MDPI

Article

Performance of Salivary Extracellular RNA Biomarker Panels for Gastric Cancer Differs between Distinct Populations

Karolina Elżbieta Kaczor-Urbanowicz [1,2,3,4], Mustafa Saad [1], Tristan R. Grogan [5], Feng Li [1,4], You Jeong Heo [1], David Elashoff [5], Robert S. Bresalier [6,*], David T. W. Wong [1,4,7,*] and Yong Kim [1,4,6,*]

1   Center for Oral and Head/Neck Oncology Research, School of Dentistry, University of California at Los Angeles, Los Angeles, CA 90095, USA; kkaczor@ucla.edu (K.E.K.-U.); mzsaad@g.ucla.edu (M.S.); fengli@ucla.edu (F.L.); heo893@naver.com (Y.J.H.)
2   UCLA Institute for Quantitative and Computational Biosciences, University of California at Los Angeles, Los Angeles, CA 90095, USA
3   UCLA Section of Orthodontics, University of California at Los Angeles, Los Angeles, CA 90095, USA
4   Section of Biosystems and Function, School of Dentistry, University of California at Los Angeles, Los Angeles, CA 90095, USA
5   Department of Medicine Statistics Core, David Geffen School of Medicine at the University of California at Los Angeles, Los Angeles, CA 90024, USA; tgrogan@mednet.ucla.edu (T.R.G.); DElashoff@mednet.ucla.edu (D.E.)
6   Department of Gastroenterology, Hepatology and Nutrition, Division of Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
7   UCLA's Jonsson Comprehensive Cancer Center, Los Angeles, CA 90024, USA
*   Correspondence: rbresali@mdanderson.org (R.S.B.); dtww@ucla.edu (D.T.W.W.); thadyk@ucla.edu (Y.K.)

Simple Summary: Gastric cancer (GC) is the fourth most common cancer that occurs worldwide, affecting specifically the Asian population. Currently, there are no available screening programs for GC in United States. Since saliva is a highly desirable body fluid for developing biomarkers of cancer screening, early detection, and monitoring, we previously reported that salivary extracellular RNAs could be developed to detect gastric cancer in a Korean cohort, and here, we validate them in a U.S. cohort. Our study emphasizes the importance of population-specific biomarker development and validation, and specifically, the noninvasive nature of salivary biomarkers for population-based screening in at-risk populations.

Abstract: Gastric cancer (GC) has the fifth highest incidence among cancers and is the fourth leading cause of cancer-related death GC has predominantly a higher number of cases in certain ethnic groups such as the Korean population. GC found at an early stage is more treatable and has a higher survival rate as compared with GC found at a late stage. However, a diagnosis of GC is often delayed due to the lack of early symptoms and available screening programs in United States. Extracellular RNA (exRNA) is an emerging paradigm; exRNAs have the potential to serve as biomarkers in panels aimed at early detection of cancer. We previously reported the successful use of a panel of salivary exRNA for detecting GC in a high-prevalence Korean cohort, and that genetic changes reflected cancer-associated salivary exRNA changes. The current study is a case-control study of salivary exRNA biomarkers for detecting GC in an ethnically distinct U.S. cohort. A model constructed for the U.S. cohort combined demographic characteristics and salivary miRNA and mRNA biomarkers for GC and yielded an area under the receiver operating characteristic (ROC) curve (AUC) of 0.78. However, the constituents of this model differed from that constructed for the Korean cohort, thus, emphasizing the importance of population-specific biomarker development and validation.

Keywords: biomarkers; validation; gastric cancer; exRNA; saliva

1. Introduction

Gastric cancer (GC) is an aggressive type of cancer that remains a healthcare burden worldwide [1]. For 2022, the American Cancer Society has estimated that there will be about

26,380 new cases of gastric cancer in the USA and about 11,090 deaths (https://seer.cancer.gov/statfacts/html/stomach.html (accessed on 10 May 2022)). Gastric cancer accounts for ~1.5% of all new cases of cancers in the USA each year. The incidence of GC in the United States has relatively decreased, however, in Western countries, approximately half of patients present with locally advanced or metastatic GC at diagnosis, and an additional 40% to 60% of those patients undergoing resection of gastric adenocarcinoma relapse after surgery [2]. Thus, early detection of this type of cancer is the main goal to reduce mortality, as the 5-year survival rate of early detected cases can reach >95% [3].

Many studies from East Asian countries have shown that screening methods, especially endoscopic screening for detecting early-stage GC, have resulted in a reduction in mortality. However, population-based screening programs do not exist in the USA, because of the low incidence of GC overall [4].

Although upper gastroesophageal endoscopy with targeted and random biopsies remains to be the gold standard for the detection of GC, other screening tools have been implemented in high-risk countries such as pepsinogens (PGs) including PG 1 and PG 2, gastrin-17, and *Helicobacter pylori* (*H. pylori*) IgG antibody tests. Blood-based tumor markers such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) have also been used for the detection of GC, but have low sensitivity and specificity for early-stage disease [5,6].

Screening programs for GC vary among countries, depending on prevalence and cost-effectiveness [6]. We previously reported the use of extracellular RNA (exRNA) biomarkers in saliva as a diagnostic tool for screening and/or risk assessment for GC [7]. In a study of a Korean cohort of subjects, we identified 30 mRNA and 12 miRNA biomarkers that were associated with the expression pattern and presence of GC. A configured biomarker panel consisted of three mRNAs (*SPINK7*, *PPL*, and *SEMA4B*) and two miRNAs (*miR-140-5p* and *miR-301a-3p*) that were all significantly downregulated in the GC group, and yielded an area under the receiver operating characteristic (ROC) curve (AUC) of 0.81 (95% CI 0.72–0.89). When combined with demographic factors, the AUC of the biomarker panel reached 0.87 (95% CI 0.80–0.93) in differentiating subjects with GC from those without cancer. Since a lower expression of these salivary markers was indicative of GC, a comprehensive cut-off validation study would be necessary to develop these markers for the screening of GC in the general population.

It is known that the pathogenesis of GC depends on multiple etiological factors and ethnicity could obviously be one of determining factors [8]. However, it is also possible that there are common biological alterations that may contribute to the pathogenesis of disease and these genetic changes may be cancer-associated salivary exRNA alterations. The miRNAs and exRNAs are endogenous, small non-coding RNA molecules that post-transcriptionally modulate gene expression [9]. Because these molecules are stable in different body fluids including saliva, analysis of these molecules can lead to an important, noninvasive diagnostic and prognostic tool for GC screening. However, the expression of biomarkers may differ based on the population investigated. A key requirement of biomarker validation for clinical and regulatory purposes is that of intended use. A biomarker or panel of biomarkers that may differentiate disease from normal in one population may not perform similarly in a different population. While we have previously demonstrated that a particular panel of salivary biomarkers may differentiate subjects with GC from those without cancer in an Asian population with a high prevalence of gastric cancer, it is unclear whether this same panel of markers would perform similarly in a low-prevalence U.S. population. The current study represents a case-control study of salivary exRNA biomarkers in a U.S. cohort.

2. Materials and Methods

2.1. Saliva Collection and Processing

Unstimulated whole saliva was collected from 50 newly diagnosed treatment-naive patients with histologically proven GC (stages I–IV) and 50 control subjects without GC,

based on recent endoscopic results at the University of Texas MD Anderson Cancer Center, USA, using standard operating procedures (SOPs) developed for our prior study of a Korean cohort [7,10]. Subjects were asked to avoid oral hygiene measures, eating, drinking, and gum chewing at least 1 h prior to saliva collection. The subjects rinsed with tap water (10 mL) for 30 s about 10 min prior to saliva collection and expectorated. Clinical samples were collected in sterile tubes, lasting 5–10 min per collection (at least 5 mL of saliva), and kept on ice through the entire process. All samples were processed, approximately 1 h after collection. First, samples were centrifuged in a refrigerated centrifuge at $2400\times g$ for 15 min at 4 °C, and the supernatant was processed immediately for the concurrent stabilization of proteins and RNA by the inclusion of a protease inhibitor cocktail (aprotinin, 3-phenylmethylsulfonyl fluoride (3-PMSF) (Sigma-Aldrich, St. Louis, MO, USA), sodium orthovanadate (Na3VO4) (Sigma-Aldrich, St. Louis, MO, USA)) and RNase inhibitor (Invitrogen SUPERase·In RNase Inhibitor, Thermo Fisher Scientific, Austin, TX, USA) based on our saliva standard operating procedure (SOP) [11]. These samples were aliquoted into smaller cryo-vials, labeled, and frozen at −80 °C.

This study, including the patient consent process, was approved by the Institutional Review Board for Human Studies at the University of Texas MD Anderson Cancer Center (IRB number PA17-0583). The control group consisted of subjects undergoing upper endoscopy for dyspepsia or gastroesophageal reflux-like symptoms and documented to have no neoplasia. Patient-level clinical demographics were obtained (age, gender, ethnicity, smoking history, staging, and diagnosis). The study was performed from 19 October 2017 to 13 June 2019.

2.2. RNA Isolation from Saliva Samples

Total RNA from 50 blinded GC subjects and 50 non-GC control subjects was isolated by using a Qiagen miRNeasy Micro kit (Qiagen, Germantown, MD, USA). The 250 μL samples of cell-free saliva was used to isolate total RNA using a modified protocol successfully used in the lab for isolating salivary RNA [10]. The final RNA was eluted in 14 μL of water.

2.3. Validation of miRNA GC Markers

The biomarker panel used in this study contained two miRNAs (*miR-140-5p* and *miR-301a-3p*); *U6* snRNA and *miR-197* were used as the reference genes. TaqMan miRNA assays (Thermo Fisher Scientific, Austin, TX, USA), containing these four small RNA genes, were ordered from Applied Biosystems (Foster City, CA, USA). The protocol was similar to that recommended by the manufacturer for creating custom reverse transcription (RT) and preamplification primer pools using TaqMan MicroRNA Assays (Thermo Fisher Scientific, Austin, TX, USA). Total RNA (3 ng) was converted to cDNA using a TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA). After RT, the product was preamplified using SsoAdvance PreAmp Supermix (Bio-Rad, Hercules, CA, USA) and preamplification primer pool. The preamplified product was diluted 2 times prior to miRNA quantification. The qPCR reactions for each candidate miRNA were performed in triplicate on a Roche LightCycler 480 II (Roche, San Francisco, CA, USA). The average threshold cycle (Cq) was examined and *U6* snRNA and *miR-197* were used as the reference genes for normalizing the data.

2.4. Validation of mRNA GC Markers

Three selected candidate mRNA biomarkers (3 mRNAs (*PPL*, *SEMA4B*, and *SPINK7*) as well as 2 reference genes (*GAPDH* and *ACTB*)) generated by microarray profiling were validated by nested real-time quantitative polymerase chain reaction (RT-qPCR) (RT-PCR followed by a separate SYBR green quantitative polymerase chain reaction (qPCR)) on the new set of samples from MD Anderson Cancer Center (blinded 50 GC and 50 non-GC). The gene accession numbers and primer sequences used for the transcriptomic biomarker validation are shown in Supplementary Materials Table S1. The qPCR primers were designed using the Primer3 software and synthesized by Sigma-Genosys after performing

a Primer-BLAST search. The primer sequences were designed to avoid any known single-nucleotide polymorphism region in the target gene. All the amplicons were intron spanning. The RT-qPCR assay followed the Minimum Information for Publication of Quantitative Real-Time PCR Experiment guidelines and was performed in duplicate with each biomarker candidate. The specificity of the PCR product for each gene was confirmed with melting curve analysis and 3% agarose gel analysis.

2.5. RT-qPCR Preamplification for Validation of mRNA Candidates

The multiplex RT-PCR preamplification was performed with an Invitrogen SuperScript III Platinum One-Step qRT-PCR System (Thermo Fisher Scientific, Austin, TX, USA) with a pool of outer primers at 100 nM each. The reaction mixture was prepared on ice, and then loaded into the preheated thermocycler. The amplification was performed as follows: 2 min at 60 °C; 30 min at 50 °C; 2 min at 95 °C; and 15 cycles of 15 s at 95 °C, 30 s at 50 °C, 10 s at 60 °C, and 10 s at 72 °C; with a final extension for 10 min at 72 °C and cooling to 4 °C. Immediately after RT-qPCR, 10 μL of the reaction was treated with 4 μL of Exo-SAP-IT (Thermo Fisher Scientific, Austin, TX, USA) for 15 min at 37 °C to remove excess primers and deoxynucleotide triphosphates (dNTPs), and then heated to 80 °C for 15 min to inactivate the enzyme mix. The preamplified complementary DNAs (cDNAs) were then diluted by adding water to 200 μL (20-fold) to enable the qPCR of all targets.

2.6. qPCR for Validation of mRNA Candidates

Singleplex qPCR was performed in 10 μL reactions with 2 μL of each preamplified cDNA sample and the inner primers at 200 nM each. The reaction was conducted with a SYBR Green I Master mix in LightCycler 480 (Roche Diagnostics, Indianapolis-Marion County, Indiana) instrument. After 10 min of polymerase activation at 95 °C, 40 cycles of 15 s at 95 °C and 60 s at 60 °C were performed, followed by melting curve analysis. Three controls including one RT control, no-template control, and positive control with universal human RNA were performed with every candidate on each sample.

2.7. Statistical Analysis for qPCR

The qPCR analyses were all done in triplicate. For the miRNA analysis, data were analyzed using the RQ Manager software version 1.2 and DataAssist software version 3.0 (Applied Biosystems). Similarly, the ΔCq value was computed using RNA polymerase III transcribed *U6* small nuclear RNA as the reference gene [7]. For the mRNA analysis, the ΔCq of each biomarker candidate was calculated by subtracting the Cq value of the housekeeping genes (*GAPDH* and *ACTB*) from the raw Cq value in the same sample. ΔCq values for mRNA and miRNA were compared between groups using the Wilcoxon rank-sum test.

3. Results

3.1. Clinicopathological Characteristics of Patients

The patients' characteristics and study variables are summarized between groups (GC vs. control) using mean (SD) and frequency (%) and compared between groups using the two-sample *t*-test or chi-square test (Table 1).

Table 1. Patients' characteristics for non-GC and GC patients from the MD Anderson Cancer Center.

Patient Characteristics	Control (*n* = 49)	GC (*n* = 51)	*p*-Value	Test
Age	60.0 (11.2%)	61.0 (12.3%)	0.666	*t*-Test
Male	16 (32.7%)	32 (62.7%)	0.003	Chi-square
Ethnicity			0.467	Fisher's
Asian	3 (6.1%)	6 (11.8%)		
Black, non-Hispanic	6 (12.2%)	10 (19.6%)		
Caucasian	32 (65.3%)	26 (51.0%)		
Hispanic	8 (16.3%)	9 (17.6%)		
Present smoker	5 (10.2%)	7 (13.7%)	0.588	Chi-square
Prior smoker	16 (32.7%)	15 (29.4%)	0.726	Chi-square
Present or prior smoker	21 (42.9%)	22 (43.1%)	0.977	Chi-square
***H. pylori* biopsy performed**	22	35	–	–
***H. pylori* positive (% of tested individuals)**	2 (9.1%)	5 (14.3%)	0.695	Fisher's

3.2. miRNA RT-qPCR

Next, we constructed a model with the demographic terms (age, gender, and smoking history) plus the two candidate miRNA markers for GC (computing the dCT by subtracting the reference gene (*U6*) from our candidate markers (*miR-140* and *miR-301a*)). This was the same reference gene (*U6*) used in our prior study [7]. In this study, we found the CT values for *U6* were 15.37 ± 2.64 in the non-GC control group and 15.53 ± 2.70 in the GC patient group ($p = 0.809$ by *t*-test). The *p*-values show no significant differences between GC patients and non-GC controls, suggesting *U6* is a good reference gene for salivary exRNA quantification. To reduce the potential bias from one reference gene, we also tested *miR-197* as an extra reference small RNA. We found the CT values for *miR-197* were 16.20 ± 1.82 in the non-GC control group and 16.19 ± 1.85 in the GC patient group ($p = 0.796$). Next, we compared the AUC between the model with only demographic factors to the model utilizing demographic and miRNA data using the DeLong's test (Table 2). Analyses were conducted using IBM SPSS V25 (Armonk, NY, USA) and R V 3.6.1 (www.r-project.org (accessed on 20 March 2021), Vienna, AU, USA) and *p*-values < 0.05 were considered to be statistically significant. The AUC (95% CI) was 0.75 (0.65–0.84) for the GC group versus the non-GC group based on these two miRNA markers together with demographic factors. The markers (dCT) were both significant in that model (*miR-140* ($p = 0.003$), miR-301a ($p = 0.002$)). Interestingly, the demographic model alone yielded an AUC of only 0.68, while the combined model (demographic data with miRNA biomarkers) resulted in an improved AUC of 0.75 (DeLong *p*-value = 0.129).

Table 2. Models: (A) Demographic model for gastric cancer and (B) demographic model with two miRNA biomarkers for gastric cancer.

A. Demographic Model for Gastric Cancer		
Terms	**Odds Ratio (95% CI)**	***p*-Value**
Age	0.99 (0.96–1.04)	0.945
Male	3.74 (1.57–8.92)	0.003
Present or prior smoker	0.75 (0.31–1.79)	0.514
B. Demographic Model with Two miRNA Biomarkers for Gastric Cancer		
Terms	**Odds Ratio (95% CI)**	***p*-Value**
Age	0.99 (0.95–1.03)	0.683
Male	5.42 (2.03–14.48)	0.001
Ever Smoker	0.82 (0.33–2.07)	0.680
dCT*miR140_U6*	2.56 (1.37–4.79)	0.003
dCT*miR301_U6*	0.36 (0.19–0.68)	0.002

Next, logistic regression models for GC status were constructed using demographic factors from our previous publication (age, gender, and smoking history) with the AUC (95% CI) and odds ratios (ORs) estimated (Table 3). Interestingly, the demographic factors in this U.S. cohort showed an AUC of 0.68 (95% CI 0.57–0.78), which was similar to the AUC of 0.69 (95% CI 0.59–0.79) in the Korean cohort that we previously reported [7].

Table 3. Performance of the panel (demographic features + miRNAs + mRNAs) according to the different stages of GC (I–IV).

| vs. Control	Overall	Stage I/II	Stage III/IV	Stage IV
Demographic features only	0.68 (0.57–0.78)	0.67 (0.51–0.84)	0.67 (0.56–0.79)	0.68 (0.55–0.80)
Demographic features + miRNAs	0.75 (0.65–0.84)	0.80 (0.63–0.96)	0.72 (0.61–0.83)	0.70 (0.58–0.83)
Demographic features + miRNAs + mRNAs	0.78 (0.69–0.87)	0.85 (0.72–0.99)	0.75 (0.64–0.85)	0.74 (0.63–0.86)

3.3. mRNA RT-qPCR

We constructed a new model for the U.S. cohort using the same variables in three different ways, as reported in our previous report based on a Korean cohort [7]:

(1) Model 1, a new model with only demographic characteristics (AUC = 0.68, sensitivity = 62.7%, and specificity = 70.8%);
(2) Model 2, a new model with demographic characteristics and miRNA biomarkers for GC (AUC = 0.75, sensitivity = 62.7%, and specificity = 81.3%);
(3) Model 3, a new model with demographic characteristics, miRNA, and mRNA biomarkers for GC (AUC = 0.78, sensitivity = 62.7%, and specificity = 83.3%).

When mRNAs were combined with miRNA biomarkers and demographic features, the new combined model yielded the best AUC of 0.78 for differentiating subjects with GC from those without GC, with the highest specificity (sensitivity = 62.7% and specificity = 83.3%) (Figure 1) as compared with the models composed of only demographic features (AUC = 0.68, sensitivity = 62.7%, and specificity = 70.8%) or demographic features together with miR-NAs (AUC = 0.75, sensitivity = 62.7%, and specificity = 81.3%). There was also a significant difference between Model 1 (AUC = 0.68) and Model 3 (AUC = 0.78) (Delong's test *p*-value = 0.037), suggesting that the model with both miRNA and mRNA biomarkers for GC combined with demographic characteristics (Model 3) performed better than the model with only demographic characteristics (Model 1) with increased specificity (83.3% for Model 3 as compared with 70.8% for Model 1) (Figure 1, Table 4).

Table 4. Demographic characteristics with miRNA and mRNA biomarkers for gastric cancer.

Demographic Features + 2 miRNA Biomarkers for GC + 3 mRNA Biomarkers for GC		
Terms	**OR (95% CI)**	***p*-Value**
Age	0.99 (0.95–1.03)	0.544
Male	5.42 (2.03–14.48)	0.001
Ever Smoker	0.82 (0.33–2.07)	0.421
dCT*miR-140_U6*	2.56 (1.37–4.79)	0.007
dCT*miR-301a_U6*	0.36 (0.19–0.68)	0.002
dCT*PPL_ACTB*	0.88 (0.66–1.18)	0.406
dCT*SEMA4B_ACTB*	0.90 (0.66–1.23)	0.497
dCTSPINK7_ACTB	1.23 (0.94–1.60)	0.132
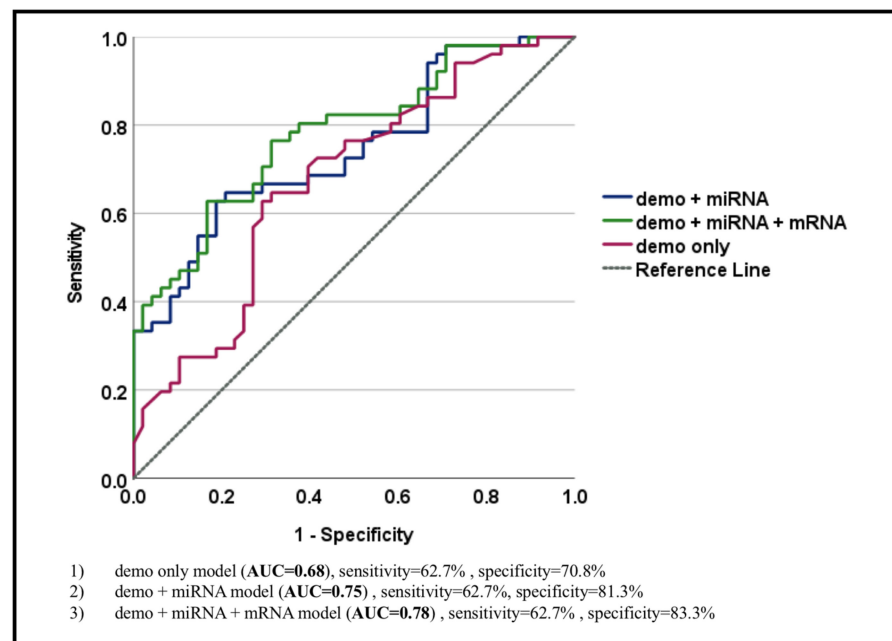
Figure 1. Performance of 3 different models: (**1**) A new model with only demographic characteristics (AUC = 0.68, sensitivity = 62.7%, and specificity = 70.8%); (**2**) a new model with demographic characteristics and miRNA biomarkers for gastric cancer (AUC = 0.75, sensitivity = 62.7%, and specificity = 81.3%); (**3**) a new model with demographic characteristics, miRNA and mRNA biomarkers for gastric cancer (AUC = 0.78, sensitivity = 62.7, and specificity = 83.3%).

However, when we applied the coefficients as estimated from the prior Korean cohort study [7], the AUC was only 0.52 because of differences in the significance of individual demographic features as well as in performance of GC miRNA and mRNA biomarkers in the current U.S. cohort as compared with the Korean cohort.

Additionally, we also assessed the panel performance (demographic characteristics (demo) + miRNAs + mRNAs) in two separate scenarios defined as controls vs. early-stage GC (I, II) as well as controls vs. late-stage GC (III, IV). It appeared that the discrimination (AUC) of the control vs. early-stage GC model was 0.85 (0.72–0.99), whereas for the control vs. late-stage GC, the performance was 0.75 (0.64–0.85). Therefore, our panel may perform better in discriminating controls from early-stage GC, although this would need to be confirmed in a follow-up study (Table 3).

4. Discussion

A growing number of studies have demonstrated the utility of exRNA as a reliable noninvasive approach for diagnosis, therapy, and prognosis of cancers [12]. Extracellular RNAs have been explored as biomarkers in a number of different biofluids and types of cancer, which include esophageal squamous cell carcinoma (ESCC) [13], lung cancer [14], brain cancers [15–18], prostate cancer [19], pancreatic cancer [20], colon cancer [21], and gastric cancer [22]. As of 2020, 45 clinical trials, in the USA and numerous other countries, have been reported that have focused on the use of exRNA and exosomes as clinical biomarkers of cancer [12]. These clinical trials have explored exRNAs as clinical biomarkers of various cancer types including lung and prostate cancers. Blood is a primary source of exRNAs that have been tested, but studies have also investigated urine. Especially saliva is being explored as an emerging biofluid that is easy to collect, and has been shown to reflect the spectrum of health and disease states found using serum [23,24].

Standards for validation of biomarkers require that they be applied in the population for which they are intended to be used [25]. A biomarker or panel of biomarkers which can differentiate disease from normal in one population may not perform similarly in a different population. We previously reported on a panel of salivary biomarkers which, when combined with specific demographic factors, differentiated subjects with GC from those without cancer in an Asian population with a high prevalence of gastric cancer. Our aim was to evaluate the performance of salivary exRNA biomarkers for GC, which we previously discovered and validated in Korean GC patients [7], in a U.S. population. Previously, 12 mRNA and 6 miRNA candidates were verified with a discovery Korean cohort by RT-qPCR and further validated with an independent Korean cohort ($n = 200$). The configured biomarker panel consisted of three mRNAs (*SPINK7*, *PPL*, and *SEMA4B*) and two miRNAs (*miR-140-5p* and *miR-301a*), which were all significantly downregulated in the GC group, and yielded an AUC of 0.81 (95% CI 0.72–0.89). When combined with demographic factors, the AUC of the biomarker panel reached 0.87 (95% CI 0.80–0.93) [7]. In our prior study, demographic characteristics (including age, gender, and smoking) were all highly significant predictors of case status [7], while, in the current U.S. MD Anderson Cancer Center cohort, only gender was significant (Table 4). However, this U.S. cohort (MD Anderson Cancer Center cohort) had more ethnic diversity including Caucasian (51% of the GC group and 65.3% of the control group), Black non-Hispanic (19.6% of the GC group and 12.2% of the control group), and Hispanic (17.6% of the GC group and 16.3% of the control group) subjects, with the Asian population as the least prevalent group (11.8% of theGC group and 6.1% of the control group). In a previous study [7], Asians constituted 100% of the GC and control groups for both mRNA and miRNA discovery and validation phases (Figure 2). In addition, in the current study, the samples were obtained from older patients (GC group, 61 years old and control group, 60 years old), fewer smokers (present or prior smoking, 43.1% of the GC group and 42.9% of the control group), and fewer males (62.7% of the GC group and 32.7% of the control group) as compared with the Korean cohort (Figure 2) [7].

It is unclear how these factors may account for differences between the two distinct populations with respect to biomarker profiles. Interestingly, the prevalence of *H. pylori* positivity in the U.S. population with gastric cancer was relatively low as compared with Asian populations with gastric cancer where *H. pylori* was a major risk factor. There was no difference, however, in the prevalence of *H. pylori* between cancer patients and controls in the U.S. cohort, but the number of subjects studied was small.

These demographic differences may be important factors to consider for validation of salivary exRNA GC biomarkers in two entirely independent patient cohorts (Korea vs. USA). Our study indicated that, overall, the demographic factors in this U.S. cohort were similar (AUC of 0.68) to those of the Korean cohort (AUC of 0.69) [7]. In our previous study with Korean subjects, we found a difference in nearly all the selected mRNAs (*ANXA1, CD24, CSTB, ERO1A, KRT4, KRT6A, PPL, RANBP9, S100A10, SEMA4B,* and *SPINK7*) and miRNAs for GC (*miR-140-5p, miR-374a, miR-454, miR-15b, miR-28-5p,* and *miR-301a*). They were all statistically significant (FDR-adjusted *p*-value <0.05). However, in the U.S. cohort, none of the miRNAs (*miR-140* and *miR-301*) and mRNAs (*SPINK7, PPL,* and *SEMA4B*) performed similarly (Supplementary Materials Figure S1). Therefore, any model constructed from the prior Korean cohort could not be generalized to the current U.S. cohort [7].

Demographic characteristics - Comparison between Korean and U.S. (MDA Anderson cohort)

Table 1. Demographic characteristics for non-GC and GC patients from the discovery phase [Feng Li et al 2018]

Demographic Variable	Characteristic	Transcriptomic Biomarker Discovery Phase [a]			miRNA Biomarker Discovery Phase [b]		
		GC (n=63)	Non-GC (n=31)	*P*-value	GC (n=10)	Non-GC (n=10)	*P*-value
Age (y)	Mean ± SD	56.2±11.1	54.8±10.4	0.56	58.4±7.6	52.4±7.9	0.10
Gender, *n* (%)	Male	43 (68.3)	13 (41.9)	0.02	6 (60.0)	4 (40.0)	0.66
	Female	20 (31.7)	18 (59.1)		4 (40.0)	6 (60.0)	
Ethnicity, *n* (%)	Asian	63 (100)	31 (100)		10 (100)	10 (100)	
Smoking, *n* (%)		27 (42.8)	5 (16.1)	0.01	3 (30.0)	0 (0.0)	0.21
Drinking, *n* (%)		28 (44.4)	9 (29.0)	0.15	4 (40.0)	5 (50.0)	1.00
H. Pylori, *n* (%)		37 (58.7)	8 (25.8)	0.003	5 (50.0)	4 (40.0)	1.00

[a] Transcriptomic profiling (n=94); [b] miRNA profiling (n=20).

Table 2. Demographic characteristics for non-GC and GC patients (n=200) from the validation phase [Feng Li et al 2018]

Demographic Variable	Characteristic	Validation Phase		
		GC (n=100)	Non-GC (n=100)	*P*-value
Age (y)	Mean ± SD	51.9±9.2	55.7±8.1	0.003
Gender, *n* (%)	Male	79 (79.0)	47 (47.0)	<0.001
	Female	21 (21.0)	53 (53.0)	
Ethnicity, *n* (%)	Asian	100 (100)	100 (100)	
Smoking, *n* (%)		66 (44.0)	33 (33.0)	<0.001
Drinking, *n* (%)		55 (55.0)	N/A	
H. Pylori, *n* (%)		N/A	N/A	

N/A: Not available

Table 3. Demographic characteristics for non-GC and GC patients from U.S. MDA cohort

Patient Characteristics	Control (n=49)	GC (n=51)	*P*-value	test
Age	60.0 (11.2%)	61.0 (12.3%)	0.666	t-test
Male	16 (32.7%)	32 (62.7%)	0.003	chi-square
Ethnicity			0.467	Fisher's
Asian	3 (6.1%)	6 (11.8%)		
African American	6 (12.2%)	10 (19.6%)		
Caucasian	32 (65.3%)	26 (51.0%)		
Hispanic	8 (16.3%)	9 (17.6%)		
Present smoker	5 (10.2%)	7 (13.7%)	0.588	chi-square
Prior smoker	16 (32.7%)	15 (29.4%)	0.726	chi-square
Present or prior smoker	21 (42.9%)	22 (43.1%)	0.977	chi-square
***H. pylori* biopsy performed**	22	35		
***H. pylori* positive (% of tested individuals)**	2 (9.1%)	5 (14.3%)	0.695	Fisher's

Figure 2. Comparison of demographic characteristics between Korean and U.S. study groups. The demographic characteristics of the Korean cohort used in our previous study [7] (Table 1 for the discovery phase and Table 2 for the validation phase) were compared with the demographic characteristics of the U.S. cohort used in this study (Table 3). (Tables 1 and 2 adapted from Li et al. [7]. Reprinted with permission of the Oxford University Press, Copyright © 2022 American Association of Clinical Chemistry. Li et al., Discovery and Validation of Salivary Extracellular RNA Biomarkers for Noninvasive Detection of Gastric Cancer. *Clin. Chem.* 2018, *64*, 1513–1521. https://doi.org/10.1373/clinchem.2018.290569).

Interestingly, there was a statistically significant difference between a new model with only demographic characteristics (AUC = 0.68) and a new model with demographic characteristics, and miRNA/mRNA biomarkers for GC (AUC = 0.78) (Figure 1) (Delong's test *p*-value = 0.037), indicating that the model with both miRNAs and mRNAs together with demographic characteristics (Model 3) was much better than the model with demographic characteristics alone (Model 1). Interestingly, our panel has a potential to perform better in discriminating non-GC controls from early-stage GC (I, II) (AUC = 0.85 (95% CI 0.72–0.99)) as compared with the late-stage GC (III, IV) (AUC = 0.75 (95% CI 0.64–0.85)), although this still would need to be confirmed in a follow-up study. Thus, we were able to validate a panel of salivary exRNA biomarkers with credible clinical performance for the detection of GC in a U.S. population. Our study confirms, again, the potential utility of salivary exRNA biomarkers in screening and risk assessment for GC.

4.1. Current Biomarker Performance

Most of the currently available published reports are based on blood-based biomarkers for GC. So et al. developed a 12-miR assay from serum specimens (*miR-140, miR-183, miR-30e, miR-103a, miR-126, miR-93, miR-142, miR-21, miR-29c, miR-424, miR-181a,* and *miR-340*) in a three-phase, multicenter study comprising 5248 subjects from Singapore and Korea in retrospective cohorts of 682 subjects [26]. Interestingly, one of the miRNAs was *miR-140*, the same as that investigated in our studies. The 12- miRNA panel yielded an AUC of 0.93 (95% CI 0.90–0.95) and an AUC of 0.92 (95% CI 0.88–0.96) in the discovery and verification cohorts, respectively. In this prospective study, overall sensitivity was 87.0% (95% CI

79.4–92.5%) at specificity of 68.4% (95% CI 67.0–69.8%). Interestingly, the AUC was 0.848 (95% CI 0.81–0.88), much higher than other more frequently used gastric tumor markers such as *H. pylori* serology (0.635), pepsinogen (PG) 1/2 ratio (0.641), PG index (0.576), ABC method (0.647), CEA (0.576), and CA19-9 (0.595) [26]. Other plasma miRNAs yielded AUCs from 0.65 to 0.75 for *miR-185*, *miR-20a*, *miR-210*, *miR-25*, and *miR-92b* [27]. In another study, *miR-181a-1* and *KAT2B* mRNA were identified as a combined predictor for GC with AUC > 0.95 [28], while the expression levels of *HOXC6* mRNA in patients with advanced GC (AGC) were found to be significantly higher than those in patients with early-stage GC (EGC) [29]. Another report suggested that the combination of three biomarkers (collagen type VI alpha 3 chain (*COL6A3*), serpin family H member 1 (*SERPINH1*), and pleckstrin homology and RhoGEF domain containing G1 (*PLEKHG1*)) yielded an elevated AUC of 0.907. A higher *COL6A3* level was also significantly correlated with lymph node metastasis and poor prognosis in GC patients, while high levels of SER*PINH1* and *PLEKHG1* mRNA expression were correlated with lower overall survival (OS) in GC patients [30]. CircRNAs may also be an auxiliary diagnostic biomarker of GC [16]. In a combined 11 studies, which included 12 types of circRNAs (11 in tissues and 5 in plasma), all were downregulated. The combined diagnostic OR (DOR) and AUC with 95% CI were 8.778 (6.108, 12.614) and 0.81 (0.78, 0.84) respectively [31]. Another type of RNA with diagnostic value for GC detection include long non-coding RNAs (lncRNAs). The ROC curve showed that the AUC of serum lncRNA HCP5 detected by qRT-PCR was 0.818 (95% CI 0.757–0.880, $p < 0.001$, 80% sensitivity, and 70% specificity), while the three combined diagnoses (HCP5, CEA, and CA199) provided the highest AUC of 0.870 (95% CI 0.819–0.921, $p < 0.001$) in distinguishing between GC and healthy donors reaching 81% sensitivity and 79% specificity [32]. In addition, it was found that LINC00941 was associated with tumor depth and distant metastasis in GC as it could discriminate GC samples from normal samples (AUC = 0.7911, 95% CI 0.7264–0.8559, $p < 0.0001$) and M1 samples from M0 samples (AUC = 0.6809, 95% CI 0.5852–0.7766, $p = 0.0031$) [33].

4.2. Limitations, Future Studies, and Advantage of the Markers Used in This Study

The abovementioned studies suggest salivary RNAs as potential biomarkers for the diagnosis of GC, but emphasize the need for validation in intended use populations. However, no study of adequate sample size for independent validation has been performed to date. There remains an unmet need to develop a noninvasive biomarker assay for identifying patients with GC from a high-risk population. Thus, the major advantage of the markers used in the current study is their noninvasive nature, which is important for population-based screening in at-risk populations.

5. Conclusions

We aimed to develop universal biomarkers for GC that could be applicable to all individuals regardless of their ethnic origin. Although we were unable to "validate" the prior model developed based on a Korean cohort [7], we were able to demonstrate that our markers had diagnostic utility above and beyond demographic factors alone. Additional studies are needed to evaluate the diagnostic utility of our models in different ethnic populations, such as a Korean cohort in a U.S. population. More importantly, our study emphasizes the importance of population-specific biomarker development and validation for salivary exRNA biomarker for GC detection.

Institutional Review Board Statement: The study was approved by the Institutional Review Board for Human Studies at the University of Texas MD Anderson Cancer Center (IRB number PA17-0583). The study was performed from 19 October 2017 to 13 June 2019.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data can be found in the text.

References

1. Rawla, P.; Barsouk, A. Epidemiology of gastric cancer: Global trends, risk factors and prevention. *Prz. Gastroenterol.* **2019**, *14*, 26–38. [CrossRef] [PubMed]

2. Bresalier, R.S. Treatment of advanced gastric cancer. In *World Gastroenterology Organization Handbook on Early Diagnosis and Treatment of GI Cancer*; Kuipers, E., Sung, J., Eds.; World Gastroenterology Organisation (WGO): Milwaukee, WI, USA, 2020; Available online: https://www.changing-transport.org/transfer-project-in-the-philippines/ (accessed on 23 March 2022).

3. Song, Z.; Wu, Y.; Yang, J.; Yang, D.; Fang, X. Progress in the treatment of advanced gastric cancer. *Tumour. Biol.* **2017**, *39*, 1010428317714626. [CrossRef] [PubMed]

4. Yang, L.; Ying, X.; Liu, S.; Lyu, G.; Xu, Z.; Zhang, X.; Li, H.; Li, Q.; Wang, N.; Ji, J. Gastric cancer: Epidemiology, risk factors and prevention strategies. *Chin. J. Cancer Res.* **2020**, *32*, 695–704. [CrossRef] [PubMed]

5. Pectasides, D.; Mylonakis, A.; Kostopoulou, M.; Papadopoulou, M.; Triantafillis, D.; Varthalitis, J.; Dimitriades, M.; Athanassiou, A. CEA, CA 19–9, and CA-50 in monitoring gastric carcinoma. *Am. J. Clin. Oncol.* **1997**, *20*, 348–353. [CrossRef] [PubMed]

6. Pasechnikov, V.; Chukov, S.; Fedorov, E.; Kikuste, I.; Leja, M. Gastric cancer: Prevention, screening and early diagnosis. *World J. Gastroenterol.* **2014**, *20*, 13842–13862. [CrossRef] [PubMed]

7. Li, F.; Yoshizawa, J.M.; Kim, K.M.; Kanjanapangka, J.; Grogan, T.R.; Wang, X.; Elashoff, D.E.; Ishikawa, S.; Chia, D.; Liao, W.; et al. Discovery and Validation of Salivary Extracellular RNA Biomarkers for Noninvasive Detection of Gastric Cancer. *Clin. Chem.* **2018**, *64*, 1513–1521. [CrossRef] [PubMed]

8. Machlowska, J.; Baj, J.; Sitarz, M.; Maciejewski, R.; Sitarz, R. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int. J. Mol. Sci.* **2020**, *21*, 4012. [CrossRef]

9. Patton, J.G.; Franklin, J.L.; Weaver, A.M.; Vickers, K.; Zhang, B.; Coffey, R.J.; Ansel, K.M.; Blelloch, R.; Goga, A.; Huang, B.; et al. Biogenesis, delivery, and function of extracellular RNA. *J. Extracell. Vesicles* **2015**, *4*, 27494. [CrossRef]

10. Kaczor-Urbanowicz, K.E.; Trivedi, H.M.; Lima, P.O.; Camargo, P.M.; Giannobile, W.V.; Grogan, T.R.; Gleber-Netto, F.O.; Whiteman, Y.; Li, F.; Lee, H.J.; et al. Salivary exRNA biomarkers to detect gingivitis and monitor disease regression. *J. Clin. Periodontol.* **2018**, *45*, 806–817. [CrossRef]

11. Henson, B.S.; Wong, D.T. Collection, storage, and processing of saliva samples for downstream molecular applications. *Methods Mol. Biol.* **2010**, *666*, 21–30. [CrossRef]

12. Happel, C.; Ganguly, A.; Tagle, D.A. Extracellular RNAs as potential biomarkers for cancer. *J. Cancer Metastasis Treat.* **2020**, *6*, 32. [CrossRef] [PubMed]

13. Tanaka, Y.; Kamohara, H.; Kinoshita, K.; Kurashige, J.; Ishimoto, T.; Iwatsuki, M.; Watanabe, M.; Baba, H. Clinical impact of serum exosomal microRNA-21 as a clinical biomarker in human esophageal squamous cell carcinoma. *Cancer* **2013**, *119*, 1159–1167. [CrossRef]

14. Rabinowits, G.; Gercel-Taylor, C.; Day, J.M.; Taylor, D.D.; Kloecker, G.H. Exosomal microRNA: A diagnostic marker for lung cancer. *Clin. Lung Cancer* **2009**, *10*, 42–46. [CrossRef]

15. Skog, J.; Wurdinger, T.; van Rijn, S.; Meijer, D.H.; Gainche, L.; Sena-Esteves, M.; Curry, W.T., Jr.; Carter, B.S.; Krichevsky, A.M.; Breakefield, X.O. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* **2008**, *10*, 1470–1476. [CrossRef] [PubMed]

16. Balaj, L.; Lessard, R.; Dai, L.; Cho, Y.J.; Pomeroy, S.L.; Breakefield, X.O.; Skog, J. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* **2011**, *2*, 180. [CrossRef]

17. Noerholm, M.; Balaj, L.; Limperg, T.; Salehi, A.; Zhu, L.D.; Hochberg, F.H.; Breakefield, X.O.; Carter, B.S.; Skog, J. RNA expression patterns in serum microvesicles from patients with glioblastoma multiforme and controls. *BMC Cancer* **2012**, *12*, 22. [CrossRef] [PubMed]

18. Chen, W.W.; Balaj, L.; Liau, L.M.; Samuels, M.L.; Kotsopoulos, S.K.; Maguire, C.A.; Loguidice, L.; Soto, H.; Garrett, M.; Zhu, L.D.; et al. BEAMing and Droplet Digital PCR Analysis of Mutant IDH1 mRNA in Glioma Patient Serum and Cerebrospinal Fluid Extracellular Vesicles. *Mol. Ther. Nucleic Acids.* **2013**, *2*, e109. [CrossRef] [PubMed]

19. Khan, S.; Jutzy, J.M.; Valenzuela, M.M.; Turay, D.; Aspe, J.R.; Ashok, A.; Mirshahidi, S.; Mercola, D.; Lilly, M.B.; Wall, N.R. Plasma-derived exosomal survivin, a plausible biomarker for early detection of prostate cancer. *PLoS ONE* **2012**, *7*, e46737. [CrossRef]

20. Lau, C.; Kim, Y.; Chia, D.; Spielmann, N.; Eibl, G.; Elashoff, D.; Wei, F.; Lin, Y.L.; Moro, A.; Grogan, T.; et al. Role of pancreatic cancer-derived exosomes in salivary biomarker development. *J. Biol. Chem.* **2013**, *288*, 26888–26897. [CrossRef]

21. Garcia, J.M.; Garcia, V.; Pena, C.; Dominguez, G.; Silva, J.; Diaz, R.; Espinosa, P.; Citores, M.J.; Collado, M.; Bonilla, F. Extracellular plasma RNA from colon cancer patients is confined in a vesicle-like structure and is mRNA-enriched. *RNA* **2008**, *14*, 1424–1432. [CrossRef]

22. Kim, H.K.; Song, K.S.; Park, Y.S.; Kang, Y.H.; Lee, Y.J.; Lee, K.R.; Kim, H.K.; Ryu, K.W.; Bae, J.M.; Kim, S. Elevated levels of circulating platelet microparticles, VEGF, IL-6 and RANTES in patients with gastric cancer: Possible role of a metastasis predictor. *Eur. J. Cancer* **2003**, *39*, 184–191. [CrossRef]

23. Slavkin, H.C. Toward molecularly based diagnostics for the oral cavity. *J. Am. Dent. Assoc.* **1998**, *129*, 1138–1143. [CrossRef] [PubMed]

24. Lee, Y.H.; Wong, D.T. Saliva: An emerging biofluid for early detection of diseases. *Am. J. Dent.* **2009**, *22*, 241–248. [PubMed]

25. Pepe, M.S.; Feng, Z.; Janes, H.; Bossuyt, P.M.; Potter, J.D. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: Standards for study design. *J. Natl. Cancer Inst.* **2008**, *100*, 1432–1438. [CrossRef]

26. So, J.B.Y.; Kapoor, R.; Zhu, F.; Koh, C.; Zhou, L.; Zou, R.; Tang, Y.C.; Goo, P.C.K.; Rha, S.Y.; Chung, H.C.; et al. Development and validation of a serum microRNA biomarker panel for detecting gastric cancer in a high-risk population. *Gut* **2021**, *70*, 829–837. [CrossRef]

27. Zhu, C.; Ren, C.; Han, J.; Ding, Y.; Du, J.; Dai, N.; Dai, J.; Ma, H.; Hu, Z.; Shen, H.; et al. A five-microRNA panel in plasma was identified as potential biomarker for early detection of gastric cancer. *Br. J. Cancer* **2014**, *110*, 2291–2299. [CrossRef]

28. Chen, S.; Zhu, J.; Yu, F.; Tian, Y.; Ma, S.; Liu, X. Combination of miRNA and RNA functions as potential biomarkers for gastric cancer. *Tumour. Biol.* **2015**, *36*, 9909–9918. [CrossRef]

29. Jung, J.; Jeong, S.; Jeong, H.; Oh, H.E.; Choi, J.W.; Lee, E.S.; Kim, Y.S.; Kwak, Y.; Kim, W.H.; Lee, J.H. Increased HOXC6 mRNA expression is a novel biomarker of gastric cancer. *PLoS ONE* **2020**, *15*, e0236811. [CrossRef]

30. Cao, W.; Zhou, D.; Tang, W.; An, H.; Zhang, Y. Discovery of plasma messenger RNA as novel biomarker for gastric cancer identified through bioinformatics analysis and clinical validation. *PeerJ* **2019**, *7*, e7025. [CrossRef]

31. Jiang, F.; Hong, F.; Shah, M.W.; Shen, X. Circular RNAs as diagnostic biomarkers in gastric cancer: A meta-analysis review. *Pathol. Res. Pract.* **2019**, *215*, 152419. [CrossRef]

32. Qin, S.; Yang, L.; Kong, S.; Xu, Y.; Liang, B.; Ju, S. LncRNA HCP5: A Potential Biomarker for Diagnosing Gastric Cancer. *Front. Oncol.* **2021**, *11*, 684531. [CrossRef] [PubMed]

33. Liu, H.; Wu, N.; Zhang, Z.; Zhong, X.; Zhang, H.; Guo, H.; Nie, Y.; Liu, Y. Long Non-coding RNA LINC00941 as a Potential Biomarker Promotes the Proliferation and Metastasis of Gastric Cancer. *Front. Genet.* **2019**, *10*, 5. [CrossRef] [PubMed]