Hispanic/Latino Patients with Gastric Adenocarcinoma Have Distinct Molecular Profiles Including a High Rate of Germline CDH1 Variants

Sam C. Wang1, Yunku Yeu2, Suntreia T.G. Hammer3, Shu Xiao1, Min Zhu4, Changjin Hong2, Jean R. Clemenceau2, Lynn Y. Yoon1, Ibrahim Nassour1, Jeanne Shen5, Deepak Agarwala6, Scott I. Reznik7, John C. Mansour1, Adam C. Yopp1, Hao Zhu4, Tae Hyun Hwang2, and Matthew R. Porembka1

ABSTRACT

Hispanic/Latino patients have a higher incidence of gastric cancer and worse cancer-related outcomes compared with patients of other backgrounds. Whether there is a molecular basis for these disparities is unknown, as very few Hispanic/Latino patients have been included in previous studies. To determine the genomic landscape of gastric cancer in Hispanic/Latino patients, we performed whole-exome sequencing (WES) and RNA sequencing on tumor samples from 57 patients; germline analysis was conducted on 83 patients. The results were compared with data from Asian and White patients published by The Cancer Genome Atlas. Hispanic/Latino patients had a significantly larger proportion of genomically stable subtype tumors compared with Asian and White patients (65% vs. 21% vs. 20%, P < 0.001). Transcriptomic analysis identified molecular signatures that were prognostic. Of the 43 Hispanic/Latino patients with diffuse-type cancer, 7 (16%) had germline variants in CDH1. Variant carriers were significantly younger than noncarriers (41 vs. 50 years, P < 0.05). In silico algorithms predicted five variants to be deleterious. For two variants that were predicted to be benign, in vitro modeling demonstrated that these mutations conferred increased migratory capability, suggesting pathogenicity. Hispanic/Latino patients with gastric cancer possess unique genomic landscapes, including a high rate of CDH1 germline variants that may partially explain their aggressive clinical phenotypes. Individualized screening, genetic counseling, and treatment protocols based on patient ethnicity and race may be necessary.

Significance: Gastric cancer in Hispanic/Latino patients has unique genomic profiles that may contribute to the aggressive clinical phenotypes seen in these patients.

Introduction

Gastric cancer is the second-deadliest cancer worldwide, causing an estimated 834,000 deaths in 2016 (1). Hispanic/Latino patients have different clinicopathologic features than patients of other ethnicities and races. In the United States, Hispanics/Latinos have twice the incidence and mortality from gastric cancer compared with non-Hispanic Whites (2). Hispanic/Latino patients with gastric cancer also tend to be diagnosed at a younger age, with more advanced-stage disease, and with a higher proportion of diffuse-type cancers (DGC; refs.3–5). While environmental exposures and socioeconomic factors likely contribute to the observed clinicopathologic differences, ethnicity/race-associated differences in tumor biology may also be involved. For example, African-American patients with breast cancer have higher rates of triple-negative cancers and a higher prevalence of TP53 mutations, as compared with White patients (6, 7).

Whether there is a molecular basis for observed outcome differences for patients with gastric cancer of different ethnicities/races has been heretofore unanswerable as previous large gastric cancer genomic studies had included very few Hispanic/Latino patients. The Cancer Genome Atlas (TCGA) has performed the largest published sequencing study of gastric adenocarcinoma and included only five Hispanic/Latino patients in its 478-patient cohort (8). Other major sequencing efforts of gastric cancer originated in East Asia, including those by Ichikawa and colleagues (207 patients) and Cristescu and colleagues (300 patients); these studies also did not include any Hispanic/Latino patients (9, 10). Given the known association between ethnicity/race and tumor biology, the under-representation of Hispanic/Latino patients in previously published studies have likely biased our current genomic understanding of gastric cancer (11).

To address this knowledge gap, we performed a large, integrated genomic analysis of samples from 83 Hispanic/Latino patients with gastric cancer. Comparative analyses were performed using data from Asian and White patients previously published by TCGA (12).
Materials and Methods

Sample acquisition and processing

This study was approved by the University of Texas Southwestern Medical Center Institutional Review Board. All patients with gastric adenocarcinoma who were self-reported as being of Hispanic/Latino ancestry were recruited to join the study. All enrolled patients provided written informed consent.

Blood samples were drawn and stored at −80°C prior to nucleic acid extraction. Tumor and adjacent nonneoplastic gastric tissue were obtained from subjects via endoscopic biopsies or gastric resections. The samples were stabilized immediately in RNAlater (Ambion) for at least 24 hours at 4°C, then stored in liquid nitrogen until nucleic acid extraction. A second set of adjacent tissue samples from both the tumor and nonneoplastic stomach were also obtained for pathologic examination to confirm the histology, and to provide a microscopic assessment of tumor cellularity and extent of tumor necrosis. These samples were evaluated by a board-certified pathologist with expertise in gastrointestinal malignancies (S.T.G. Hammer). No samples were excluded on the basis of tumor cellularity. Samples with greater than 10% necrosis were excluded. For some samples, RNA was isolated with mirVana miRNA Isolation Kits (Ambion) and DNA was isolated with QuickGene DNA Tissue Kits (Kurabo). Other samples were processed using the AllPrep DNA/RNA Kits (Qiagen). Nucleic acid quality control was ensured with NanoDrop (Thermo Fisher Scientific) spectrophotometric quantitation and visualization on an agarose gel.

**CDH1 promoter methylation**

Tumor and nontumor DNA were prepared with the EpiTect II DNA Methylation Enzyme Kit (Qiagen). Quantification of methylated DNA was then performed using qPCR based assay with the EpiTect Methyl II PCR Primer Assay for Human CDH1, CpG Island 45769 (Qiagen) with the RT2 SYBR Green qPCR Mastermix (Qiagen). DNA was then performed using the EpiTect Methylation PCR Kit (Qiagen) with the RT2 SYBR Green qPCR Mastermix (Qiagen).

**Generation of CDH1 mutants**

Chinese hamster ovary (CHO) cells (ATCC, CCL-61) were maintained in F-12K medium (Gibco) with 10% FBS supplementation. Mycoplasma testing was performed upon receipt of the cells. hE-cadherin-pcDNA3 was a gift from Barry Gumbiner (Addgene plasmid #45769; http://n2t.net/addgene:45769; RRID:Addgene_45769). Variants were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). Plasmid transfection was performed with Lipofectamine 3000 (Invitrogen). Selection was performed with 4-18G (Sigma). Sanger sequencing was performed to confirm sequences using the primers (Genewiz) in Table 1.

**Immunofluorescence**

CHO cells were fixed on glass slides with precooled methanol for 15 minutes at −20°C and blocked by 1% BSA in PBS-T for 1 hour at room temperature. The slides were then incubated with anti-E-cadherin antibody (Abcam, ab76055; 1:1,000 dilution) at 4°C overnight, followed by secondary antibody at room temperature for 1 hour. DAPI was used as a nuclear stain (Vector Laboratories). Images were captured on a Zeiss confocal microscope.

**IHC**

Antigen retrieval was performed with sodium citrate buffer, followed by incubation with anti-E-cadherin antibody (Abcam, ab76055; 1:1,000) at 4°C overnight. Detection was performed with the ABC Kit (Vector Laboratories) and DAB Kit (Vector Laboratories) or MOM Kit (Vector Laboratories).

---

Table 1. Primers.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>PCR primer F</th>
<th>PCR primer R</th>
<th>Sanger sequencing primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>P15</td>
<td>TGTGCCCAATC-</td>
<td>CAGCGTGACT-</td>
<td>TCAGAGCACA-</td>
</tr>
<tr>
<td></td>
<td>GAGAAGTTA</td>
<td>AGCCGAGCTT</td>
<td>AGGAAGTCA-</td>
</tr>
<tr>
<td>P16</td>
<td>CTTCTCCAAAG-</td>
<td>TCAAAGGCCTG-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td></td>
<td>CCTTAGACC</td>
<td>AGTCTGCTC</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td>P20</td>
<td>TGTAACACGGC-</td>
<td>GGAAGGAAAGT</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td></td>
<td>CAGAGACCT</td>
<td>GAGTCTGGGA-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td>P30</td>
<td>CCCACACATCC-</td>
<td>CCTTAGCCAAG</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td></td>
<td>AGTTCTGAT</td>
<td>CCTTACGCAA-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td>P33</td>
<td>CTGTGGTGTTTC-</td>
<td>GCTCCAACCC-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td></td>
<td>GGTGACGAGC</td>
<td>TCCTTCTCTT-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td>P50</td>
<td>GACCCAGAGCA-</td>
<td>CTTCCATGAA-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td></td>
<td>GTTTCACCC</td>
<td>CAGACCCCTT-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td>P71</td>
<td>AGTCGGGGAAG-</td>
<td>CTCAAGGAGG-</td>
<td>ACAAAGGAAAT-</td>
</tr>
<tr>
<td></td>
<td>ATGTGTCGTA</td>
<td>GGAGCTGAA-</td>
<td>ACAAAGGAAAT-</td>
</tr>
</tbody>
</table>

Scratch assay

CHO cells transfected with a given plasmid were grown to confluence. A scratch was made and three images were taken of each well. Twenty-four hours later, three more images of each well were taken. The distance between the wound edges was measured using cellSens Dimension Software (Olympus). The average of the three images from each time point was used as one biological replicate. Two independent experiments with at least four biological replicates for each genotype were performed.

**Statistical analysis**

The Mann–Whitney U test was used to compare continuous variables. Categorical variables were presented as counts and proportions and compared with Fisher exact tests. Survival was estimated using the Kaplan–Meier method and compared via the log-rank test. For the epidemiologic studies, data were presented as medians with interquartile ranges and full ranges in box and whisker plots and compared with the Kruskal–Wallis test.

**Whole-exome sequencing, RNA sequencing, and bioinformatic analyses**

See the Supplementary Methods section for details regarding the whole-exome sequencing (WES), RNA sequencing (RNA-seq), and bioinformatic analyses. The data have been deposited with links to BioProject accession number PRJNA611545 in the NCBI BioProject database.

**Patient and public involvement statement**

Neither patients nor the public were involved in the design, conduct, reporting, or dissemination of our research.

**Results**

We performed WES and RNA-seq on tissue samples from 57 patients, 55 of whom had not received any treatment. Blood samples were also obtained from 52 of these patients and used as normal controls. For the 5 patients for whom blood samples were unavailable, we used nonneoplastic gastric tissue as controls. We also performed WES on blood samples from an additional 26 patients (Table 2; Supplementary Table S1). The mean coverage for WES was 267 × for the 57 tumor samples, 209 × for the 5 nonneoplastic gastric samples,
Table 2. Clinicopathologic characteristics of Hispanic/Latino patients with gastric cancer in this study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>83</td>
</tr>
<tr>
<td>Sample sequenced Tissue</td>
<td></td>
</tr>
<tr>
<td>Blood only</td>
<td>57 (69%)</td>
</tr>
<tr>
<td>Age (median, IQR, range)</td>
<td>53, 45–61, 23–85</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>54 (65%)</td>
</tr>
<tr>
<td>Female</td>
<td>29 (35%)</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
</tr>
<tr>
<td>Cardia</td>
<td>24 (29%)</td>
</tr>
<tr>
<td>Body</td>
<td>30 (36%)</td>
</tr>
<tr>
<td>Antrum</td>
<td>24 (29%)</td>
</tr>
<tr>
<td>Overlapping</td>
<td>5 (6%)</td>
</tr>
<tr>
<td>Clinical stage</td>
<td></td>
</tr>
<tr>
<td>Stage I (T1-2N0M0)</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Stage II and III (T3-4N0M0, TanyN1-3M0)</td>
<td>46 (55.5%)</td>
</tr>
<tr>
<td>Stage IV (TanyNanyM1)</td>
<td>34 (41%)</td>
</tr>
<tr>
<td>Recurrent</td>
<td>2 (2.5%)</td>
</tr>
<tr>
<td>Differentiation</td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Moderate</td>
<td>18 (22%)</td>
</tr>
<tr>
<td>Moderate/poor</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>Poor</td>
<td>57 (69%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 (7%)</td>
</tr>
<tr>
<td>Lauren classification</td>
<td></td>
</tr>
<tr>
<td>Diffuse</td>
<td>43 (52%)</td>
</tr>
<tr>
<td>Mixed</td>
<td>7 (8.5%)</td>
</tr>
<tr>
<td>Intestinal</td>
<td>26 (31%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>7 (8.5%)</td>
</tr>
<tr>
<td>Helicobacter pylori infection (by histology)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15 (18%)</td>
</tr>
<tr>
<td>No</td>
<td>59 (71%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>9 (11%)</td>
</tr>
</tbody>
</table>

Abbreviation: IQR, interquartile range.

and 67 × 10^9 for the 78 blood samples. For RNA-seq, the average number of reads was 96.9 million, with an average mapping rate of 97.6% for the 57 tumors and 3 nonneoplastic gastric samples.

Consistent with previous reports, the median age at time of diagnosis for the 83-patient Hispanic/Latino patient cohort was younger than that for the 77 Asian and 172 White patients analyzed by the TCGA (53 years vs. 66 and 66, respectively, P < 0.0001, Supplementary Fig. S1A). To confirm that the Hispanic/Latino cohort’s self-reported ancestry was unique from that of the TCGA Asian and White patients, we compared WES data from each of the three groups to reference data available through the Human Genome Diversity Project (HGDP; ref. 13). Using principal component analysis, we found that the Hispanic/Latino cohort clustered independently from the Asian and White patients in the TCGA groups and was related most closely to the HGDP samples from Central and South America (Fig. 1; Supplementary Fig. S1B).

Gastric cancers in Hispanic/Latino patients are enriched for the genomically stable subtype

We next classified the 57 Hispanic/Latino gastric cancer samples into one of the four molecular subtypes established by the TCGA (Supplementary Fig. S2A; ref. 12). We did not include African-Americans in this analysis, as there were only four African-American patients in the TCGA cohort. Tumors were first characterized on the basis of Epstein–Barr virus (EBV) infection status, which was determined bioinformatically with PathoScope 2.0 (14). We found no EBV infections, whereas 10% of the TCGA cohort was EBV-positive (12). Next, microsatellite instability (MSI) was assessed bioinformatically using MSI Sensor, which has previously demonstrated near-perfect concordance with the results of PCR or IHC analysis (15, 16). Three of the 57 samples (5%) had MSI sensor scores of greater than 10, indicating microsatellite instability (Supplementary Fig. S2B). Accordingly, these three samples showed mutation burdens greater than 13 mutations per megabase (Mb), whereas the average mutation burden for the 54 non-MSI samples was 2.5 mutations per Mb.

The remaining samples underwent somatic copy number alteration (SCNA) analysis (10). Seventeen samples (30%) had high SCNA scores, which placed them into the chromosomal instability (CIN) group, and 37 patients (65%) had low scores and were categorized as genomically stable (GS; Fig. 2A). When compared with the Asian (20%) and White (21%) patients, Hispanic/Latinos had a significantly higher proportion of GS tumors (65%, P < 0.001; Fig. 2B). There were no significant differences between Asian and White patients in the proportions of subtypes. CIN samples showed an average of 3.5 mutations per Mb, while GS tumors had 2.0 mutations per Mb. This is consistent with the TCGA data found on the Broad Firehose, which showed CIN and GS samples as having 3.3 and 1.8 mutations per Mb, respectively (http://firebrowse.org/cohort=STAD).

In the TCGA analysis, the GS subtype was found to be enriched for tumors with diffuse-type histology (12). Accordingly, we found that of the 37 GS patients, 78% had diffuse-type, 16% had intestinal-type, and 6% had mixed-type tumors. In contrast, the CIN cohort was comprised of 23.5% diffuse, 53% intestinal, and 23.5% mixed-type tumors (P < 0.001; Fig. 2A).

Hispanic/Latino gastric cancers recapitulate key genomic features identified by the TCGA

Although the Hispanic/Latino samples were significantly enriched for GS tumors, many defining genomic alterations previously identified by the TCGA were recapitulated in the current cohort. For example, the most common recurrent mutation in Hispanic/Latino gastric cancer samples was TP53, as was the case in the TCGA (Fig. 3A). We also found similar structural variations. The TCGA identified CLDN18-ARHGAP fusions in 15% of their GS-type tumors. These rearrangements lead to dysregulated RHOA signaling and loss of an epithelial phenotype (12, 17). Using FusionCatcher and STAR-fusion to evaluate our RNA-seq data, we found that 4 tumors had this rearrangement (Fig. 3B; Supplementary Table S2; refs. 18, 19). All four were GS and diffuse-type. We also observed that 76% of CIN tumors, which were enriched for intestinal-type histology, had amplifications in the 8q24.21 region. This was significantly higher than seen in GS samples, in which only 19% had this copy number abnormality (Fig. 3B; P < 0.001). We confirmed this finding in the TCGA cohort, which similarly showed an enrichment of this structural alteration in CIN samples (12). The 8q24.21 region most notably carries the MYC oncogene, and other groups have noted that amplifications in this region are common in intestinal-type gastric cancers and are associated with worse outcomes in patients with gastric cancer (20–22). Finally, we identified five instances of KRAS amplification (12p12.1), all of which occurred in CIN patients, consistent with previous reports (Fig. 3B; ref. 23).

One major difference between the cohorts is the incidence of PIK3CA mutations, which were mainly found in EBV-type tumors.
that were lacking in the Hispanic/Latino patients. We also found differences in the mutation rates of some key signaling pathway members stratified by molecular subtype (Fig. 3C). We observed that Hispanic/Latino CIN tumors had a lower rate of TP53 mutations (35% vs. 70%) but a higher incidence of APC mutations (29% vs. 10%). In the Hispanic/Latino patients we also found a lower rate of alterations in RHOA (3% vs. 18%, sum of both CIN and GS) and ARID1A (8% vs. 25%, sum of both CIN and GS).

Gene expression profiling is prognostic
To further interrogate the RNA-seq dataset, we selected the top 50 most variably expressed genes and performed unbiased consensus clustering to identify patient subgroups. We found five clusters with distinct clinicopathologic profiles (Fig. 4A; Supplementary Table S3). Patients in clusters 2 and 3 tended to be younger, while clusters 1 and 5 patients were older. Cluster 3 patients had tumors enriched for diffuse-type and GS tumors. When we compared the survival of each cluster, we found the grouping provided significant prognostic capability (Supplementary Fig. S3A, \( P < 0.01 \)). Cluster 1 patients had the shortest median survival at 7.7 months, whereas cluster 4 patients had the longest survival, with median survival not reached. Patients in clusters 2, 3, and 5 had similar survival that were intermediate to clusters 1 and 4. When we grouped clusters 2, 3, and 5 as an intermediate-risk category, its median survival was 19.7 months (Fig. 4B; \( P < 0.001 \)).
Figure 3.
Key genomic features of gastric cancer are identified in Hispanic/Latino (Hs/L) samples. A, Recurrent somatic mutations identified by the TCGA in nonhypermutated gastric cancer samples from Hispanic/Latino patients. B, Structural variations seen in Hispanic/Latino gastric cancer samples. C, Comparison of incidence of somatic alterations in select genes involved in RTK/RAS/PI(3)K signaling, cell cycle, cell adhesion, Wnt signaling, and chromatin remodeling in the TCGA and Hs/L cohorts, stratified by CIN and GS subtypes.
Importantly, the prognostic value of mRNA clustering was maintained when patients were stratified by molecular subtype or by Lauren classification (Supplementary Fig. S3B–S3E; \( P < 0.05 \) for each).

When we performed Gene Set Enrichment Analysis (24) to identify pathways that were uniquely overexpressed in cluster 1 and 4 tumors, we found that the upregulated pathways in cluster 1 were involved in cell-cycle regulation, cell growth, and epithelial–mesenchymal transition (Fig. 4C) while upregulated pathways in cluster 4 were associated with an activated immune response (Fig. 4D).

Hispanic/Latino patients with diffuse-type tumors have frequent germline CDH1 variants

We analyzed the WES data from either blood or nonneoplastic stomach from 83 patients and identified seven germline CDH1 variants (Fig. 5A; Table 3). All seven variants were identified in patients with DGC (16%) and were confirmed with Sanger sequencing (Supplementary Fig. S4A; Supplementary Table S4). Two variants were deletions and five were missense. In patients with DGC, the median age of variant carriers was 41 years (range: 36–54 years) while the median age of CDH1 wild-type (WT) patients was 50 years.

Figure 4. Transcriptomic signatures of gastric cancer from Hispanic/Latino patients are prognostic. A, Unsupervised consensus clustering based on the top 50 most variably expressed genes. B, Kaplan–Meier curves comparing overall survival based on clusters. \( P < 0.001 \). C, Normalized enrichment scores from gene set enrichment analysis comparing cluster 1 to clusters 2, 3, 4, and 5. Orange dots denote Hallmark gene sets related to cell cycle, cell growth, and epithelial–mesenchymal transition, all of which had FDR \( q \)-value \(< 0.01 \). D, Normalized enrichment scores from gene set enrichment analysis comparing cluster 4 to clusters 1, 2, 3, and 5. Red dots denote immune-related Hallmark gene sets, all of which had FDR \( q \)-value \(< 0.01 \).
Hispanic/Latino patients with gastric cancer have high rates of germline CDH1 variants. A, Seven germline CDH1 variants were identified in patients with diffuse gastric cancer. B, Western blot showing E-cadherin expression level upon transfection of plasmids carrying WT CDH1, A286G variant, or G1849A variant into Chinese hamster ovary cells. C, Representative pictures of scratch assays. Distance between the wound edges were measured after 24 hours. D, Quantification of remaining distance between wound edges, relative to 0 h. N ≥ 9 per group, with at least two independent experiments. ***, P < 0.001; ****, P < 0.0001.

(continued)

Pathogenic CDH1 germline variants are known to cause hereditary DGC syndrome. However, none of the germline variant carriers in our Hispanic/Latino cohort had a family history of gastric cancer or lobular breast cancer, which is another manifestation of the mutations (25). Previous reports have suggested that germline CDH1 alterations contribute to early-onset gastric cancer in patients without family histories of cancer (26). We performed a literature search to estimate the rate of germline CDH1 variants in patients with gastric cancer without family histories of gastric cancer, and identified four studies with relatively large cohorts. These included patients from Italy (27), Canada (28), China (29), and Korea (30). Out of 350 patients with DGC, 12 germline mutations in the coding region of CDH1 were identified across 13 patients (3.7%), with 3 patients having deletions, and 10 having missense alterations (Supplementary Table S4). Thus, the prevalence of germline CDH1 variants in patients without a relevant family history was markedly higher in the Hispanic/Latino cohort than what has been reported in other ethnic/racial groups.

To determine whether the identified CDH1 mutations were pathogenic, we first checked the population frequency of these variants in the Genome Aggregation Database (https://gnomad.broadinstitute.org). All seven variants were found in less than 1% of both the general population and in the Latino cohort, which would be consistent with pathogenicity (Table 3). We next queried the annotations of the five missense alterations in the ClinVar database (31). Two were classified as benign (P15 and P20) while the rest were either of uncertain significance or had conflicting data (P30, P50, and P71). Next, we used SIFT and PolyPhen-2 to predict the variant functionality via a bioinformatic approach. Consistent with ClinVar annotation, P15 and P20 were predicted to be benign, but P30, P50, and P71 were projected to be pathogenic (Table 3). Of the 6 patients whose tissue samples were available for analysis, we performed IHC for E-cadherin, and found that there was either decreased (P15, P16, and P33) or near-complete loss (P20 and P50) of protein expression in 5 of the 6 patients, including in P15 and P20, who harbored putatively benign variants (Supplementary Fig. S5).

The variant found in P15, who was a 51-year-old man presenting with locally advanced disease, was a c.286 A>G transition that resulted in an I96V amino acid change. Patient 20, who was a 37-year-old woman presenting with metastatic disease, had an c.1849 G>A change that led to an A617T amino acid change. To test the effects of these variants in vitro, we generated plasmids carrying WT CDH1 or these variants and transfected them into CHO cells, which do not express E-cadherin at baseline and has been used extensively by other groups to test CDH1 variant function (32–34). Sanger sequencing confirmed that the mutations were generated correctly. We found that both 286 A>G and 1849 G>A variants generated protein products that were normal in size and cellular localization (Fig. 5B; Supplementary Fig. S6A). There was no difference in protein expression levels.

E-cadherin is involved in cell–cell adhesion and its loss can result in increased cellular migration. We performed scratch assays to test whether 286 A>G or 1849 G>A affected the migratory ability of CHO cells. After 24 hours, parental CHO cells had completely covered the scratch. As expected, CHO cells expressing WT CDH1 led to significantly reduced cellular migration, with 68% of the...
## Table 3. Germline CDH1 variants found in the Hispanic/Latino gastric cancer cohort.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>Mutation</th>
<th>Chromosome position</th>
<th>Amino acid changes</th>
<th>Mutation type</th>
<th>ClinVar</th>
<th>Predicted effect by SIFT and PolyPhen-2</th>
<th>gnomAD (Latino)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P15</td>
<td>51</td>
<td>Germline Exon3:c.286A&gt;G</td>
<td>I96V</td>
<td>Missense</td>
<td>Benign</td>
<td>Benign</td>
<td>0.016%</td>
<td>0.12%</td>
</tr>
<tr>
<td>P16</td>
<td>36</td>
<td>Germline Exon13:c.1988_2011del, 2012A&gt;C</td>
<td>663-671del EVGDYKINLKLMD-</td>
<td>Deletion - in-frame</td>
<td>N/A</td>
<td>Deleterious</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>P20</td>
<td>37</td>
<td>Germline Exon12:c.1849G&gt;A</td>
<td>A617T</td>
<td>Missense</td>
<td>Benign</td>
<td>Benign</td>
<td>0.45%</td>
<td>0.30%</td>
</tr>
<tr>
<td>P30</td>
<td>41</td>
<td>Germline Exon14:c.2276G&gt;C</td>
<td>G759A</td>
<td>Missense</td>
<td>Uncertain</td>
<td>Deleterious</td>
<td>0.0004%</td>
<td>0.0029%</td>
</tr>
<tr>
<td>P33</td>
<td>40</td>
<td>Germline Exon2:c.135delC</td>
<td>H45fs</td>
<td>Deletion - frameshift</td>
<td>N/A</td>
<td>N/A</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>P34</td>
<td>54</td>
<td>Germline Exon6:c.715G&gt;A</td>
<td>G239R</td>
<td>Missense</td>
<td>Confl</td>
<td>N/A</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>P50</td>
<td>41</td>
<td>Germline Exon16:c.2558C&gt;T</td>
<td>S853L</td>
<td>Missense</td>
<td>Uncertain</td>
<td>Deleterious</td>
<td>0.0016%</td>
<td>0.009%</td>
</tr>
</tbody>
</table>

Abbreviation: N/A, not available.

## Discussion

Hispanic/Latino patients experience significant gastric cancer outcome disparities. Whether there is a molecular basis for these differences is unknown, as previous gastric cancer genomic studies included very few Hispanic/Latino patients (9, 10, 12). To our knowledge, the only study to date that had included a large number of Hispanic/Latino patients was performed by Sahasrabude and colleagues (37). However, their analysis of 333 patients from Latin America was limited to targeted sequencing of only five genes involved in DNA repair.

In this study, we have performed a large, integrated analysis of Hispanic/Latino gastric cancer samples and compared our results to those from Asian patients. Previous studies in other cancer types also identified genomic differences based on ethnicity and race. Shi and colleagues found that more than 50% of Asian patients with non–small cell lung adenocarcinoma had EGR mutations, as compared with 20% of White patients (38). In a study of African-American patients with prostate cancer, Yamash and colleagues identified genomic markers related to race that were highly prognostic (39). Thus, having ethnically and racially representative study cohorts will enhance our understanding of fundamental disease biology and ensure that the efficacy of a selected treatment has been tested and confirmed for the patient’s ethnic/racial background (11, 40).

The high rate of germline CDH1 variants in our Hispanic/Latino DGC cohort is striking. Of the seven mutations we identified, which represented 16% of the patients with DGC, two had not been previously reported in ClinVar, three were annotated as uncertain or conflicting, and two were designated as benign. Thus, these variants would likely have been excluded as pathogenic. However, several lines of evidence indicate that these mutations have deleterious effects. First, previous studies have suggested that germline CDH1 variants may contribute to early-onset DGC (26). The variant carriers in our cohort had a median age of diagnosis of 41 years as compared with patients with DGC with WT CDH1 who had a median age of 50 at diagnosis. Second, E-cadherin protein expression was decreased or lost in 5 of the 6 tumors from CDH1 variant carriers that were available for analysis. Third, in silico analysis predicted that three of the five missense mutations were pathogenic. Finally, functional modeling of the two
missense variants annotated by ClinVar and predicted to be benign by both SIFT and PolyPhen2 demonstrated pathogenic cellular migration phenotype. Our findings speak to the limitations of the currently available tools to predict accurately the pathogenicity of a given variant. When germline CDH1 variants are identified in patients who have a high pretest probability of carrying a pathogenic variant, such as in a young patient with DGC, more rigorous functional testing should be utilized to determine pathogenicity.

Germline CDH1 variants are one of the causes of hereditary DGC syndrome. Because none of the seven Hispanic/Latino CDH1 variant carriers had a family history of gastric cancer or lobular breast cancer, these variants are either de novo or exhibited low penetrance. Previous estimates that carriers of pathogenic CDH1 variants have a lifetime risk of up to 80% of developing DGC are likely overestimations as they are based on families that fulfill the International Gastric Cancer Linkage Consortium (IGCLC) guidelines and thus subjected to ascertainment bias (41). Recent studies that examined DGC penetrance in carriers of pathogenic CDH1 variants that do not fulfill IGCLC criteria indicate a lower lifetime gastric cancer risk. Nicolai and colleagues found a lifetime risk of 37% in their cohort while Roberts and colleagues estimated risk at 42% for men and 33% for women by age 80 (42, 43). These recent reports along with our findings suggest that some germline CDH1 variants require other oncogenic molecular and/or environmental factors to drive DGC formation. This represents an opportunity for precision treatment strategies as we hypothesize that different variants may produce varied biological effects and targets for therapy. Finally, while 5 of our 7 patients would have undergone genetic testing based on IGCLC recommendations to test patients with DGC diagnosed before age 50, we did not meet criteria. A recent study by Low et al. and colleagues found that 65% of CDH1 variant carriers did not meet IGCLC guidelines for testing (44). This suggests that revisions will be necessary to improve the sensitivity of guidelines for genetic testing to identify germline CDH1 carriers.

Previous analyses of early-onset gastric cancer have identified DGC as being associated with young age (12, 45). The high rate of DGC in Hispanic/Latino patients is consistent with the younger age of diagnosis in this cohort. However, the molecular mechanism behind early-onset carcinogenesis in this subgroup is unknown. As discussed above, the high rate of germline CDH1 variants in the Hispanic/Latino cohort may play a role. Previous studies in non-Hispanic/Latino cohorts showed that germline CDH1 variants occur in about 1–3% of nonfamilial patients with gastric cancer (27–30). In addition, the TCGA reported only two CDH1 variants in 295 patients, and these were nonpathogenic (12). Other factors unrelated to Lauren classification and GS subtype clearly affect the etiology of early-onset gastric cancer because three of four very young (<35 years old) patients in our cohort were in the CIN group, with two of them having intestinal-type cancers. This will be an important area for future study, as the incidence of gastric cancer is rising in the United States only among young patients and thus will likely disproportionally affect the Hispanic/Latino population and exacerbate gastric cancer outcome disparities (46).

While molecular classification systems proposed by the TCGA and others have provided new paradigms to study gastric cancers, the practical implications of this scheme for patient care remain elusive. Recently, Sohn and colleagues reported that the TCGA classification may provide both prognostic and predictive value in Korean patients (47). They found that EBV tumors had the best outcomes and GS cancers had the worst. In addition, Kim and colleagues showed that immunotherapy was effective mainly in EBV or MSI-type tumors, while CIN and GS cancers were generally resistant (48). These findings have significant implications for the Hispanic/Latino cohort that we analyzed because 95% of the patients had either CIN or GS tumors. Using consensus clustering of RNA-seq data, we identified transcriptional signatures that were prognostic and thus can aid in risk stratification and treatment planning. Importantly, we found that these signatures were prognostic for subgroups stratified on the basis of both molecular subtypes and by Lauren classification, suggesting of their wide applicability across tumor types. Intriguingly, patients in the low-risk, favorable prognosis group had a gene signature indicative of an activated immune response. Whether these patients will benefit from immune therapy is a possibility that should be tested. Finally, validation in patients of all ancestry will be required to determine whether this signature is ethnicity/race–specific.

Future studies of Hispanic/Latino populations will benefit from more refined definitions of ancestry mix. Hispanic and Latino groups encompass a geographically diverse population exhibiting significant demographic and socioeconomic heterogeneity, due to the differential admixture of European, Indigenous American, and African populations. Previous studies have shown that ancestry proportions in Hispanic/Latino patients are associated with breast cancer incidence and outcomes (49, 50). However, while our study cohort is derived from patients living in North Texas, the country of origin is heterogeneous, as denoted by the relatively broad cluster seen in the ancestry analysis.

In conclusion, while gastric cancer outcome disparities may result from a combination of environmental exposures and socioeconomic factors, inherent tumor biology is also an essential component. Our study analyzing a large cohort of Hispanic/Latino patients with gastric cancer is an important step in addressing the outcome disparity that these patients face by providing a genomic context for their disease. We have found that gastric cancers arising in Hispanic/Latino patients exhibit significantly different genomic landscapes than those developing in Asian and White patients. There is an enrichment for GS tumors and a high rate of germline CDH1 variants. Our findings should be considered in establishing guidelines for screening, genetic counseling, and treatment of Hispanic/Latino patients with gastric cancer.

**Disclosure of Potential Conflicts of Interest**

H. Zhu is a paid consultant for and receives commercial research grant from Twenty-Eight Seven Therapeutics and has ownership interest (including patents) in Ionis. T.H. Hwang is a paid consultant for MEDICALIP, reports receiving commercial research grant from AITRICS and Pfizer, and has ownership interest (including patents) in Patent 1 and Patent 2. M.R. Porembka is an unpaid consultant for Debbie's Dream Foundation Medical Scientific Advisory Board. No potential conflicts of interest were disclosed by the other authors.

**Authors' Contributions**


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.T.G. Hammer, M.R. Porembka

Study supervision: S.C. Wang, D. Agarwal, T.H. Hwang, M.R. Porembka

Published OnlineFirst April 8, 2020; DOI: 10.1158/0008-5472.CAN-19-2918
Gastric Cancer in Hispanic/Latino Patients

Acknowledgments

The authors thank Drs. Herbert J. Zeh III and Todd A. Aguilera for critical reading of the manuscript. The authors acknowledge the assistance of the University of Texas Southwestern Tissue Resource, a shared resource at the Simmons Comprehensive Cancer Center (Dallas TX), which is supported, in part, by the NCI, under award number 5P30CA142543. S.C. Wang was a UT Southwestern Disease-Oriented Scholar and an American College of Surgeons Research Faculty Fellow and was supported by the Society for Surgery of the Alimentary Tract Health Disparity Research Award and the NCI/NIH (K08 CA222611). M.R. Poremba is a Dedman Family Scholar in Clinical Care. I. Nassour was supported by the National Center for Advancing Translational Sciences of the NIH, under Award No. UL1TR001155. T.H. Hwang was supported by the University of Texas Lung Cancer SPORE cancer development grant (P50CA070907), the American Cancer Society Institutional Research Award (ACS-IRG-02-196), and the UT Southwestern Simmons Comprehensive Cancer Center Support Grant (P30CA142543).

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.

Received September 17, 2019; revised February 5, 2020; accepted March 30, 2020; published first April 8, 2020.

References

Wang et al.


Hispanic/Latino Patients with Gastric Adenocarcinoma Have Distinct Molecular Profiles Including a High Rate of Germline CDH1 Variants

Sam C. Wang, Yunku Yeu, Suntrea T.G. Hammer, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-19-2918

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2020/04/08/0008-5472.CAN-19-2918.DC1

Cited articles
This article cites 47 articles, 11 of which you can access for free at:
http://cancerres.aacrjournals.org/content/80/11/2114.full#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/80/11/2114.
Click on “Request Permissions” which will take you to the Copyright Clearance Center’s (CCC) Rightslink site.