Short Communication

Molecular Quality of Exfoliated Cervical Cells: Implications for Molecular Epidemiology and Biomarker Discovery

Angelique H. Habis,[1] Suzanne D. Vernon,[1] Daisy R. Lee,[1] Mukesh Verma,[2] and Elizabeth R. Unger[1]

[1]Centers for Disease Control and Prevention, U.S. Public Health Service, U.S. Department of Health and Human Services, Atlanta, GA and [2]Cancer Biomarkers Research Group, National Cancer Institute, NIH, Bethesda, MD

Abstract

The biologic sample collected in molecular epidemiology studies must accurately reflect the disease being studied and have sufficient molecular quality for the intended assays. Noninvasive sampling methods, such as scrapes or brushes, are increasingly used. In this study, we evaluate the impact of sample collection media and extraction methods on the quality and yield of RNA from routine exfoliated cervical cytology. Excess cellular material remaining on the cytologic collection device after preparation of the routine screening Papanicolaou smear was placed in a variety of collection media and extracted using two commercial kits. The collection media had the largest impact on the yield and quality of RNA as evaluated by denaturing agarose gel electrophoresis and image analysis. Two collection media, PAXgene and PreservCyt, yielded RNA from most samples. The RNA showed some degree of degradation as evidenced by the reduced size of the higher molecular weight ribosomal band. However, with a sensitive gold particle-based detection method, reproducible microarray results were obtained using this RNA. (Cancer Epidemiol Biomarkers Prev 2004; 13(3):492–496)

Introduction

Molecular epidemiologic studies require careful attention to biologic sampling as well as population sampling. The biologic sample must accurately reflect the disease being studied and have sufficient molecular quality for the intended assays. Human materials available for epidemiologic research are either excess diagnostic samples or samples collected using relatively noninvasive methods. Noninvasive sampling methods are increasingly used because they can be uniformly applied to a population, whereas diagnostic collections have limited representation of healthy controls. Peripheral blood, urine, saliva, stool, and exfoliated cells obtained by scraping or brushing are examples of noninvasive samples used in molecular epidemiology. These samples are also ideal for screening assays; therefore, biomarker discovery and validation studies increasingly rely on the same types of material.

Experience with Papanicolaou (Pap) smear screening demonstrates that exfoliated cervical cells reflect cervical neoplasia. The molecular quality of this sample is clearly adequate for DNA-based molecular testing, such as PCR and hybrid capture. However, the use of exfoliated cervical cells in RNA-based gene expression techniques required for the discovery and validation of early cancer detection markers is unclear. In this study, we evaluate the impact of sample collection media and extraction methods on the quality and yield of RNA from routine exfoliated cervical cytology.

Materials and Methods

Cervical Specimens and Collection Methods. Anonymous specimens were collected from excess cellular material remaining on cytologic collection devices (Papette broom; Wallach Surgical Devices, Orange, CT) after preparation of routine screening Pap smears. Cells were dislodged into collection media, using vigorous agitation, and stored at ambient temperature for up to 24 h prior to transportation to the laboratory. At the laboratory, samples were stored at 4°C for 24–48 h before extraction.

The collection media examined included five different tissue culture media [DMEM, F-12, AIM-V, keratinocyte serum-free media (Invitrogen, Carlsbad, CA), and organ transport medium (Mediatech, Inc., Herndon, VA)], two

Received 7/7/03; revised 11/1/03; accepted 11/11/03.

Grant support: Supported in part by the National Cancer Institute's Early Detection Research Network, Interagency Agreement Y1-CN-0101-01; American Society for Microbiology and National Center for Infectious Diseases postdoctoral fellowship (A. H. H.).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Elizabeth R. Unger, Viral Exanthems and Herpesvirus Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G-41, Atlanta, GA 30333. Phone: (404) 639-3533; Fax: (404) 639-3540. E-mail: eru0@cdc.gov

alcohol-based collection fluids [95% ethanol and PreservCyt solution for Thin Prep cytology (Cytyc Corp., Boston, MA)], two high-salt solutions [RNA*later* (Ambion, Inc., Austin, TX) and MTM (a gift of MTM Laboratories AG, Heidelberg, Germany)], and a cationic detergent-based surfactant (PAXgene stabilizer; a gift of PreAnalytiX, Franklin Lakes, NJ). The volume was maintained at 6–20 ml, depending on the collection media, and one-fourth of the total volume was used for each RNA extraction. Results from 378 aliquots from 288 women are reported.

Nucleic Acid Extraction. For samples collected in PAXgene, cells were pelleted with centrifugation at 10,000 × *g* for 6 min. After the supernatant fluid was discarded, the cell pellets were washed by resuspension in 1 ml of dimethylpyrocarbonate (DMPC)-treated water, and then repelleted by centrifugation at 10,000 × *g* for 6 min. For all other media, cells were pelleted by centrifugation at 2000 × *g* for 30 min and used after removal of supernatant. All cell pellets were disrupted by vortexing in the lysis media appropriate for each extraction method.

A variety of extraction methods were successful in preliminary experiments, and we chose two methods amenable to processing multiple small samples for more detailed evaluation. The first method, MasterPure Complete DNA and RNA Purification Method (Epicentre Technologies, Inc., Madison, WI), uses a rapid desalting process to purify total nucleic acid (TNA; *i.e.*, DNA and RNA). For this method, the manufacturer's protocol for lysis was modified by increasing the proteinase K concentration to 200 μg/sample and by modifying the lysis conditions to 60°C for 50 min. The remainder of the protocol was unaltered. The TNA extracts were resuspended in 25 μl of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA. The second method, RNA*wiz* (Ambion), uses a proprietary reagent containing denaturants and RNase inhibitors to isolate total RNA. The manufacturer's protocol was used without modification. The resulting RNA was resuspended in 11 μl of DMPC-treated water. All TNA and RNA extracts were stored at −70°C.

RNA Analysis. RNA was evaluated with UV spectrophotometry and denaturing agarose gel electrophoresis (6.5% formaldehyde and 1% agarose). Samples were scored positive for RNA if ribosomal bands were visible on the ethidium bromide-stained gel, and they were scored positive for DNA if a high molecular weight DNA band was present. Total RNA yield was quantified by densitometric assessment of both ribosomal bands using the FlourChem Digital Imaging System (Alpha Inotek, Inc., San Leandro, CA) with comparison with a known amount of standard 28S and 18S RNA control marker.

Probe Preparation and Hybridization to Microarrays. The TNA from MasterPure extracts of two to three samples collected in PreservCyt or PAXgene were pooled to yield sufficient material for three technical replicates. For each probe synthesis, 500 ng of total RNA were washed with DMPC-treated water, concentrated to 10 μl using Microcon-100 filters (Millipore, Bedford, MA), and then treated for 30 min at 37°C with 1 unit of DNase I (GenHunter Corporation, Nashville, TN). Biotin was incorporated during cDNA synthesis by adding 1 μl of biotin-11-dUTP (1 mM; Enzo Diagnostics, Inc., Farmingdale, NY) to 1 μl of nucleotide mix (100 mM dGTP, dATP, and dCTP and 10 mM dTTP) during the reverse transcription (RT) reaction [3 μg of random hexamer primers and 200 units of Superscript-II (Invitrogen) following the manufacturer's protocol]. Unincorporated nucleotides were removed using QIAquick PCR purification column (Qiagen, Inc., Valencia, CA), and all of the biotinylated cDNA was used for hybridization to Atlas Human 3.8I oligonucleotide glass microarrays (Clontech Laboratories, Inc., Palo Alto, CA).

We used the automated Ventana Discovery System (Ventana Medical Systems, Tucson, AZ) with a 12-h, 42°C hybridization and three 10-min high-stringency washes (0.1× SSC, 42°C). Anti-biotin antibodies coated with resonance light scattering (RLS) gold particles (a gift of Genicon Sciences Corporation, San Diego, CA) were used for detection. The microarrays were prepared for imaging and storage by dipping in archiving solution (a gift of Genicon Sciences) and scanned using the GSD-501 scanner (Genicon Sciences) at 10-μm resolution. The TIFF images were analyzed with GenePix Pro version 4.0 software (Axon Instruments, Union City, CA), and raw median feature and mean local background intensities were exported to the Excel (Microsoft Corp., Redmond, WA) spreadsheet program for analysis.

Microarray Data Analysis. The cutoff value (COV) for feature intensities on each array was calculated as the average + 2 SD of each local background (1). The array's average signal and background were calculated as the average of all feature intensities above COV and all local backgrounds, respectively. Values below COV were set to 0.01. The feature intensities were normalized to a maximal value of 100 and centralized by setting the 75th percentile to 1. Scatter plots of $\log_2$ intensities were evaluated using the Pearson correlation coefficient to measure the linear relationship and the Lin concordance coefficient to measure the level of agreement in relation to the identity line (2).

Results

The impact of collection media and extraction methods on successfully obtaining RNA from exfoliated cervical cells is summarized in Table 1. The collection media had the largest impact on the yield and quality of RNA. With both extraction methods, the yield of RNA and the proportion of samples yielding ribosomal bands was highest for those collected in PAXgene followed by PreservCyt. By contrast, no sample collected in RNA*later* or 95% ethanol yielded RNA. Because the five tissue culture media performed similarly, aggregate results are presented in Table 1.

The most notable differences between the two extraction methods were that most MasterPure extracts included DNA, and RNA*wiz* was not compatible with the high-salt MTM media. The RNA yield was generally higher for MasterPure extraction than for RNA*wiz*. The A_{260}/A_{280} ratio, often used as a measure of RNA purity, averaged 1.8 and higher for each combination of collection media and extraction but showed a wide

Table 1. RNA from exfoliated cervical cells: impact of collection media and extraction

Collection medium	Extraction (n[c])	No. (%) with ribosomal bands	Average yield[a] (μg) [range]	Average A_{260}/A_{280} ratio[b] [range]
PAXgene stabilizer	MP (43)	42 (98%)	7.2 [0.3–17.2]	1.9 [1.6–2.2]
	Rw (41)	37 (90%)	4.8 [0.2–1.7]	1.8 [1.5–2.1]
PreservCyt	MP (125)	63 (50%)	4.8 [0.5–17.2]	1.8 [1.4–1.9]
	Rw (23)	16 (70%)	1.3 [0.2–6.2]	1.9 [1.6–2.2]
Tissue culture media[d]	MP (73)	33 (45%)	2.2 [0.1–8.8]	2.0 [1.8–2.4]
	Rw (27)	12 (44%)	1.6 [0.1–2.9]	2.1 [1.8–2.3]
MTM high salt	MP (16)	8 (50%)	2.1 [0.7–4.8]	1.9 [1.7–2.0]
	Rw (7)	0	0	ND
RNA*later*	MP (12)	0	0	ND
95% Ethanol	MP (11)	0	0	ND

Note: MP, Masterpure; Rw, RNA*wiz*; ND, not determined.
[a]Based on densitometric assessment of ribosomal bands, calculated to represent the whole starting sample.
[b]For samples with ribosomal bands.
[c]The number of aliquots tested by each combination of collection media and extraction varied because combinations that were not effective were eliminated earlier than those that appeared promising.
[d]Includes results for DMEM, F12, AIM-V, keratinocyte serum-free media, and organ transport medium.

range (1.4–2.4), and absorbance did not correlate with RNA concentration as assessed by gel analysis.

Figure 1 illustrates gels representative of the full range of RNA preservation and yield for MasterPure and RNA*wiz* extracts. Some of the variability in results is attributable to variation in the cell number, composition, and preservation that occurs in cervical sampling. Samples yielding a smear without discernable bands were scored as unsuccessful RNA extractions because of extensive degradation. While samples with any visible ribosomal bands were scored as yielding RNA, the variation in the size and intensity of these bands indicates a range of preservation. Compared with the intact control RNA (Fig. 1, lane M), all extracts showed some degradation. A small amount of residual intact 28S RNA was identified in most samples collected with PAXgene, but it was only rarely identified in samples collected in the other media. Rather than being entirely eliminated, in most samples, the 28S RNA band was truncated to a lower molecular weight. DNA was clearly visible in most MasterPure extracts but largely absent from RNA*wiz* extracts.

Three technical replicates of cDNA preparations from pooled extracts of cells collected with PreservCyt and PAXgene were hybridized to glass microarrays. The gel of the pooled extracts (Fig. 2A) shows that both samples had high molecular weight DNA. The PAXgene extract had residual intact 28S rRNA, whereas the rRNA in the PreservCyt extract pool was truncated. The average raw signal intensity of PreservCyt arrays was about 3-fold higher than that for PAXgene arrays (Fig. 2B). The PreservCyt arrays also had lower background, 5–7% *versus* 11–14% for PAXgene. As shown in the scatter

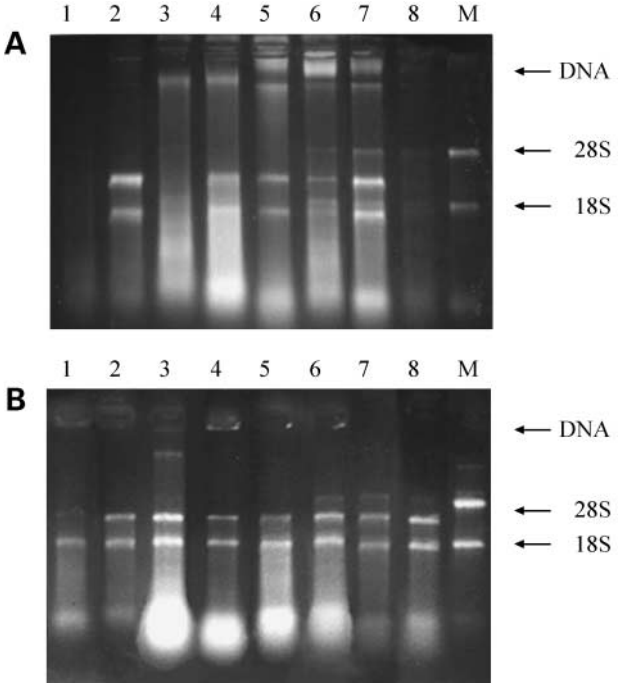

Fig. 1. Ethidium bromide-stained denaturing agarose gel electrophoresis of extracts of exfoliated cervical cells. **A.** MasterPure extracts. *Lanes 1* and *2,* tissue culture media; *lanes 3* and *4,* PreservCyt; *lanes 5* and *6,* high-salt solution; *lanes 7* and *8,* PAXgene stabilizer; *M,* marker. **B.** RNA*wiz* extracts. *Lanes 1* and *2,* tissue culture media; *lanes 3*, *4*, and *5,* PreservCyt; *lanes 6*, *7*, and *8,* PAXgene stabilizer; *M,* marker.

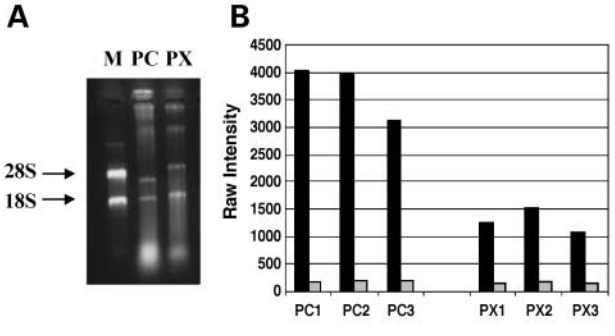

Fig. 2. Impact of collection media on microarray hybridization. **A.** Ethidium bromide-stained denaturing agarose gel of pooled MasterPure extracts used for microarray assays. *M,* marker; *PC,* PreservCyt; *PX,* PAXgene. **B.** Average raw intensity (■) and background (□) of replicate Atlas Human 3.8I microarrays prepared from MasterPure extracts of exfoliated cervical cells collected in PreservCyt (*PC1*, *PC2*, and *PC3*) and PAXgene (*PX1*, *PX2*, and *PX3*).

plots in Fig. 3, replicate arrays gave reproducible results, with PreservCyt arrays showing a slightly tighter distribution. The Pearson and Lin coefficients for pairwise comparisons were 0.95–0.96 and 0.94–0.95, respectively, for PreservCyt and 0.88–0.92 and 0.88–0.92, respectively, for PAXgene arrays.

Discussion

To our knowledge, this report is the first to document the yield and quality of RNA obtained from exfoliated cervical cells collected and processed with methods amenable to molecular epidemiologic studies and to demonstrate use of this RNA for microarray analysis. An early report describing the vision of a ''molecular Pap test'' (3) performed RT-PCR assays on 11 samples, but like reports from other laboratories that relied on RT-PCR assays of cervical material (4, 5), it did not provide data on yield or RNA quality. To apply quantitative RT-PCR to cervical cytology, Lamarq *et al.* (6) explored several options for collection and extraction. Based on results from a few samples, they selected Trizol reagent (Life Technologies, Inc.) for collection and extraction. The Trizol suspensions were immediately snap frozen in liquid nitrogen, an approach presenting difficulty for smooth interfacing in a clinical environment.

Two collection media, PAXgene and PreservCyt, combined with two different extraction methods yielded RNA from most samples. However, this RNA showed some degree of degradation as evidenced by the reduced size of the higher molecular weight ribosomal band and the presence of low molecular weight RNA. This degradation, which was less than that shown in the method adopted by Lamarcq *et al.* (6), could be attributed to autolysis that occurs during normal epithelial cell maturation (7) as well as to changes that occur after exfoliation (8). Our prior experience with tissue culture cells fixed in PreservCyt clearly demonstrated that intact RNA with good preservation of ribosomal bands can be obtained from material stored for up to 1 year prior to extraction (9).

Prior studies have demonstrated that modifications to the design of RT-PCR assays allow the reliable use of degraded RNA (6, 10). However, it was not clear that the cervical RNA would give reproducible results in microarray analysis. We introduced modifications in probe labeling and detection allowing microarray profiling of 500 ng of exfoliated cervical RNA. Random hexamer priming of RT during cDNA synthesis allowed fragmented RNA without polyadenylate tails to be labeled. Detection with scattered light from the RLS gold particles is 10^4 to 10^6 times greater than that by fluorescence (11, 12). The RLS signal does not fade; the microarrays become a permanent record. The reproducibility of the microarray results from our report, as evidenced by the Pearson and Lin coefficients for technical replicates (0.88–0.96 and 0.88–0.95, respectively), is comparable with those we observed with optimally prepared RNA from peripheral blood (13).

Preservation of exfoliated cervical cells was not the intended use of PAXgene, but our results suggest that the reagent could be adapted for this purpose. RNA was consistently better preserved in PAXgene than

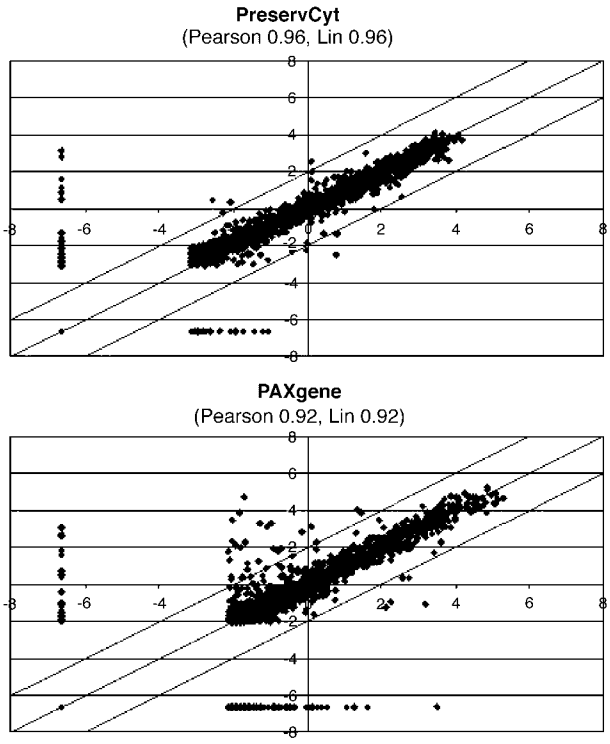

Fig. 3. Reproducibility of microarray results. Technical replicates of cDNAs prepared from cervical cytology samples collected in PreservCyt (*above*) or PAXgene (*below*) media. Scatter plot of the $\log_2$ intensity of each gene, with the line of identity in the *center*. *Outer lines,* 4-fold difference in intensity.

PreservCyt samples; however, the reproducibility of microarray results was slightly less. The labeling efficiency of PAXgene extracts may possibly be reduced due to carryover of inhibitory materials, and further optimization may improve results.

PreservCyt is a methanol-based fixative that is widely used in clinical settings and has the advantage of preserving cellular morphology so that cytology and immunocytochemistry can be performed on the same sample. We are currently using PreservCyt collection media and MasterPure extraction in an ongoing epidemiologic study of cervical neoplasia and have been able to obtain RNA with at least partial preservation of ribosomal bands in 70% of over 900 samples collected to date. The improved success rate may reflect the improved cellular yield that can be anticipated from premenopausal women with dysplasia.

Acknowledgments

We thank Jane Wamsher and Denise Marshall for collection of the cervical samples and Irina Dimulescu, Naovarath Taysavang, and Brian Allen for technical contributions.

References

1. Hedge P, Qi R, Abernathy K, et al. A concise guide to cDNA microarray analysis. Biotechniques, 2000;29:548–62.
2. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 1989;45:255–68.
3. Lin WM, Ashfaq R, Michalopulos EA, Maitra A, Gazdar AF, Muller CY. Molecular Papanicolaou tests in the twenty-first century:

molecular analyses with fluid-based Papanicolaou technology. Am J Obstet Gynecol, 2000;183:39–45.
4. Klaes R, Woerner SM, Ridder R, et al. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res, 1999;59:6132–6.
5. Sotlar K, Selinka HC, Menton M, Kandolf R, Bultmann B. Detection of human papillomavirus type 16 E6/E7 oncogene transcripts in dysplastic and nondysplastic cervical scrapes by nested RT-PCR. Gynecol Oncol, 1998;69:114–21.
6. Lamarcq L, Deeds J, Ginzinger D, Perry J, Padmanabha S, Smith-McCune K. Measurements of human papillomavirus transcripts by real time quantitative reverse transcription-polymerase chain reaction in samples collected for cervical cancer screening. J Mol Diagn, 2002;4:97–102.
7. Bhattacharya PK, Pappelis AJ. Changes in nucleic acid and protein content in nuclei of human cervical cells. Mech Ageing Dev, 1984;27:135–42.
8. Klaassen I, Copper MP, Brakenhoff RH, Smeets SJ, Snow GB, Braakhuis BJ. Exfoliated oral cell messenger RNA: suitability for biomarker studies. Cancer Epidemiol Biomarkers & Prev, 1998;7:469–72.
9. Tarkowski TA, Rajeevan MS, Lee DR, Unger ER. Improved detection of viral RNA isolated from liquid-based cytology samples. Mol Diagn, 2001;6:125–30.
10. Tong D, Schneeberger C, Leodolter S, Zeillinger R. Quantitative determination of gene expression by competitive reverse transcription-polymerase chain reaction in degraded RNA samples. Anal Biochem, 1997;251:173–7.
11. Yguerabide J, Yguerabide EE. Light-scattering submicroscopic particles as highly fluorescent analogs and their use as tracer labels in clinical and biological applications. I. Theory. Anal Biochem, 1998;262:137–56.
12. Yguerabide J, Yguerabide EE. Light-scattering submicroscopic particles as highly fluorescent analogs and their use as tracer labels in clinical and biological applications. II. Experimental characterization. Anal Biochem, 1998;262:157–76.
13. Ojaniemi H, Evengard B, Lee DR, Unger ER, Vernon SD. Impact of RNA extraction from limited samples on microarray results. Biotechniques, 2003;35:968–73.