Optimal Markers for Real-Time Quantitative Reverse Transcription PCR Detection of Circulating Tumor Cells from Melanoma, Breast, Colon, Esophageal, Head and Neck, and Lung Cancers

Liqiang Xi,1 Daniel G. Nicastri,1 Talal El-Hefnawy,2 Steven J. Hughes,2 James D. Luketich,2,3 and Tony E. Godfrey1*

Background: The detection of circulating tumor cells (CTCs) may prove useful for screening, prognostication, and monitoring of response to therapy. However, given the large background of circulating cells, it is probably necessary to detect 1 cancer cell in >10^6 leukocytes. Although reverse transcription (RT)-PCR is potentially sensitive and specific enough to achieve this goal, success will require the use of appropriate mRNA markers. The goal of this study was to identify optimal marker combinations for detection of CTCs.

Methods: An extensive literature and internet database survey was conducted to identify potential markers. We then used real-time quantitative RT-PCR to test for expression of selected potential markers in tissue samples from primary tumors of breast, colon, esophagus, head and neck, lung, and melanoma and normal blood samples. Markers with high expression in tumors and a median 1000-fold lower expression in normal blood were considered potentially useful for CTC detection and were tested further in an expanded sample set.

Results: A total of 52 potential markers were screened, and 3–8 potentially useful markers were identified for each tumor type. The mRNAs for all but 2 markers were found in normal blood. Marker combinations were identified for each tumor type that had a minimum 1000-fold higher expression in tumors than in normal blood.

Conclusions: Several mRNA markers may be useful for RT-PCR–based detection of CTCs from each of 6 cancer types. Quantification of these mRNAs is essential to distinguish normal expression in blood from that due to the presence of CTCs. Few markers provide adequate sensitivity individually, but combinations of markers may produce good sensitivity for detection of the presence of these 6 neoplasms.

© 2007 American Association for Clinical Chemistry

Despite surgical resection and adjuvant therapy, the path of many cancer patients unfortunately ends with distant metastatic disease. The prognosis of patients with distant metastases in cancers such as breast, colon, esophagus, head and neck, lung, and melanoma is extremely poor. Five-year survival probability for stage IV patients, for example in lung cancer, is typically <10%, although some reports suggest survival of 30%–35% for selected patients after complete metastectomy (1, 2). Presumably, distant metastases are the result of hematogenous dissemination of tumor cells, and in this context, it is assumed that stage IV patients, and some earlier-stage patients, have circulating tumor cells (CTCs) even before clinical manifestation of distant metastasis. Reproducible detection of these rare CTCs is extremely challenging but could potentially be useful for diagnosis, risk stratification, recurrence prediction, development of novel therapies, and treat-
ment monitoring (3, 4). Indeed, recent reports indicate that reproducible detection and quantification of CTCs in stage IV breast cancer patients is an independent prognostic factor that can be used to monitor therapeutic response and guide treatment options (4–8). Similar studies are now being performed for other tumor types (9), and CTC detection has the potential to become a standard part of cancer staging.

CTCs may represent only 1 cell among 10^6 peripheral blood mononuclear cells (10). In general, methods for their detection comprise the use of a CTC enrichment step and a detection step that may or may not be separate from the enrichment. Enrichment methods include both positive and negative selection using antibodies and either flow cytometry or immunomagnetic separation, density gradient centrifugation using Ficoll or modified gradients such as Oncoquick, and filtration using various types of membranes or porous devices. Detection is typically performed with epithelial or tissue-specific markers rather than cancer-specific markers and relies on the assumption that these markers should not be present and detectable in the blood unless CTCs are present. In many cases, however, this assumption is not valid, and most detection methods have some background levels in blood from individuals without cancer. The most common CTC detection methods are immunostaining (immunohistochemistry or immunofluorescence) against epithelial or tissue-specific antigens and reverse transcription (RT)-PCR to detect mRNA from the genes coding for these antigens. Immunostaining has the added benefit of using cell morphology to increase specificity but requires manual interpretation by trained individuals and is somewhat subjective. RT-PCR, on the other hand, is extremely sensitive and less subjective and can be automated, but its specificity is completely dependent on the cancer specificity of the marker that is used. Furthermore, because of ectopic expression of mRNAs (11), some background expression can be expected for any marker and therefore quantitative assays are essential to differentiate true CTC signals from pure background.

In this study, we screened potential mRNA markers for each of 6 major tumor types and identified both individual markers and marker combinations that may be useful for CTC detection in each tumor type.

Materials and Methods

Identification of Potential Markers

An extensive literature and public database survey was conducted to identify any potential markers relevant to breast, colon, esophageal, head and neck, and lung cancer and melanoma. Resources for this survey included PubMed, OMIM, UniGene (http://www.ncbi.nlm.nih.gov/), GeneCards (http://bioinfo.weizmann.ac.il/cards), and CGAP (http://cgap.nci.nih.gov). Our survey criteria were somewhat flexible, but the goal was to identify genes with moderate to high expression in cancer tissue and low expression in normal peripheral blood. In addition, genes reported to be upregulated in cancer and genes with restricted tissue distribution were considered potentially useful. Finally, we evaluated genes reported to be cancer-specific, such as the cancer testis antigens and hTERT.

Tissues and Pathological Evaluation

 Archived tumor specimens were obtained from tissue banks at the University of Pittsburgh Medical Center through protocols approved by the institutional review board. All samples were of primary tumor, not metastatic disease, and patients were predominantly those with early-stage disease for which surgery was performed with curative intent. A small piece of each primary tumor was snap frozen in liquid nitrogen and later embedded in OCT for frozen section. Twenty 5-micron sections were cut from each frozen tissue piece for RNA isolation. In addition, sections for RNA isolation were cut and placed on slides for hematoxylin and eosin analysis at the beginning, middle (between the 10th and 11th sections for RNA), and end of the sections. All 3 hematoxylin and eosin slides from each specimen underwent pathological review to confirm the presence and percentage of tumor and to identify any contaminating tissues.

Blood Sample Collection and Processing

Normal blood samples were obtained from individuals participating in a lung cancer screening program at the University of Pittsburgh Cancer Center. Only patients with negative screening results were selected as controls for this study. All individuals signed institutional review board approval forms and gave written consent for blood collection and use in research studies. For each individual, 10 mL of blood was drawn by venipuncture into Becton Dickinson Vacutainer Tubes containing EDTA anticoagulant. Blood samples were processed within 30 min by centrifugation at 2000 g for 15 min. Buffy coat was collected and washed once with 1× PBS (0.14 g/L NaH2PO4, 9 g/L NaCl, 0.8 g/L Na2 HPO4, pH 7.4) (Mediatech), and cells were pelleted by centrifugation and lysed with TRIzol Reagent (Invitrogen). The cell lysate was stored at −80 °C until RNA isolation (see methods below).

Marker Screening

The screening was conducted in 2 phases. All potential markers entered the primary screening phase and expression was analyzed in 6 primary tumors (at least 50% tumor determined by pathologic evaluation) and 5 normal blood samples (5 pooled RNAs). Markers with high expression in primary tumors and a median of 1000-fold lower expression in normal blood were considered potentially useful for CTC detection. In the secondary screening, we tested 3–8 potential markers from the primary screening for each tumor type with ~20–30 primary tumor samples and 19 blood samples from individual healthy donors. All tumors were from independent patients.
RNA ISOLATION AND cDNA SYNTHESIS

Tumor RNA was isolated using the RNaseasy minikit (Qiagen) essentially as described by the manufacturer. The only modification was that we doubled the volume of lysis reagent and loaded the column in 2 steps. This procedure was found to provide better RNA yield and purity. RNA extraction from blood buffy coat was performed with TRIzol Reagent and followed by a purification step with RNaseasy columns. All RNAs were DNase treated with the DNA-free Kit (Ambion), and RT was performed in 100 μL reaction volumes with random hexamer priming and Superscript II (Invitrogen) reverse transcriptase. For the primary screen, 2 RT reactions were performed, each with 1000 ng of RNA. The cDNAs were combined and quantitative PCR (QPCR) was performed with the equivalent of 20 ng RNA per reaction. For the secondary screen, the RNA input for primary tumors was 1200 ng per RT reaction and 20 ng per QPCR, but this was increased to 5000 ng per RT reaction and 160 ng per QPCR for the blood to improve sensitivity for detection of low background expression.

QPCR

All QPCR was performed on the ABI Prism 7700 Sequence Detection Instrument (Applied Biosystems). Relative expression of the marker genes was calculated with the delta-CT (threshold cycle) methods previously described. The expression of the marker genes was calculated with the Detection Instrument (Applied Biosystems). Relative expression was determined by testing primers for generation of a single band on gels by the use of cDNA templates and annealing temperatures of 60, 62, and 64 °C. Additional details describing our methods for primer design and testing have recently been published. PCR efficiency was estimated with SYBR Green quantification before use in the primary screen. Further optimization and more precise estimates of efficiency were performed with 5′ nuclease probes for all assays used in the secondary screen.

A mixture of the Universal Human Reference RNA (Stratagene) and RNAs from human placenta, thyroid, heart, colon, and PCI13 and SKBR3 cell lines served as a universal positive expression control for all the genes in the marker screening process. Tissue RNAs were batched for RT, and the positive control RNA was included in each batch. Batches of cDNAs were then analyzed by QPCR one gene at a time on 96-well plates. Thus, the positive control RNA acted as a control for both RT and QPCR of each batch. The positive control was not used for calculation of relative expression.

QUANTIFICATION WITH SYBR GREEN (PRIMARY SCREEN)

For SYBR Green I–based QPCR, each 50-μL reaction contained 5 μL 10× TaqMan Buffer A [500 mmol/L KCl, 100 mmol/L Tris-HCl, 0.1 mol/L EDTA, 600 mmol/L Passive Reference 1 (Applied Biosystems)], 300 nmol/L each dNTP, 3.5 mmol/L MgCl2, 0.06 units/μL AmpliTaq Gold (Applied Biosystems), 0.25× SYBR Green I (Molecular Probes), and 200 nmol/L each primer. The amplification program comprised 2 stages, with an initial 95 °C Taq activation stage for 12 min followed by 40 cycles of 95 °C denaturation for 15 s, 60, 62 or 64 °C anneal/extend for 60-s and 10-s data collection steps at a temperature 2–4°C below the Tm of the specific PCR product being amplified. After amplification, a melting curve analysis was performed by collecting fluorescence data while increasing the temperature from 60°C to 95°C over a 20-min period.

QUANTIFICATION WITH 5′ NUCLEASE PROBES (SECONDARY SCREEN)

Probe-based QPCR was performed as described previously. Briefly, reactions were performed with a probe concentration of 200 nmol/L and a 60-s anneal/extend phase at 60 °C for GUSB, CK19, CK20, EGFR, MGB1, PTHrP, SCCA, and SFTPB; 62 °C for CK7, MAGEA, MART1, and MGB2; and 64 °C for CEA, LUNX, PIP, TACSTD1, TM4SF3, and TYR. The sequences of primers and probes (purchased from IDT) for genes evaluated in the secondary screen are listed in Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol53/issue7. The primer sequences for markers used in the primary screen will be provided upon request.

DATA ANALYSIS

In the primary screen, data from the melting curve were analyzed with ABI Prism 7700 Dissociation Curve Analysis 1.0 software (Applied Biosystems). The first derivative of the melting curve was used to determine the product Tm as well as to establish the presence of the specific product in each sample. In general, samples were

5 Human genes: GUSB, β-glucuronidase; CK7, cytokeratin 7; CK19, cytokeratin 19; TACSTD1, tumor-associated calcium signal transducer 1; MGB1, mammaglobin 1; MGB2, mammaglobin 2; PIP, prolactin-inducible protein; CEA, carcinoembryonic antigen–related cell adhesion molecule 5; CK20, cytokeratin 20; TM4SF3, transmembrane 4 superfamily, member 3, tumor associated antigen; SCCA, squamous cell carcinoma antigen; PTHrP, parathyroid hormone–related protein; LUNX, lung-specific X protein; SFTPB, surfactant, pulmonary–associated protein B; MAGEA, melanoma antigen gene family A; MART1, melanoma antigen recognized by T cells 1; TYR, tyrosinase; EGFR, epidermal growth factor receptor.
Fig. 1. Marker expression profiles for primary tumor from 6 tumor types and normal blood in secondary screening.

Each ♦ indicates a gene expression value for an independent tumor sample, and each ○ indicates a gene expression value for an independent normal blood sample.
analyzed in duplicate QPCRs and the mean C_T value was used in the expression analysis. However, in the secondary screen, triplicate reactions were performed for each individual sample, and the lowest C_T value was used to calculate relative expression in normal blood and thus obtain the highest (most conservative) estimate of background expression for the sample. Fold changes of median expression value in tumors vs median expression values in blood, median and minimum expression values in tumors vs maximum expression value in blood, and percentage of tumor samples with minimum 1000-fold higher than maximum expression value in blood were calculated for each marker in these 6 cancer types.

**Results**

**Primary Screen**

Our literature and database surveys identified a total of 52 genes for evaluation in the primary screen. All of these genes were analyzed for expression in 6 primary tumors and 5 normal blood samples. The resulting data of relative expression for all genes is shown in Fig. 1 in the online Data Supplement. Median relative expression (relative to the endogenous control gene) in the primary tumors and in normal blood was calculated for each gene in the primary screen. In addition, we calculated the ratio of relative expression between the lowest expression in tumors and the highest expression in normal blood and between the median expression in tumors and the highest expression in normal blood. Some genes, such as most genes for testis cancer antigens, had no detectable expression in normal blood but also had very low expression or low frequency of expression in the primary tumors. As a result, these genes are unlikely to be sensitive markers for detection of CTCs. When we used median expression in the tumors as the numerator (median tumor/highest normal), 6 genes (CK7, CK19, TACSTD1, MGB1, MGB2, and PIP) for breast cancer, 5 genes (CEA, CK19, CK20, TACSTD1, and TM4SF3) for colon cancer, 6 genes (CEA, CK19, CK7, TACSTD1, CK20, and TM4SF3) for esophageal cancer, 4 genes (CEA, CK19, SCCA, and PTHrP) for head and neck cancer, 7 genes (CEA, CK7, CK19, LUNX, SCCA, SFTPB, and TACSTD1) for lung cancer, and 3 genes (MAGEA-plex, MART1, and TYR) for melanoma clearly stood out as having mean tumor:highest normal blood ratios ~1500–25,000. When the minimum difference between tumors and benign nodes was calculated as a ratio (lowest tumor:highest normal blood), most of these genes still had ratios >1000. Thus, these genes were selected for further analysis in the secondary screen. EGFR was inadvertently omitted from the primary screen but was added in the secondary screen.

**Secondary Screen**

In this stage of the screening procedure, pathologic evaluation of primary tumor specimens revealed a median
tumor percentage of 75% (range, 5%–95%) for 25 breast cancer samples, 50% (range, 20%–80%) for 28 colon cancer samples, 70% (range, 5%–100%) for 19 esophageal cancer samples, 40% (range, 5%–95%) for 26 head and neck cancer samples, 70% (range, 5%–95%) for 22 lung cancer samples, and 80% (range, 5%–95%) for 24 melanoma samples.

The relative expression profiles of all markers in the primary tumors of 6 tumor types and normal blood are shown in Fig. 1. In primary tumors, the expression of all genes was variable and was not significantly correlated with percentage of tumor (data not shown). With the exception of MGB2 and MAGEA-plex, low expression of all markers was also observed in normal blood. The median and minimum expression among the tumors, median and maximum expression among the blood, tumor median expression:blood median expression ratio, tumor median expression:blood maximum expression ratio, and tumor minimum expression:blood maximum expression were calculated and are shown in Table 1. By looking at 2 criteria, the percentage of tumor samples with a minimum expression 1000-fold above the highest expression in normal blood and the minimum expression difference (minimum tumor expression:blood maximum expression) for each marker and tumor type, we easily identified the best single markers. They were MGB2 for breast cancer (96% of 25 tumors had >1000-fold expression above highest expression in normal; minimum expression ratio was 6.19 × 10^5), TM4SF3 for colon cancer (61.4% of 28 tumors; 1.57 × 10^5), CK19 for esophageal cancer (94.7% of 19 tumors; 6.32 × 10^5), EGFR for head and neck cancer (100% of 26 tumors; 1.63 × 10^5), SFTPB for lung cancer (86.6% of 22 tumors; 9.54 × 10^5), and TYR for melanoma (87.5% of 24 tumors; 9.75 × 10^5).

**Multimarker Analysis**

From the data presented, it appears that single marker assays may be capable of producing highly sensitive analysis of blood samples for CTCs. However, because there was both variable expression in tumors and variable background expression in normal blood even in these relatively small data sets, it is reasonable to consider that a single-gene assay may lack adequate sensitivity for CTC detection in larger patient sets. As such, multimarker analyses were performed to identify complementary markers that, when used in combination, would result in 100% of the tumor samples having expression >1000-fold above the highest observed expression in normal blood. To identify such combinations, we calculated the ratio of each marker expression over the highest expression of this marker among all 19 normal blood samples for each specific tumor sample and for all tumors. These data are graphically represented for each tumor type to facilitate visualization of which marker(s) provide the highest fold changes with the highest frequency (see Fig. 2 in the online Data Supplement). For example, in breast cancer, MGB2 provided the highest fold change in 19 of 25 tumors, followed by CK19 (2 of 25), CK7 (2 of 25), EGFR (1 of 25), and PIP (1 of 25). In a multimarker assay, the best 2 markers for breast cancer would be MGB2 and CK7 or CK19 (although CK19 may not be the best choice because the presence of multiple pseudogenes precludes cDNA specificity) and the best 3 would be MGB2, CK7, and EGFR. Combined, these 3 markers give a minimum fold change of 2300 in this data set (Table 1 and Fig. 2). Similarly, the best 2- to 4-marker combination for the other cancer types would be TM4SF3 and TACSTD1 for colon cancer (100% of 28 tumors >1000-fold; minimum, 1.23 × 10^5), TM4SF3 and CK7 for esophageal cancer (95.9% of 19 tumors; 4.64 × 10^5), PTHrP, EGFR, and SCCA for head and neck cancer (100% of 26 tumors; 3.28 × 10^5), CK7, EGFR, SCCA, and SFTPB for lung cancer (100% of 22 tumors; 2.26 × 10^5), and MAGEA-plex, MART1, and TYR for melanoma (100% of 24 tumors; 3.27 × 10^5).

**Discussion**

RT-PCR has been widely used for the detection of CTCs in the peripheral blood in a variety of cancer types (17–21). Even among reports limited to a specific tumor type, however, the sensitivity and specificity for detecting CTCs is extremely variable (22, 23), possibly reflecting significant differences in assay design, choice of markers, and enrichment methods. One of the key factors is certainly the choice of markers for RT-PCR detection of these CTCs. We have used neoplastic breast, colon, head and neck, esophageal, and lung cancer and melanoma samples along with blood samples from individuals without cancer to identify optimal markers and marker combinations for CTC detection by RT-PCR in each of these 6 tumor types. Primary screening of 52 markers identified sets of 3–8 markers for each tumor type that showed potential for CTC detection. These markers were then evaluated in larger sample sets.

Two findings are immediately clear from analysis of these data. First, considerable heterogeneity exists in the expression levels of all markers detected in the primary tumors, and expression level did not correlate with tumor cell percentage in this sample set. This result indicates that individual markers are not likely to be robust in large patient sets and that efforts to use quantitative RT-PCR (qRT-PCR) to specifically enumerate CTCs (quantify number of CTC per unit volume) in blood of cancer patients will probably not be successful. qRT-PCR estimates of disease burden may prove informative for tracking changes on an individual patient basis, but comparisons between patients may not be valid. Second, all but 2 of the markers in the secondary screen (MGB2 and MAGEA-plex) were expressed at low but detectable levels in blood from normal individuals, indicating that quantification is essential for qRT-PCR–based detection of CTCs. Furthermore, this background level of expression, along with the expression level of the marker in tumors, defines the potential sensitivity of any given marker for CTC detection and can be used to estimate the physical CTC...
enrichment that is required before qRT-PCR–based detection.

Our study has identified several novel markers for detection of CTCs, including TM4SF3 in colon cancer and esophageal cancer, PTHrP in head and neck cancer, and SFTP8 in lung cancer. TM4SF3 (Transmembrane 4 superfamily, member 3, tumor associated antigen) is also known as Tetraspanin 8 (TSPAN8) in GenBank. This gene codes for the cell surface glycoprotein defined by the monoclonal antibody CO-029, which is a 27- to 34-kDa membrane protein expressed in gastric, colon, rectal, and pancreatic carcinomas but not in most normal tissues.
Fig. 2. Marker expression differences between primary tumor and normal blood for best combination markers.
The best combination of 3 or 4 markers was identified in an attempt to obtain a minimum fold difference \(>1000\) for 100% of tumor samples. Each panel shows fold change of expression value for each primary tumor sample calculated with the highest detected expression value in normal blood samples as the denominator. Different symbols represent the different markers (listed in the key to the right of each panel), and the histogram column extends to the marker with the highest fold change for that particular sample.
according to the original report (24). PTHrP (Parathyroid hormone-related protein), also known as PTHHLH and HHM, regulates endochondral bone development and epithelial-mesenchymal interactions. SFTPB (Surfactant protein B) is one of a family of surfactant molecules secreted by type II alveolar cells, and as such it is a particularly strong marker for the adenocarcinoma subtype of non–small cell lung cancer. We have previously reported usefulness of both genes for detecting tumor cells in lymph nodes from patients with head and neck cancer and lung cancer (25,26).

We found that cytokeratin 19 (CK19) seems to be a very good marker for CTC detection in several tumor types, although this finding is not novel. Unfortunately, this gene has at least 2 pseudogenes that make it difficult to develop a cDNA-specific assay; therefore, the utility of this marker may be limited in clinical settings, given the potential for false positive results. Recently Balducci et al. (27) reported that they successfully designed cDNA-specific primers and improved specificity of CK19 mRNA detection by RT-PCR. It may be important to develop cDNA-specific CK19 assays for use in esophageal cancer, because we have found that the alternative combination of CK7 and TM4SF3 is not superior to CK19 alone.

In summary, our data demonstrate that gene expression differs between tumor type and individual tumors of the same type and that most epithelial or cancer-related markers have low expression in normal circulating blood cells. We have identified marker combinations that maximize differences in expression in tumors compared with cells isolated from normal blood. When performed in conjunction with appropriate enrichment, qRT-PCR for combinations of appropriate, tumor-type–specific markers should increase sensitivity for CTC detection in cancer patients. As we enter the era of targeted therapeutics, future studies should be aimed at identifying markers or characteristics of CTCs to determine tumorigenic capacity of the cells and to guide appropriate treatment selection.

Grant/funding support: This work was supported by a Cooperative Research and Development Agreement between Cepheid (Sunnyvale, CA) and the University of Pittsburgh and by National Institutes of Health Grant CA90665.

Financial disclosures: S.J.H. and T.E.G. own Cepheid publicly traded stock, value <$10,000. Both authors have also received speaker’s honoraria at several industrial workshop presentations.

Acknowledgments: We thank the following pathologists for reviewing tumor specimens: Drs. Sydney D. Finkelstein (esophageal tumors), Kenneth McCarty (breast, colon tumors, and melanoma), Jennifer L. Hunt (head and neck tumors), and Samuel A. Yousem (lung tumors).
References


