–167.
2. Bhatia,P., Taylor,W.R., Greenberg,A.H. and Wright,J.A. (1994) *Anal. Biochem.*, **216**, 223–226.
3. Urrutia,R., McNiven,M.A. and Kachar,B. (1993) *J. Biochem. Biophys. Methods*, **26**, 113–120.
4. Gallagh,S.R. (1997) In Coligan,J.E., Kruisbeek,A.M., Margulies,D.H., Shevach,E.M. and Strober,W. (eds), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York, NY, Appendix A3D.
5. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.
6. Toscani,A., Soprano,D.R., Cosenza,S.C., Owen,T.A. and Soprano,K.J. (1987) *Anal. Biochem.*, **165**, 309–319.
7. Chomczynski,P. and Sacchi,N. (1997) In Coligan,J.E., Kruisbeek,A.M., Margulies,D.H., Shevach,E.M. and Strober,W. (eds), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York, NY, Chapter 4.2.