Short Technical Report

Statistical Validation of the Mammaglobin-Nested RT-PCR Assay for Tumor Cell Detection in Blood of Breast Cancer Patients

BioTechniques 31:1358-1362 (December 2001)

**O. Zach, H. Wagner¹,
H. Kasparu, O. Krieger,
and D. Lutz**

Elisabethinen Hospital ¹Institute
for Applied Statistics, University
of Linz, Linz, Austria

ABSTRACT

A stochastic model was developed to validate the results obtained with the mammaglobin-nested RT-PCR assay for tumor cell detection in peripheral blood of breast cancer patients. Since the assay consists of four PCR setups per peripheral blood sample, the probabilities for receiving 0, 1, 2, 3, or 4 positive setups were calculated. In this model, samples with just 500 mammaglobin mRNA molecules are highly probable to result in at least three positive setups, whereas lower quantities shift the probabilities towards one or two positive setups. In the clinical trial, samples with one or two mammaglobin positive setups were detected in 6/143 (4%) patients with benign lesions of the breast, in 41/310 (13%) breast cancer patients with no evidence of disease and in 39/157 (25%) breast cancer patients with metastatic disease. On the contrary, no sample from patients with benign lesions of the breast resulted in three or four positive setups, but 5/310 (2%) breast cancer patients with no evidence of disease and 46/157 (29%) with metastatic disease. These results correspond with the model: an increased number of tumor cells in peripheral blood lead to a higher amount of mammaglobin mRNA molecules, and these samples may result in at least three positive

setups. Samples with three or four positive setups were mainly derived from breast cancer patients with metastatic disease and only occasionally from patients with no evidence of disease. On account of these results, samples with at least three positive setups are of prognostic value and regarded as tumor cell positive.

INTRODUCTION

RT-PCR is a sensitive technique that employs oligonucleotide primers to amplify a distinct cDNA target within a heterogeneous cDNA sample (3). This enables the development of assays for various specific targets to monitor metastatic spread in peripheral blood (pB) of cancer patients.

We developed a sensitive and specific nested RT-PCR assay to detect mammaglobin (MGB1, SCGB2A2, and UGB2) mRNA transcripts in the pB of breast cancer (BC) patients and demonstrated a correlation between mammaglobin positivity, stage of disease, and carcinoembryonic antigen plasma levels (8). However, at the maximum sensitivity required for the detection of disseminated tumor cells, statistical effects increasingly become important for the consistency of results (5).

To assess the results of the mammaglobin-nested RT-PCR assay obtained with pB samples from BC patients, a stochastic model was developed to calculate probabilities of obtaining positive PCR results, depending on the number of mammaglobin mRNA molecules in the RNA preparation and considering the efficiency of cDNA synthesis.

MATERIALS AND METHODS

Patients

Peripheral blood samples were taken from female patients with benign lesions of the breast (21–84 years, mean 50) and from BC patients (26–94 years, mean 56) after informed consent according to local clinical legacy had been taken. BC patients were classified into two defined clinical subgroups: patients who had no evidence of disease (NED; with and without any adjuvant treatment) and patients with metastatic disease (MD; stage IV or relapse from any earlier stages).

Mammaglobin-Nested RT-PCR

Ten milliliters of pB of each patient were divided into two equal amounts; RNA was prepared from each fraction and dissolved in 50 μ L DEPC-treated water (Figure 1). Two microliters of RNA were reverse transcribed into first-strand cDNA in a volume of 20 μ L, and $2 \times 4 \mu$ L each cDNA were used for nested RT-PCR as described (8), leading to four PCR setups per sample. Each PCR setup was subjected to electrophoresis on 1.5% agarose gels stained with ethidium bromide.

Stochastic Model

To assess the quality of the assay, a stochastic model was developed that allows calculation of the probabilities for achieving 0, 1, 2, 3, or 4 positive PCR setups. The detection limit of DNA on agarose gels is at least 5 ng (2). Fifty

micrograms of a 1-kb DNA consists of 4.74×10^{13} molecules (6); therefore, 5 ng of the 201-bp mammaglobin PCR product consists of 2.37×10^{10} molecules, which is the minimum number of molecules required for a positive PCR result. Consequently, a sufficient number of mammaglobin mRNA molecules in the RNA preparation and the efficiency of reverse transcription are essential for reaching that number of amplified molecules by nested RT-PCR.

Let R denote the number of mammaglobin mRNA molecules in 50 μL water. Assuming that the mRNA molecules are distributed uniformly in the solution, then the amount S contained in 2 μL taken for cDNA synthesis has a binomial distribution with R and $\pi = 2/50 = 0.04$, that is,

$$Pr(S = s) = \binom{R}{s} \pi^s (1 - \pi)^{R-s}$$

Let eff denote the mean percentage of mRNA transcribed into cDNA (4). The distribution of the resulting amount of cDNA D is then binomially distributed with parameters S and $\pi = \text{eff}$.

It can easily be shown that, under these assumptions, the resulting amounts U_1 and U_2 taken from the same cDNA solution ($2 \times 4 \mu\text{L}$ out of 20 μL) for PCR are multinomially distributed with parameters $\pi_1 = \pi_2 = \pi = 0.04 \times 0.2 \times \text{eff}$, that is,

$$Pr(U_1 = u_1; U_2 = u_2) = \binom{R}{u_1, u_2} \pi^{u_1 + u_2} (1 - \pi)^{R - u_1 - u_2}$$

Thirty PCR cycles result in an amount V_i of DNA molecules, whereby V_i is a function of U_i . Two microliters of the first PCR product are taken for the second amplification. The number of DNA molecules in this volume, denoted as

Figure 1. Schematic presentation of the mammaglobin nested RT-PCR assay.

Molecular Diagnostic Techniques

W_i has a binomial distribution with parameters V_i and $\pi = 0.04$. The number of DNA molecules after the second amplification is X_i . To calculate the probabilities for $k = 0, 1, 2, 3$, and 4 positive results of a sample, the relationship between V_i and U_i , and X_i and W_i , respectively, has to be specified.

Theoretically, the amount of DNA is doubled in each PCR cycle, but a rapid perfect PCR simulation shows that the amount of final PCR products after n cycles is $2^{n-1} - n$ when the target consists of a single-stranded cDNA molecule and $2^n - 2n$ when the target consists of a dsDNA molecule, that is, $V_i = (2^{29} - 29)U_i$ for the first PCR and $X_i = (2^{30} - 60)W_i$ for the second. Consideration of the efficiency of amplification in a sense that only an amount of a percent of DNA is amplified leads to the model

$$V_i = ((1+a)^{29}-29)U_i \text{ and} \\ X_i = ((1+a)^{30}-60)W_i$$

A single PCR setup is positive if $X_i > 2.37 \times 10^{10}$. An analysis of the effect of the amplification parameter a shows, under the condition that $U_i \geq 1$, that the probability of detection is one whenever $a > 0.584$. In other words, a single cDNA molecule is sufficient for a positive result after nested PCR if the efficiency of PCR $> 58.4\%$. We assume that $a > 0.584$; therefore, the probability of k detections in a sample is: $\Pr(k=0) = \Pr(U_1=0; U_2=0) = (1-2\pi)^R$; $\Pr(k=1) = \Pr(U_1 \geq 1; U_2=0) + \Pr(U_1=0; U_2 \geq 1) = 2 \times [(1-\pi)^R - (1-2\pi)^R]$; $\Pr(k=2) = 1 - \Pr(k=0) - \Pr(k=1)$. As PCR setups from different pB samples are independent, the probabilities for 0, 1, 2, 3, or 4 detections can easily be calculated combining the joint distributions of (U_1, U_2) for both probes.

$$\Pr(0 \text{ detections}) = \Pr(k=0) \times \Pr(k=0) \text{ etc.}$$

RESULTS

As calculated in our model, amplification of a single mammaglobin cDNA molecule leads to a positive result if the efficiency of PCR exceeds 58.4%. Accordingly, only distinct but never weak bands were detected on agarose gels after optimization of the PCR conditions (8). This reflects an adequate efficiency

Table 1. Calculated Probabilities for Achieving 0, 1, 2, 3, or 4 Positive PCR Setups

mRNA Number	cDNA Efficiency	Number of positive PCR setups				
		0	1	2	3	4
50	0.3	0.6181	0.3164	0.0603	0.0051	0.0002
50	0.4	0.5262	0.3671	0.0953	0.0109	0.0005
50	0.5	0.4479	0.3993	0.1324	0.0194	0.0011
50	0.6	0.3811	0.4169	0.1696	0.0304	0.0020
50	0.7	0.3242	0.4231	0.2053	0.0439	0.0035
50	0.8	0.2757	0.4207	0.2385	0.0595	0.0055
50	0.9	0.2345	0.4117	0.2685	0.0771	0.0082
50	1.0	0.1993	0.3979	0.2950	0.0962	0.0117
500	0.3	0.0081	0.0760	0.2655	0.4116	0.2388
500	0.4	0.0016	0.0260	0.1552	0.4107	0.4065
500	0.5	0.0003	0.0084	0.0815	0.3496	0.5602
500	0.6	0.0001	0.0026	0.0403	0.2720	0.6850
500	0.7	0.0000	0.0008	0.0192	0.2004	0.7795
500	0.8	0.0000	0.0002	0.0089	0.1427	0.8481
500	0.9	0.0000	0.0001	0.0041	0.0995	0.8964
500	1.0	0.0000	0.0000	0.0019	0.0683	0.9298

Probabilities depend on the number of mammaglobin mRNA molecules in the RNA preparation and the cDNA synthesis efficiency.

of PCR, permitting a clear evaluation of each PCR setup as positive or negative. Consequently, the amount of mammaglobin mRNA molecules in the RNA preparation and the efficiency of reverse transcription into cDNA are decisive for the number of positive PCR setups.

To clarify the relationship between the number of mammaglobin mRNA molecules in the RNA preparation and the outcome of the assay, the probabilities for achieving 0, 1, 2, 3, or 4 positive PCR setups were calculated for 50 and 500 mRNA molecules, depending on various cDNA synthesis efficiencies (Table 1). For 50 mRNA molecules and an efficiency of 100%, the probabilities for 1, 2, 3, or 4 positive setups are 39.79%, 29.5%, 9.62%, and 1.17%, respectively, but still 19.93% for a negative result. Reduction of the efficiency leads to an increase of the probabilities for one positive setup or a negative result. Interestingly, only efficiencies of cDNA synthesis below 50% lead to a higher probability for a negative than to any positive result.

On the other hand, for 500 mRNA

molecules and an efficiency of 100%, the probabilities for a negative or a 1 positive setup result are 0, but 0.19% for 2, 6.83% for 3, and 92.98% for 4 positive setups. Lower efficiencies of cDNA synthesis shift the probabilities mainly towards three positive setups; even with an efficiency of 30%, the probability for a negative result is just 0.81%.

Results of the mammaglobin-nested RT-PCR assay with pB samples are listed in Table 2: samples with one or two mammaglobin positive PCR setups were detected in 6/143 (4%) patients with benign lesions of the breast, in 41/310 (13%) BC patients with NED, and in 39/157 (25%) BC patients with MD. On the contrary, no sample from patients with benign lesions of the breast resulted in 3 or 4 positive PCR setups, but 5/310 (2%) BC patients with NED and 46/157 (29%) with MD.

The 217 BC patients with NED were followed for at least 12 months. Out of 41 patients with one or two mammaglobin-positive PCR setups, five relapsed 5–21 months afterwards. This rate of relapse is similar to patients

with mammaglobin-negative setups (12% vs. 13%). On the contrary, all five NED patients with at least three positive setups relapsed 1–13 months after first examination (Table 3).

DISCUSSION

The use of RT-PCR for the detection of circulating tumor cells in the pB of BC patients is a powerful technique because it is remarkably sensitive. It permits the analysis of gene expression from very small amounts of RNA even at the level of a single tumor cell. To extract reliable information from this method, consideration of the mathematical background is necessary (4).

Assuming that a single tumor cell contains only 10 mammaglobin molecules, a concentration of one cell per milliliter of pB leads to about 50 mRNA molecules prepared from 5 mL pB. This number of molecules is likely to result in one or two positive PCR setups if the efficiency of cDNA synthesis exceeds 50%. On the other hand, a 10-fold increase of the number of mRNA molecules leads to three or four positive setups, even at an efficiency of reverse transcription equal to 30%. Hence, the efficiency of cDNA synthesis seems to have only little effect on the result of the assay.

In addition, these calculations point to a theoretical detectability of as little as a single tumor cell per milliliter of pB. In dilution experiments of SKBR5 breast cancer cells in pB of healthy individuals, the calculated detection limit of the mammaglobin-nested RT-PCR assay was one tumor cell in 10^6 – 10^7 white blood cells (8), which corresponds to the theoretical detection limit.

Clinical results of pB samples tested via the mammaglobin-nested RT-PCR assay for the presence of tumor cells reflect the outcome of the stochastic model: the number of positive PCR setups corresponds with the stage of disease of BC patients. Samples from patients with MD are expected to have a higher incidence of tumor cells in pB than samples from patients with NED. Indeed, most samples with at least three positive PCR setups were derived from BC patients with MD (29%) but were never detected in patients with benign

lesions of the breast and only rarely in patients with NED (2%), whereby all of the latter relapsed within 13 months. On account of these results, we suggest that samples with at least three positive PCR setups are likely to contain tumor cells with metastatic potential; hence, only these samples are evaluated as tumor cell positive.

On the contrary, samples with one or

two positive PCR setups were detected in all clinical subgroups of BC patients and even in patients with benign lesions of the breast. This points to the existence of occasionally mammaglobin-expressing cells in pB being not necessarily related to tumor cells. However, samples with one positive PCR setup were more often detected in BC patients than in patients with benign lesions of

Molecular Diagnostic Techniques

Table 2. Numbers of Mammaglobin-Positive PCR Setups of pB Samples

Number of Positive PCR Setups	Benign	NED	MD
1	3 (2%)	30 (10%)	20 (13%)
2	3 (2%)	11 (3%)	19 (12%)
3	0	2 (1%)	14 (9%)
4	0	3 (1%)	32 (20%)
Number of Patients	143	310	157

Table 3. Relapse Rate of BC Patients with NED Followed for at least 12 Months

Mammaglobin Results	No. Patients	No. Relapsed Patients
4 negative setups	171	22 (13%)
1 or 2 positive setups	41	5 (12%)
3 or 4 positive setups	5	5 (100%)

the breast. In addition, samples with two positive setups were detected with higher frequency in BC patients with MD than in patients without metastases (Table 2). Therefore, samples with one or two positive PCR setups might contain a low number of tumor cells, but for clinical evaluation these samples are not considered as tumor cell positive.

The amount of mRNA molecules in the RNA preparation depends on the number of tumor cells in pB and/or the mammaglobin expression level, both of which may vary between BC patients or even during the course of disease in a single patient. The expression levels of most genes range between 5 and 15 mRNA molecules per cell (1), but mammaglobin mRNA levels are increased at least 10-fold relative to normal breast tissue in 23% of BC patients (7). Hence, samples containing tumor cells with a 10-fold overexpression of mammaglobin and a concentration of one cell per milliliter of pB would lead to about 500 mRNA molecules in the RNA preparation and therefore result most probably in three or four positive setups. In this respect, even small variations in the tumor cell count between samples can lead to differences in the number of positive PCR setups in the mammaglobin-nested RT-PCR assay. Therefore, a single blood draw of a BC

patient yields just a result of the situation at a specific time point; only examinations along the course of disease will lead to reliable results.

Mathematical considerations of nested RT-PCR assays are a substantial prerequisite for the correct assessment of results obtained from clinical samples. The results of pB samples tested for the presence of tumor cells via the mammaglobin-nested RT-PCR assay correspond with the stochastic model: the number of positive setups reflects the amount of tumor cells.

ACKNOWLEDGMENTS

The authors wish to thank Diane Thompson for critically reviewing the manuscript and Anneliese Kolb for excellent technical assistance.

REFERENCES

1. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. 1983. *Molecular Biology of the Cell*. Garland Publishing, New York.
2. Ausubel, F.M., R. Brent, R.E. Kingston, D.D. More, J.G. Seidman, J.A. Smith, and K. Struhl. 1994. *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
3. Connolly, A.R., L.G. Cleland, and B.W. Kirkham. 1995. Mathematical considerations

of competitive polymerase chain reaction. *J. Immunol. Methods* 187:201-211.

4. Freeman, W.M., S.J. Walker, and K.E. Vrana. 1999. Quantitative RT-PCR: pitfalls and potential. *BioTechniques* 26:112-125.
5. Jung, R., P. Ahmad-Nejad, M. Wimmer, M. Gerhard, C. Wagener, and M. Neumaier. 1997. Quality management and influential factors for the detection of single metastatic cancer cells by reverse transcriptase polymerase chain reaction. *Eur. J. Clin. Chem. Clin. Biochem.* 35:3-10.
6. Sambrook, J., I. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. 2nd ed. CSH Laboratory Press, Cold Spring Harbor, NY.
7. Watson, M.A. and T.P. Fleming. 1996. Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer. *Cancer Res.* 56:860-865.
8. Zach, O., H. Kasparu, O. Krieger, W. Hehenwarter, M. Girschikofsky, and D. Lutz. 1999. Detection of circulating mammary carcinoma cells in the peripheral blood of breast cancer patients via a nested reverse transcriptase polymerase chain reaction assay for mammaglobin mRNA. *J. Clin. Oncol.* 17:2015-2019.

Received 15 May 2001; accepted 5 October 2001.

Address correspondence to:

Dr. Otto Zach
Elisabethinen Hospital
1st Department of Internal Medicine
Fadingerstr. 1
A-4010 Linz
Austria
e-mail: otto.zach@elisabethinen.or.at

**For reprints of this or
any other article, contact
Reprints@BioTechniques.com**