Short Technical Report

Quantitative Detection of Low-Copy-Number mRNAs Differing at Single Nucleotide Positions

BioTechniques 34:172-177 (January 2003)

J. Stenman¹, S. Lintula¹,
O. Rissanen¹, P. Finne¹,
J. Hedström¹, A. Palotie^{1,2},
and A. Orpana¹

¹Helsinki University Central
Hospital, Helsinki, Finland, and
²UCLA, Los Angeles, CA, USA

ABSTRACT

Accurate analysis of mRNA expression levels of SNPs, highly homologous genes, and splicing variants requires techniques capable of quantifying low-copy-number mRNAs differing at single nucleotide positions. We have used an RT-PCR-based technique based on co-amplification of closely related target mRNA transcripts and assessed the effect of the stochastic distribution of low-copy-number templates on sampling variation when quantifying rare mRNA transcripts. The technique was optimized for maximal sensitivity to enable the analysis of samples containing a subpopulation of target cells and small microdissected samples. We demonstrate that the input level of template molecules is a critical determinant of the achievable assay precision. A minimum of approximately 50 molecules of template is required to discriminate between 2-fold differences in the expression levels of two transcripts. At levels above 1000 molecules of input template, the stochastic effects on sampling variation become negligible.

INTRODUCTION

High-throughput techniques for analyzing genetic variation are rapidly providing novel data on variants causative of, or related to, human disease. The most common form of human genetic variation is SNPs, which can, when located in the coding region of a gene, cause an alteration in the specific gene product and thereby predisposition to a disease. In many cases, however, the predisposing variant is located in the regulatory region of a gene, resulting only in an altered expression level of the affected allele. Unbalanced expression of the alleles, indicating differences in the control of expression, could thus be used as an aid in finding the diseases associated with genetic variation. One strategy for the identification and verification of the biological role of these polymorphisms is to acquire marker SNPs in the coding region of the gene from available databases and to use them to quantify experimentally the relative SNP expression patterns of the alleles. This requires techniques able to accurately distinguish templates differing at single nucleotide positions, sensitive enough for quantification of the mRNA expression levels in a subpopulation of target cells in a clinical sample.

Several amplification techniques enable quantitative detection of rare mRNA transcripts in subpopulations of target cells in clinical samples (1–3,7,8,10,13,21) and from more abundant transcripts deriving from single cells (14). However, the assay precision decreases dramatically at low levels of target transcripts (5,11,18). This phenomenon is caused in part by methodological inaccuracy, but it is also to a

great extent affected by the sampling variation arising from the stochastic distribution of low-copy-number template molecules in internal standards, calibrators, and clinical samples (17).

The aim of this study was to assess the effect of the stochastic distribution of low-copy-number templates on sampling variation. For this purpose, we have optimized a highly sensitive RT-PCR-based technique for co-amplification and quantitative detection of squamous cell carcinoma antigen (SCCA) mRNA transcripts. The wild-type transcripts and internal standards are distinguished by single nucleotide differences using post-amplification minisequencing. This experimental setup provides a model for quantitative SNP expression analysis. To determine the imprecision imposed by stochastic sampling variation, we have related experimentally obtained results to the statistically predicted optimal accuracy at different input levels of target templates.

MATERIALS AND METHODS

Five-milliliter blood samples were hemolyzed by hypotonic shock, leukocytes were pelleted, and total RNA was extracted using RNeasy[®] spin columns (Qiagen, Hilden, Germany). DNase digestion was carried out in the column membranes as described by the manufacturer. The total RNA preparations were stored at -80°C until analysis. mRNA was oligo-dT-captured on the wall of 0.2-mL streptavidin-coated PCR tubes (Roche Diagnostics, Basel, Switzerland) from the 0.5–1.0 µg total RNA preparations. Reverse transcription and first-round PCR of the SCCA1 and

SCCA2 templates were carried out in a 50- μ L reaction, in the presence of 20 molecules of a nonlinearized plasmid containing the internal DNA standard

sequence (SuperScript IITM reverse transcriptase; Invitrogen, Carlsbad, CA, USA/AmpliTa[®] Gold DNA polymerase; Applied Biosystems, Foster City, CA, USA). Following amplification, 2 μ L PCR product were transferred to new reaction tubes, and all reactions were amplified to the plateau in a 100- μ L nested PCR using DyNAzyme IITM DNA polymerase (Finnzymes, Espoo, Finland) and a 5'-biotinylated downstream primer. Following amplification, the PCR products were captured in streptavidin-coated scintillating 96-well microplates (Perkin Elmer, Norwalk, CT, USA) and minisequenced using DyNAzyme II- and 3H-labeled nucleotides (4). The ratio of wild-type and internal standard tem-

plates was quantified in a beta counter. Primer sequences, sequence information, and cycling conditions are available at http://www.nature.com/nbt/web_extras/ (supplement to Reference 15).

Complementary RNA (cRNA) standard templates were used for determining the measuring range of the assay (Figure 1), and the SCCA1 and SCCA2 expression levels in cultured A431 cells were synthesized as follows. The specific SCCA2 and SCCA1 PCR amplicons were cloned into a pCR II vector (Invitrogen) and linearized with *KpnI* restriction endonuclease. In vitro transcription was performed using the Ampliscribe T7-transcription kit (EPI-CENTRE, Madison, WI, USA) according to the manufacturer's instructions. The cRNA standard templates were extracted using RNeasy spin columns, and DNase digestion was carried out in the column membranes. The final cRNA templates were diluted in DEPC-treated water and stored in aliquots at -80°C until RT-PCR analysis.

To determine the precision of the assay, a standard curve of PCR calibrators consisted of a 5-fold dilution series of 1- μ g aliquots of pooled leukocyte total RNA spiked with 4–500 cultured A431 epidermoid carcinoma cells (ATCC,

Figure 1. Measuring the range of the assay. Twenty molecules of double-stranded plasmid DNA containing the internal standard sequence were co-amplified with a serial dilution of cRNA standard templates ranging from 10 to 10⁶ copies spiked into mRNA from 1 μ g leukocyte total RNA. The cRNA standard template contained T7-RNA polymerase-generated templates of both SCCA1 and SCCA2 mixed in a 1:1 ratio. Upper panel, ethidium bromide-stained gel showing the 108-bp PCR products containing the products of both SCCA RNA and the internal standard templates. All PCRs are amplified to the plateau (30 + 35 cycles) to bring the PCR products into the measuring range of the minisequencing detection assay, irrespectively of the amount of input template. Middle panel, cpm measured from the dilution series of standard cRNA templates and corresponding RT-controls. Lower panel, the ratio of internal standard to SCCA wild-type mRNA was linear from 10 to 10⁴ copies of cRNA standard products.

Figure 2. Statistically estimated sampling variation of rare template molecules. Means and 95% confidence intervals of single templates versus two equimolar templates occurring in the same sample. The imprecision caused by stochastic distribution of highly diluted template molecules is greatly pronounced when low-copy-number target templates are quantified against a reference template. CI, confidence interval.

DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

Rockville, MD, USA). This dilution series was amplified separately 12 times on separate days, and the confidence intervals of the mean values at each level of input templates were compared to the statistically estimated confidence interval of the mean for the sampling variation at the corresponding levels.

The sampling variation attributable to the stochastic variation at low template concentrations was calculated as follows. The probability of a certain ratio of wild-type to reference templates to be present in an aliquot pipetted from a sample containing a known ratio of templates was expressed as $P_{\text{Ratio}} = P_{\text{WT}} \times P_{\text{Ref}}$, where P_{WT} and P_{Ref} describe the probabilities of the absolute copy numbers of wild-type and reference template molecules present in the resulting aliquot. As the same ratio of wild-type to reference templates can arise from different combinations of absolute copy numbers, the probability of a certain ratio was estimated by summing the probabilities of the different copy number combinations, resulting in this ratio. The probability distributions and confidence intervals of a ratio of two templates in an aliquot were estimated from the cumulative probabilities of the most frequent ratios. A Poisson distribution was assumed for templates present at levels of 50 molecules or less, whereas a normal distribution was assumed for templates present at levels above 50 molecules/reaction (Figure 2).

To test the reproducibility of the assay, 385 blood samples from 131 individuals were analyzed. Informed consent was obtained from all patients participating in the study. The study was approved by the ethical committees of the department of Gynecology and Obstetrics and the department of Otorhinolaryngology, Helsinki University Central Hospital (Helsinki, Finland).

RESULTS

The degree of sampling variation caused by the stochastic distribution of rare template molecules occurring in a solution was statistically estimated for different levels of input template (Figure 2). When sampling a target transcript without a reference template, the 95% confidence intervals at 8, 40, 200,

and 1000 molecule levels were $\pm 62\%$, $\pm 30\%$, $\pm 14\%$, and $\pm 6\%$, respectively. However, when the sample contains a target (wild-type) and a reference template, the variation in the ratio between the two templates is dependent on the stochastic distribution of both of these templates, resulting in a multiplication of stochastic effects. This variation increases dramatically when the amount of template copies decrease; at a level of eight copies of both target and reference templates, the 95% confidence interval spans a range of -70% to $+200\%$.

To optimize the number of internal standard molecules for sensitive quantification purposes, we co-amplified 20 test samples containing mRNA from approximately 1 μg leukocyte total RNA with 500, 100, and 20 molecules of double-stranded plasmid DNA containing the internal standard sequence. Between these levels, there were no significant differences in amplification reproducibility. When using 20 molecules of internal standard DNA, amplification failed in only 4 of 385 test samples and thus provided verification of the sensitivity of the PCR in the individual samples. At this level, no amplification of endogenous background SCCA transcripts or possible contaminating templates was observed in any of the 385 test samples. No attempt was made to decrease the amount of internal standard below this level.

The sensitivity and measuring range of the assay was tested by preparing a serial dilution of wild-type SCCA cRNA templates transcribed in vitro, ranging from 10 to 10^6 copies, in mRNA captured from 1 μg leukocyte total RNA (Figure 1, upper panel). The PCR product ratio of internal standard to SCCA wild-type RNA was linear from 10 to 10^4

copies of wild-type RNA when using 20 molecules of internal standard DNA (Figure 1, middle and lower panels). We did not observe amplification from contaminating DNA in any of the corresponding reverse transcriptase-deficient controls (Figure 1, middle and lower panels). It is evident that some degree of DNA contamination will remain in the cRNA preparations transcribed in vitro, despite RNA extraction and DNase digestion. In a real-time PCR setup with no competing template, this could occasionally give rise to an amplification of single or a few contaminating DNA molecules in the RT- controls. What seems to be critical is that the 20 molecules of double-stranded plasmid DNA used as an internal standard in all reactions was sufficient to suppress competitively the amplification of single molecules of DNA that might occasionally be present in the cRNA preparations. The fact that no background amplification of endogenous SCCA transcripts was observed in any of the 385 blood samples studied strengthens this hypothesis.

To evaluate the performance of the

Figure 3. The observed imprecision in relation to the statistically estimated variation caused by sampling of highly diluted template molecules. A 1:5 dilution series of A431 epidermoid carcinoma cells containing approximately two copies of SCCA mRNA/cell was amplified separately 12 times, and the 95% confidence intervals of the mean were plotted against the statistically estimated 95% confidence intervals for the corresponding sampling variation at each level. The confidence intervals for the sampling variation were calculated from the probability distributions of the ratio of the internal standard (20 dsDNA molecules) and the number of wild-type SCCA mRNA templates at the given levels. CI, confidence interval.

DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

RT-PCR assay at low target transcript levels, we prepared a dilution series of mRNA from cultured A431 epidermoid carcinoma cells spiked in 1 μ g normal leukocyte total RNA. The A431 cells contained approximately two copies of SCCA mRNA/cell, and the ratio of SCCA2 to SCCA1 mRNA was 0.9. Cultured cells expressing the SCCA genes were preferred in this experiment to mimic naturally occurring full-length mRNA transcripts. mRNA from this dilution series, containing 8, 40, 200, and 1000 copies of the SCCA transcripts, was captured, reverse-transcribed, and amplified separately 12 times on separate days. To assess the effect of the stochastic distribution of low-copy-number templates on sampling variation, the experimentally obtained confidence intervals of the mean values at each level of input templates were compared to the statistically estimated confidence intervals for the sampling variation at the corresponding

levels (Figure 3). At wild-type SCCA mRNA levels exceeding 40 molecules, the observed confidence intervals were very close to those statistically predicted. This indicates that the sampling variation arising from the stochastic distribution of low-copy-number template molecules in the highly diluted samples was responsible for a significant portion of the quantitative variation in the observed wild-type template to internal standard ratios.

DISCUSSION

The rapid progress in discovering random SNPs covering the entire genome has created a need for sensitive techniques to analyze quantitatively the relative expression levels of alleles in different experimental settings. The highly homologous SNP alleles provide optimal targets for accurate quantification of relative expression levels in

a co-amplification system. When quantifying a target mRNA transcript against a highly homologous reference template, such as the transcript of an SNP allele, a closely related gene, a splicing variant, or an internal standard, the amplification reactions can be designed to produce PCR products of equal length and close to equal base-pair composition. In this way, secondary structures caused by folding of the single-stranded PCR products after denaturation are unified, and differences in annealing and elongation efficiencies of the primers to the target and reference templates are subsequently minimized. As a result, replication efficiencies of the different templates are equalized. This is a critical to achieving optimal precision in amplification-based mRNA quantification systems for measuring relative levels of two templates (9,12).

We have used minisequencing (single nucleotide primer extension) for

post-amplification quantitative detection of PCR products. This technique provides highly discriminate detection of single nucleotide sequence differences and thus enables accurate quantitative detection of co-amplified templates differing only minimally from each other (6,15,16,19). As the amplification efficiencies of the almost identical wild-type and reference templates are equally affected by factors limiting exponential amplification, the ratio of the templates remains unchanged despite saturation of the amplification reaction. PCR can thus be run to the plateau without losing quantitative accuracy (12,15), and amplification products are conveniently brought into the dynamic range of detection method regardless of the input amount of target mRNA transcripts. Quantitative PCR techniques based on kinetic data acquisition during cycling have become the mainstream methodology for amplification-based quantification of rare mRNA transcripts. In applications where quantification against a highly homologous reference template is preferred, co-amplification techniques provide unmatched accuracy. Molecular beacon technology (20) allow sufficient discrimination of single nucleotide differences and could be used to combine the beneficial features of co-amplification and kinetic PCR as well as the use of other RNA amplification technologies.

Regardless of the amplification and detection methods used, stochastic sampling variation at low template concentrations is a crucial limiting factor in quantitative PCR systems optimized for maximal sensitivity (17). Assuming that the template molecules are evenly distributed in a solution of a known concentration, the probability of a certain number of molecules to be present in an aliquot pipetted from this solution can be calculated according to the Poisson distribution, as shown in limiting dilution assays (18). These effects are greatly pronounced when a wild-type mRNA template is quantified against a reference template such as an internal standard or another mRNA species or SNP allele present in the sample. In this case, the ratio of the two templates is dependent on the probability distribution of the copy numbers of each of the respective template molecules. Because of this

multiplication of stochastic effects, the achievable precision of any amplification system in which the wild-type copy number is quantified against a reference template is extremely dependent on the amount of input template when copy numbers approach unity (Figure 2). At levels above 1000 molecules of input template, the effects of stochastic sampling variation are moderate and affect single templates and ratios between a target and reference template similarly. Below this level, however, sampling variation is substantial, and approximately 50 molecules of input template are required to discriminate 2-fold differences in expression levels. This phenomenon explains, in part, the variable results commonly obtained with sensitive amplification-based assays. It is essential that the absolute amount input template can be determined, when the ratio of a rare target and a reference template is being quantified.

REFERENCES

1. Compton, J. 1991. Nucleic acid sequence-based amplification. *Nature* 350:91-92.
2. Guatelli, J.C., K.M. Whitfield, D.Y. Kwok, K.J. Barringer, D.D. Richman, and T.R. Gingeras. 1990. Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc. Natl. Acad. Sci. USA* 87:7797.
3. Haase, A.T., E.F. Retzel, and K.A. Staskus. 1990. Amplification and detection of lentiviral DNA inside cells. *Proc. Natl. Acad. Sci. USA* 87:4971-4975.
4. Ihalaenen, J., H. Siitari, S. Laine, A.C. Syvanen, and A. Palotie. 1994. Towards automatic detection of point mutations: use of scintillating microplates in solid-phase minisequencing. *BioTechniques* 16:938-943.
5. Karrer, E.E., J.E. Lincoln, S. Hogenhout, A.B. Bennett, R.M. Bostock, B. Martineau, W.J. Lucas, D.G. Gilchrist, and D. Alexander. 1995. In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries. *Proc. Natl. Acad. Sci. USA* 92:3814-3818.
6. Karttunen, L., L. Lonnqvist, M. Godfrey, L. Peltonen, and A.C. Syvanen. 1996. An accurate method for comparing transcript levels of two alleles or highly homologous genes: application to fibrillin transcripts in Marfan patients' fibroblasts. *Genome Res.* 6:392-403.
7. Lizardi, P.M., X. Huang, Z. Zhu, P. Bray-Ward, D.C. Thomas, and D.C. Ward. 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 19:225-232.
8. Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51:263-273.
9. Nedelman, J., P. Heagerty, and C. Lawrence. 1992. Quantitative PCR with internal controls. *Comput. Appl. Biosci.* 8:65-70.
10. Nuovo, G.J., G.A. Gorgone, P. MacConnell, M. Margiotta, and P.D. Gorevic. 1992. In situ localization of PCR-amplified human and viral cDNAs. *PCR Methods Appl.* 2:117-123.
11. Peccoud, J. and C. Jacob. 1996. Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophys. J.* 71:101-108.
12. Raeymaekers, L. 2000. Basic principles of quantitative PCR. *Mol. Biotechnol.* 15:115-122.
13. Saiki, R.K., T.L. Bugawan, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1986. Analysis of enzymatically amplified β -globin and HLA-DQ α DNA with allele-specific oligonucleotide probes. *Nature* 324:163-166.
14. Schutze, K. and G. Lahr. 1998. Identification of expressed genes by laser-mediated manipulation of single cells. *Nat. Biotechnol.* 16:737-742.
15. Stenman, J., P. Finne, A. Stahls, R. Grenman, U.H. Stenman, A. Palotie, and A. Orpana. 1999. Accurate determination of relative messenger RNA levels by RT-PCR. *Nat. Biotechnol.* 17:720-722.
16. Stenman, J., J. Hedstrom, R. Grenman, I. Leivo, P. Finne, A. Palotie, and A. Orpana. 2001. Relative levels of SCCA2 and SCCA1 mRNA in primary tumors predicts recurrent disease in squamous cell cancer of the head and neck. *Int. J. Cancer* 95:39-43.
17. Stenman, J. and A. Orpana. 2001. Accuracy in amplification. *Nat. Biotechnol.* 19:1011-1012.
18. Sykes, P.J., S.H. Neoh, M.J. Brisco, E. Hughes, J. Condon, and A.A. Morley. 1992. Quantitation of targets for PCR by use of limiting dilution. *BioTechniques* 13:444-449.
19. Syvanen, A.C., K. Aalto-Setälä, L. Harju, K. Kontula, and H. Soderlund. 1990. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* 8:684-692.
20. Tyagi, S. and F.R. Kramer. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14:303-308.
21. Zhou, Y., M. Calciano, S. Hamann, J.H. Leamon, T. Strugnelli, M.W. Christian, and P.M. Lizardi. 2001. In situ detection of messenger RNA using digoxigenin-labeled oligonucleotides and rolling circle amplification. *Exp. Mol. Pathol.* 70:281-288.

Received 11 June 2002; accepted 8 October 2002.

Address correspondence to:

Dr. Jakob Stenman
Dept. of Clinical Chemistry
Helsinki University Central Hospital
FIN-00029 HUS, Helsinki, Finland
e-mail: jakob.stenman@hus.fi