Contents available at **PubMed**
www.clinicalandtranslationalinvestigation.com

Check for updates

Rev Invest Clin. 2021;73(3):172-81 ORIGINAL ARTICLE

Predicting Pathogenicity of *CDH1* Gene Variants in Patients with Early-onset Diffuse Gastric Cancer from Western Mexico

Azaria García-Ruvalcaba[1,2], Lourdes del C. Rizo de la Torre[3], María T. Magaña-Torres[1], Ernesto Prado-Montes-de-Oca[4,5,6], Andrea V. Ruiz-Ramírez[1,2,4], Héctor Rangel-Villalobos[7], José A. Aguilar-Velázquez[2,7], Andrea M. García-Muro[1,2], and Josefina Y. Sánchez-López[1]*

[1]Genetics Division, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jal.; [2]Doctorate Program in Human Genetics, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara, Jal.; [3]Molecular Medicine Division, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Jal.; [4]Laboratory of Regulatory SNPs, Personalized Medicine Laboratory, Medical and Pharmaceutical Biotechnology, Research and Assistance Center in Technology and Design of Jalisco A.C. (CIATEJ AC), Consejo Nacional de Ciencia y Tecnología (CONACyT), Guadalajara, Jal.; [5]Scripps Research Translational Institute, La Jolla, California, USA; [6]Integrative Structural and Computational Biology, Scripps Research La Jolla, California, USA; [7]Department of Medical and Life Sciences, Instituto de Investigación en Genética Molecular, Centro Universitario de la Ciénega, Universidad de Guadalajara, Guadalajara, Jal., Mexico

ABSTRACT

Background: Early-onset diffuse gastric cancer (EODGC) occurs at or before 50 years of age. Pathogenic mutations and germline deletions in the *CDH1* gene (E-cadherin) are well-documented genetic factors associated with the causes of EODGC. **Objective:** The objective of the study was to study *CDH1* germline variants and their potential functional impact in patients with EODGC in a Mexican population. **Methods:** We studied seven EODGC patients from a biomedical research center in western Mexico. Variants were identified by Sanger sequencing and multiplex ligation-dependent probe amplification. The DeepSEA and SNPClinic v.1.0 software and the Ensembl (1000 Genomes Project, 1kGP) and ClinVar databases were used to predict functional single-nucleotide polymorphisms (SNPs). The genetic admixture of the Mexican patients was corroborated by 22 short tandem repeat loci genotyping and structure analysis. **Results:** We found 12 germline *CDH1* variants in all EODGC patients, and all of them are considered as polymorphisms: rs34561447, rs5030625, rs16260, rs1330727101, rs28372783, rs942269593, rs3743674, rs1801552, rs34939176, rs33964119, rs3556654, and rs1801026. The prediction of regulatory SNPs in the promoter suggests a role for a retrovirus in EODGC that induces the transcription of interferon-related genes through toll-like receptor-interferon response factor 3 signaling, as three SNPs in the *CDH1* promoter alter three binding sites for this transcription factor. In addition, SNPs rs28372783 and rs1801026 could alter upstream stimulatory factors 1 (USF1)/USF2-mediated telomerase-dependent lymphocyte activation in EODGC. Other interesting result is a CTCF-dependent shorter *CDH1* isoform lacking exon 14, probably due to exon-skipping mediated by rs33964119. **Conclusions:** Classical pathogenic germline mutations in the *CDH1* gene were not found in these 7 EODGC patients. However, the in silico approaches revealed the possible involvement of a retrovirus and a shorter E-cadherin isoform in EODGC. Nevertheless, further *in vitro* and in vivo assays are needed to confirm these predictions. (REV INVEST CLIN. 2021;73(3):XX-XX)

Key words: Early-onset diffuse gastric cancer. *CDH1* gene. E-cadherin. Gastric cancer. Variants. Bioinformatics.

***Corresponding author:**
Josefina Y. Sánchez-López
E-mail: yosalo1795@yahoo.com

Received for publication: 21-09-2020
Approved for publication: 04-12-2020
DOI: 10.24875/RIC.20000470

INTRODUCTION

Early-onset gastric cancer (EOGC) is defined as any GC that occurs at the age of 50 years or earlier. EOGC comprises approximately 10% of all patients with GC, and their reported frequencies vary between 2.7% and 15% in various studied populations[1,2]. Germline pathogenic variants of the *CDH1* gene are well-documented genetic factors associated with early-onset diffuse gastric cancer (EODGC)[1,2]. The *CDH1* gene, located on chromosome 16q22.1, has 16 exons and a length of 98,250 bp. The most common isoform of the protein encoded by this gene is translated from a 4.5-kb RNA transcript. The protein E-cadherin, encoded by the *CDH1* gene, is a cell adhesion molecule involved in the maintenance and homeostasis of normal epithelial tissue[3].

Worldwide, GC is the fifth leading cause of cancer mortality, with 8.2 cases/100,000[4]. Although GC is one of the main causes of mortality by cancer in the world, few studies have investigated *CDH1* variants in EODGC patients in the Mexican population[5-9]. To the best of our knowledge, the only variants in the *CDH1* gene that has been reported so far in Mexican patients with EODGC are c.377del and the SNP rs16260. Our objective was to study *CDH1* germline variants and their potential functional impact in patients with EODGC in a Mexican population.

METHODS

We studied seven patients (five men and two women) with EODGC. A diagnosis was made by a histopathologist who analyzed the histopathology of the gastric tumors obtained by endoscopy as part of their medical diagnosis (all patients had diffuse-type tumors and exhibited signet-ring cells). The patients were recruited by the Gastroenterology Department of the Hospital de Especialidades at Centro Médico Nacional de Occidente of Instituto Mexicano del Seguro Social located in Guadalajara City, Mexico. Patients were invited to participate in the study on a consecutive basis if they met three criteria: (i) the patients and their parents were born in Western Mexico; (ii) the patients were unrelated to each other; and (iii) they were of Mexican Mestizo ethnicity. Those patients who agreed to participate signed an informed consent letter. An Institutional Review Board and Ethics Committee approved the study. The mean age of the patients was 39.7 years with a range of 22-48 years (we chose an age < 50 years for this study, similar to Corso *et al.*[1]). Three patients met the criteria for suspected hereditary diffuse gastric cancer because they were younger than 40 years of age and had tumors of diffuse histology according to the International Gastric Cancer Linkage Consortium[10]. Furthermore, two patients had a family history of cancer (the father of patient five died of prostate cancer and the mother of patient seven had lung cancer) (Table 1).

Genomic DNA samples were obtained from peripheral blood leukocytes by the salting out method. *CDH1* variants were identified by polymerase chain reaction (PCR), followed by Sanger sequencing. Eighteen fragments were amplified, which included 16 exons, the promoter, and the 3'UTR region of the *CDH1* gene. The primers used to amplify exons 2, 6, 7, 9, 10, 13, and 16 were previously described by Corso *et al.*[1] The remaining primers (the promoter region and exons 1, 3, 4, 5, 8, 11, 12, 14, and 15) were designed by our group (sequencing primers and conditions can be provided on request). Sanger sequencing is a robust testing strategy able to determine whether a point mutation or a small deletion/duplication is present. PCR amplification followed by sequencing is considered the diagnostic gold standard. A Ready Reaction Big Dye Terminator kit v. 3.1 (Applied Biosystems, Foster City, CA, USA) was used for sequencing. Multiplex ligation-dependent probe amplification (MLPA) analyses can reveal large deletions and rearrangements not detectable by sequencing, this technique was employed for the identification of large exonic deletions in the *CDH1* gene using the SALSA® MLPA® Probemix P083-C1 *CDH1* kit C1-0114 following the manufacturer's recommendations (MRC-Holland, the Netherlands). An ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) was used for capillary electrophoresis; the ChromasLite v.2.6.6 and Coffalyzer programs were used for data analysis.

The prediction of regulatory single-nucleotide polymorphisms (SNPs) in the promoter region of the *CDH1* gene was performed with DeepSEA and SNPClinic v.1.0 software. DeepSEA is a deep learning-based algorithm that can accurately predict the epigenetic state of a sequence, including transcription

Table 1. Characteristics of Mexican early-onset diffuse gastric cancer patients and germline *CDH1* variants found per patient and predicted differentiated molecular mechanisms

Patient	Age	Sex	Blood group	History of cancer	Smoking (>6 cigarettes/ day)	Alcoholism (>1 day/ week)	c.-612_ –611insA	c.-472 delA	c.-285 C>A	c.-273 G>A	c.-197 A> C/G	c.-146 C> G/T	c.48+ 6C>T	c.2076T >C A692A	c.2164+ 17_2164+ 18insA	c.2253 C> T N751N	c.2439+ 177delT	c.*54 C> T/A/G	MLPA analysis	Predicted differentiated molecular mechanism
1	38	M	O+	No	Yes	Yes	–/insA				A/C		C/T	T/C			–/insT	C/T	N	Alteration in telomerase activity
2	44	M	O+	Nd	Nd	Nd	–/insA				A/C		C/T	T/C	–/insA	C/T	–/insT	C/T	N	Alterations in telomerase activity and in noncoding RNA
3	48	M	Nd	No	No	Yes		A/delA						C/C	–/insA	C/T	–/insT	C/T	N	Alterations in noncoding RNA and histones
4	33	F	A–	No	No	Yes	–/insA						C/T	C/C	–/insA	C/T	–/insT	T/T	N	Alterations in noncoding RNA
5	47	M	A+	Yes[a]	Yes	Yes	–/insA		C/A				T/T	T/C			–/insT		N	Alterations of MZG1 binding site
6	22	F	O+	Yes[b]	Yes	Yes	–/insA				A/C		C/T	C/C		C/T	–/insT		Nd	Alteration in splicing and/ or antisense–mediated decay
7	46	M	O+	Yes[c]	No	No	–/insA			G/A		C/G	T/T	T/C	–/insA	C/T	–/insT		N	If chromatin is accessible, alteration of IRF3, NFYA, and/or SP2 transcription factor binding sites

[a]Prostate cancer in the father of the patient.
[b]Cervical cancer in the patient.
[c]Lung cancer in the mother of the patient.
F: female; M: male; N: Normal; Nd: no data.

factor binding, DNase I sensitivities, and histone marks in multiple cell types and can further utilize this capability to predict chromatin effects of sequence variants and prioritize regulatory variants[11]. SNPClinic v.1.0 software calculates the impact of SNPs on the alteration of transcription factor binding sites (TFBSs), according to the JASPAR database, when chromatin is accessible in the input cell line/tissue[12]. To perform the SNPClinic analysis, the following 14 ENCODE cell lines potentially involved in inflammation, carcinogenesis and/or metastases (not only gastric cancer) were tested for DNase I HUP chromatin accessibility: Caco-2 (colorectal adenocarcinoma), H1 hesc and H7 hesc (embryonic stem cells), Hct116 (colon cancer), Hepg2 (liver cancer), Hpde6e6e7 (pancreatic duct), Hvmf (connective tissue), Osteobl (osteoblasts), Be2c (bone marrow), Medullo (brain), Th0, THh1, and Th2 (T helper lymphocytes), and chronic lymphocytic leukemia. To filter out SNPs not impacting TFBSs, only putative TFBSs that had relative binding scores (RBSs) ≥ 0.8 in the major allele were selected as binding TFs. Student's t-test with $p \leq 0.05$ on the null hypothesis was used to test whether the list of RBSs above the threshold was equal to those RBSs below the threshold. As an additional filtering step, only regulatory SNPs (rSNPs) with a functional impact factor (homotypic redundance weight factor × ΔRBS) ≥ an absolute value of ten were considered true positive rSNPs, according to our previous validation[12]. Because SNPClinic v.1.0 was validated for proximal promoters for SNPs located in introns, exons, and 3'UTR regions, and due to the effect of insertions/deletions, we used the DeepSEA software[11] and both the Ensembl[13] and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) databases. For the DeepSEA software, the most important chromatin features were selected by first applying a filtering threshold of an E-value < 0.01 and then by applying a threshold of log two-fold change of 0.02.

We corroborated the ancestry of Mexican EODGC patients by means of PCR genotyping of 22 autosomal short tandem repeat (STRs) with the PowerPlex® Fusion System (Promega Corp., Madison, WI, USA), followed by capillary electrophoresis in the ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Genotype assignment was performed with allelic ladders assisted by GeneMapper v.3.2 software (Applied Biosystems, Foster City, CA). The admixture analysis based on STR genotype was performed with Structure[14,15]. For this purpose, STR population databases that included Mexican Native Americans[16], as well as Europeans and Africans from the USA, were employed as ancestral references[17]. The structure parameters employed herein offered consistent admixture estimates regarding those based on AIMs and genome-wide SNPs, as previously demonstrated in Mexican populations[18].

RESULTS

The genetic admixture of the 7 Mexican EODGC patients mainly included European (73-87%) and Native American (8-22%) ancestries (Supplementary Figure S1). Because these results are in agreement with the previous descriptions of the Mexican population[19], further discussion of this finding will be omitted.

All EODGC patients presented from 5 to 8 germline *CDH1* gene variants, and all patients had at least one variant in the promoter region. A total of 12 different variants were identified in the *CDH1* gene, all of which are already known as SNPs: six in the promoter regions c.–612_-611insA (rs34561447), c.–472delA (rs5030625), c.–285C>A (rs16260), c.–273G>A (rs1330727101), c.–197A>C (rs28372783), and c.–146C>G (rs942269593); two in exons 13 and 14 c.2076T>C (rs1801552) and c.2253C>T (rs33964119); three in introns 1, 13, and 15 c.48+6C>T (rs3743674), c.2164+17_2164+18insA (rs34939176), and c.2439+177delT (rs3556654); and one in 3'UTR c.*54C>A (rs1801026) (Table 2).

The most frequent variants were c.2076T>C (A692A) and c.2439+177delT, which were observed in all subjects. The variants c.–472delA, c.–285C>A, c. –273G>A, and c. –146C>G, located in the promoter region, were found in only one patient (Table 2). The allelic frequencies of these variants reported in other populations are listed in table 3.

Conclusive results of the MLPA analysis were obtained for only six subjects because one DNA sample was of poor quality (Table 1). No deletions or duplications of the *CDH1* gene were observed in any of the six patients, since the amplified probes were observed within a radius of 1.

Table 2. *CDH1* variants found in seven early-onset diffuse gastric cancer patients and their putative functional impact according to ClinVar and Ensembl (ENCODE project data) databases and *in silico* prediction with SNPClinic and DeepSEA software

Gene region	Variant	Rs	Genotype frequencies (n = 7)		Clinvar/ Franklin*	Ensembl cell line: regulatory activity	SNPClinic cell line (TF:FIF)	DeepSEA TF**: E-value
Promoter	c. −612_−611insA	rs34561447	−|−:	0.143 (1)	NR/B	CTCF binding site (insulator)	NA	H2AZ, H3K27ac, H3K23ac
			−|ins(A)$^{9-12}$:	0.857 (6)				
			insA|insA	0.000 (0)				
	c. −472delA	rs5030625	A|A:	0.857 (6)	NR/B		NA	H4K5ac, H3K27ac, H2AK5ac
			A|delA:	0.143 (1)				
			dela|delA:	0.000 (0)				
	c. −285C>A	rs16260	C|C:	0.857 (6)	B/B	HCT116, PC9, NPC_2 and osteoblasts: Repressed	Osteoblasts MZF1: −2.76 FOSL2: −1.24 MAFG::NFE2L1:−1.04	IRF3, NF-YA, C-Fos
			C|A:	0.143 (1)				
			A|A:	0.000 (0)				
	c. −273G>A	rs1330727101	G|G:	0.857 (6)	NR/VUS	–	NA	IRF3, SP2
			G|A:	0.143 (1)				
			A|A:	0.000 (0)				
	c. −197A>C/G	rs28372783	A|A:	0.571 (4)	LB/B	–	NA	USF1, USF2, NF-E2
			A|C:	0.429 (3)				
			C|C:	0.000 (0)				
	c. −146C>G/T	rs942269593	C|C:	0.857 (6)	NR/VUS	–	NA	IRF3, NF-YA, SP2
			C|G:	0.143 (1)				
			G|G:	0.000 (0)				
Intron 1	c.48+6C>T	rs3743674	C|C:	0.143 (1)	B/B	SR, AMD, lncRNA	NT	ZEB1, TAF7, HDAC6
			C|T:	0.571 (4)				
			T|T:	0.286 (2)				
Exon 13	c.2076T>C A692A	rs1801552	T|T:	0.000 (0)	B/B	SR, AMD, lncRNA	NT	TAF1, H2BK12ac, JunD
			T|C:	0.571 (4)				
			C|C:	0.429 (3)				

(Continues)

Table 2. *CDH1* variants found in seven early-onset diffuse gastric cancer patients and their putative functional impact according to ClinVar and Ensembl (ENCODE project data) databases and *in silico* prediction with SNPClinic and DeepSEA software *(continued)*

Gene region	Variant	Rs	Genotype frequencies (n = 7)		Clinvar/ Franklin*	Ensembl cell line: regulatory activity	SNPClinic cell line (TF:FIF)	DeepSEA TF**: E-value
Intron 13	c.2164+17_2164 +18insA	rs34939176	–|–:	0.429 (3)	B/B	SR, AMD, lncRNA	NT	MafK, Bach1, NF-E2
			–|ins(A)$^{2-3}$:	0.571 (4)				
			ins(A)$^{2-3}$ |ins(A)$^{2-3}$	0.000 (0)				
Exon 14	c.2253C>T/G/A	rs33964119	C|C:	0.286 (2)	B/B	Nonsynonymous only when G or A	NT	CTCF, Rad21, CTCFL
	T (N751N)		C|T:	0.714 (5)				
			T|T:	0.000 (0)				
Intron 15	c.2439+177delT	rs35566564	–|–:	0.000 (0)	NR/B	IR, lncRNA, SR, AMD	NT	TFIIIC-110, RPC155, MBD4
			–| Ins(T)^{7}:	1.000 (7)				
			ins Ins(T)^{7} Ins(T)^{7}|:	0.000 (0)				
3'UTR	c.*54C>T/A/G	rs1801026	C|C:	0.429 (3)	B/B	IR, SR, AMD	NT	USF1, USF2, Max
			C|T:	0.429 (3)				
			T|T:	0.143 (1)				

*Genoox's Franklin tool (https://franklin.genoox.com/clinical-db/home).
**According to the DeepSEA ranking method, the top three features were selected.
AMD: antisense-mediated decay; B: benign; Bach1: transcription regulator protein BACH1; C-Fos: proto-oncogene c-Fos; CTCF: transcriptional repressor CTCF; CTCFL: transcriptional repressor CTCFL; FIF: functional impact factor; H2AK5ac: acetylation at lysine 5 histone H2A; H2AZ: variant histone H2A; H2BK12ac: acetylation at lysine 12 histone H2B; H3K23ac: acetylation at lysine 23 histone 3; H3K27ac: acetylation at lysine 27 histone 3; H4K5ac: acetylation at lysine 5 histone 4; HDAC6: histone deacetylase 6; IR: intron retention; IRF3: interferon response factor 3; JunD: transcription factor jun-D; LB: Likely benign; lncRNA: long noncoding RNA; MafK: transcription factor MafK; Max: protein max; MBD4: methyl-CpG-binding domain protein 4; NA: non-accessible; NF-E2: transcription factor NF-E2 45 kDa subunit; NF-YA: nuclear transcription factor Y alpha; NR: not reported; NT: not tested; Rad21: double-strand-break repair protein rad21 homolog; RPC155: RNA polymerase III subunit RPC155-C-; SP2: transcription factor specificity protein 2; SR: splicing region; TAF1: transcription initiation factor TFIID subunit 1; TAF7: transcription initiation factor TFIID subunit 7; TF: transcription factor; TFIIIC-110: general transcription factor IIIC polypeptide 2 (beta subunit, 110 kD); USF1; USF2: upstream stimulatory factors 1 and 2; VUS: Variant of uncertain significance; ZEB1: zinc finger E-box-binding homeobox 1.

Table 3. Minor allele frequencies of polymorphisms found in Mexican patients with early-onset diffuse gastric cancer according to the 1kGP and gnomAD databases in five super populations

Rs code	Allele	1kGP (Phase 3)				
		AFR	AMR	EAS	EUR	SAS
rs34561447	A_{12}=	0.087	0.011	0	0	0
rs5030625	A=	0.363	0.223	0.23	0.119	0.207
rs16260	A=	0.126	0.248	0.308	0.281	0.255
rs1330727101*	A=	0.0002	0	0	0	0
rs28372783	C=	0.016	0.085	0.11	0.017	0.025
rs942269593*	T=	0.00007	0	0	0	0
rs3743674	C=	0.360	0.223	0.231	0.119	0.206
rs1801552	T=	0.062	0.415	0.355	0.355	0.331
rs34939176	AAA=	0.002	0.072	0.07	0.044	0.064
rs33964119	T=	0.058	0.069	0.069	0.035	0.044
rs35566564	No population frequencies available					
rs1801026	T=	0.2	0.146	0.155	0.161	0.093

*Data from The Genome Aggregation Database (gnomAD). No evidence in 1kGP. Rs: reference SNP; 1kGP: the 1000 Genomes Project; AFR: African population; AMR: ad mixed American population; EAS: East Asian population; EUR: European population; SAS: South Asian population.

The putative functional impact of the identified variants in the EODGC patients is shown in table 2. rSNPs reveal which TFBSs are putatively affected, thereby decreasing or increasing the affinity of DNA to the TF (Table 1). Our results reveal that SNP c.-612_611insA (rs34561447) alters a TFBS for CTCF. In addition, flanking this SNP, we found two active chromatin signatures, H3K27Ac and H3K23Ac. The H3K27Ac signature also overlapped with the SNP c.-472delA (rs5030625) for which we found additional markers of active chromatin, such as H4K5Ac and H2AK5Ac. Notably, this signature was undetected by the SNPClinic software because it was designed to detect DNase I sites rather than histone acetylation.

The SNP c.−285C>A (rs16260) alters TFBS for myeloid zinc finger gene 1 (MZF1, formerly ZNF42 or MZF1B), interferon response factor 3 (IRF3), NF-YA, and c-Fos. Furthermore, we observed that *CDH1* was downregulated in the 14 ENCODE cell lines in which c.−285C>A was found, including osteoblasts, Hct116 (colorectal cancer), PC9 (non-small-cell lung cancer), and NPC_2 (nasopharyngeal carcinoma) cell lines, as well as in the four cell lines observed in ENSEMBL. The SNP c.−273G>A (rs1330727101) modifies at least three TFBSs for IRF3 and specificity protein 2 (SP2) (Table 2).

DISCUSSION

The molecular basis for EODGC has not yet been completely elucidated. Although alterations in genes, such as *CDH1*, have been reported, germline mutations in the *CDH1* gene are less frequent than somatic mutations[20]. Pathogenic germline mutations occur in up to 8% of EODGC cases[2]. However, in most cases, the germinal variants are nonpathogenic or are of uncertain significance[1,20,21].

In our study, *CDH1* pathogenic mutations in EODGC patients were ruled out; however, we identified 12 different germinal variants previously reported as SNPs (rs34561447, rs5030625, rs16260, rs1330727101, rs28372783, rs942269593, rs3743674, rs1801552, rs34939176, rs33964119, rs3556654 and rs1801026) that could contribute to the phenotypes of these patients. Some of them have been associated with pathological processes, for example: the variant rs34561447 has been found in patients with non-small cell lung cancer with a low frequency[22]; the variant rs1801552 has been identified in various pathologies such as orofacial clefts[23], primary infertility[24], and colorectal cancer[25]; and the rs1801026 was associated to poorer survival in breast cancer patients[26].

Regarding the identified variants in regulatory regions, the SNP c.−612_-611insA (rs34561447) alters a TFBS for the protein CTCF, which delimits 3D boundaries of insulators by mediating chromatin looping between its binding sites. Two active chromatin signatures were also found flanking this SNP: H327Ac, which is present in active enhancers[27] and H3K23Ac, which is recognized by both the oncoprotein TRIM24[28] and monocytic leukemic zinc finger-related factor[29].

The SNP c.−285C>A (rs16260) alters the TFBS for MZF1, which is reported as tumorigenic for GC; even the mRNA levels of MZF1 have been proposed as a prognostic marker[30]. SNPClinic analysis revealed at least 8 TFBSs for MZF1 in the proximal promoter (−2 kb) of *CDH1*, providing stronger evidence that MZF1 regulates this gene. SNP rs16260 also alters the TFBS for the proto-oncogene c-Fos. One function of c-Fos is the induction of epithelial-mesenchymal transition (EMT), at least in colorectal cancer[31]. The transcription factor interferon response element 3 (IRF3) activates transcription of interferon-related genes with antiviral activity against double-stranded DNA and sRNA viruses in a toll-like receptor 3-dependent manner[32]. This result is interesting, because even though Epstein-Barr virus and hepatitis B virus have been associated with GC and precancerous lesions, respectively[33,34], their genomes comprise double-stranded DNA, not sRNA. To date, the only sRNA virus known to be related to GC is the retrovirus human T-cell lymphotropic virus type I; surprisingly, infection by this virus diminishes the relative risk for this type of cancer[35]. rs16260 has been reported as a risk factor for prostate cancer[36]; in our study, this variant was observed in a patient whose father had prostate cancer (Table 1).

The SNP c.−273G>A (rs1330727101) modifies at least 3 TFBSs for IRF3 and specificity protein 2 (SP2). SP2 is a ubiquitous factor that binds to GC boxes and has been suggested as a regulator of T-cell antigen receptor α[37]. SP2 is also important in the phenotype maintenance of T helper 17 cells (Th17), which play an important role in maintaining mucosal barriers and contributing to pathogen clearance at mucosal surfaces[38].

The SNP c.−197A>C (rs28372783) altered the TFBS for upstream stimulatory factors 1 and 2 (USF1 and USF2). USF2 and a truncated isoform were shown to have a dominant-negative effect on telomerase reverse transcriptase (*TERT*) expression and on overall telomerase activity during lymphocyte activation[39]. Furthermore, USF1 and USF2 regulate other functions of the TERT enzyme, such as angiogenesis, inflammation, cancer cell stemness, and EMT. These extended functions are relevant in the dynamics and homeostasis of the tumor microenvironment[40]. In addition, in the presence of c.-197A>C, nuclear factor erythroid 2 (NF-E2) is overexpressed in patients with myeloproliferative neoplasms[41].

The SNP c.−146C>G (rs942269593) changes the TFBS for IRF3, NF-YA, and SP2. The impact of this SNP could be quite similar to that of the abovementioned c.−273G>A (rs1330727101). The additional feature of this SNP is that NFYA binding motifs (CCAAT boxes) are enriched (high homotypic redundancy) in the promoters of overexpressed genes in breast, colon, thyroid, and prostate carcinomas[20].

Regarding the analyzed SNPs in non-regulatory regions, these SNPs were in the ClinVar categories "benign" and "likely benign;" however, Ensembl showed that these SNPs overlap with sequences of transcripts involved in antisense-mediated decay and splicing alteration, including but not limited to intron retention and long non-coding RNA (lncRNA) alteration. A previous effort was made to understand the role of the lncRNA-miRNA-mRNA network in GC; however, the report did not consider the effects of SNPs[42]. To the best of our knowledge, there is no available *in silico* tool to quantify the effect of these SNPs in the overlapping of these long non-coding regions; therefore, additional research on this topic is needed.

The DeepSEA software gave at least two relevant predictions. First, the CTCF binding site is altered when the synonymous variant c.2253C>T (rs33964119) is present. When CTCF is not bound to its target DNA sequence, the RNA elongation rate is accelerated and can result in exon exclusion and alternative splicing[43]. The Ensembl database includes two shorter *CDH1* isoforms that lack exon 14, comprising 647 residues; therefore, it is possible that this variant could cause skipping of exon 14 by a previous alteration in the CTFC binding site. However, this was not confirmed in our studied EODGC patients because in their case, the sequence of exon 14 can be deleted in mRNA and protein but not deleted in DNA. The

second relevant prediction was that SNP c.*54C>T (rs1801026) alters the TFBS for USF1 and USF2, resulting in similar mechanisms as the promoter SNP c.−197A>C (rs28372783).

In the Mexican EODGC patients investigated in this study, predictive bioinformatic analysis presented a plausible explanation of the potential differentiated molecular mechanisms for the phenotypes observed in each of the patients (Table 1). Our results indicate that pathogenic *CDH1* germline variants are not common in EODGC, suggesting that the variants observed in these EODGC patients can contribute to the phenotypes in the patients, and we must consider that other genes, such as *ARID1A* or *RHOA*, can be carriers of pathogenic mutations, as has been previously suggested[44].

Regarding searching for deletions or duplications in the *CDH1* gene, these germinal alterations were not identified in the studied EODGC patients. Other studies have also not found deletions in this gene, including a study of 25 Korean EODGC patients[2] and a study of 88 Brazilian EOGC patients[45].

Finally, the number of patients is quite limited for drawing conclusions about the exact molecular etiology of EODGC. However, these *in silico* results are interesting and have strong *in vitro* support, primarily because SNPClinic and DeepSEA are programs that use ENCODE, Roadmap Epigenomics, chromatin profiles and JASPAR databases as inputs, which are supported by a large number of *in vitro* experiments. The findings of this study will be validated in future in vitro and in vivo investigations.

ACKNOWLEDGMENTS

Azaria García Ruvalcaba, Andrea V. Ruiz Ramírez, Andrea M. García Muro, and José A. Aguilar-Velázquez received scholarships from CONACYT, Mexico. The authors would like to thank Andrea Rebeca Bustos Carpinteyro PhD., for her contribution in this study.

SUPPLEMENTARY DATA

Supplementary data are available at Revista de Investigación Clínica online (www.clinicalandtranslationalinvestigation.com). These data are provided by the corresponding author and published online for the benefit of the reader. The contents of supplementary data are the sole responsibility of the authors.

REFERENCES

1. Corso G, Pedrazzani C, Pinheiro H, Fernandes E, Marrelli D, Rinnovati A, et al. E-cadherin genetic screening and clinicopathologic characteristics of early onset gastric cancer. Eur J Cancer. 2011;47:631-9.
2. Kim S, Chung J, Jeong T, Park YS, Lee JH, Ahn JY, et al. Searching for E-cadherin gene mutations in early onset diffuse gastric cancer and hereditary diffuse gastric cancer in Korean patients. Fam Cancer. 2012;12:503-7.
3. Graziano F. The E-cadherin gene, structure and function. In: Corso G, Roviello F, editors. Spotlight on Familial and Hereditary Gastric Cancer. Dordrecht: Springer; 2013. p. 27-33.
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394-424.
5. Medina-Franco H, Barreto-Zuñiga R, García-Alvarez MN. Preemptive total gastrectomy for hereditary gastric cancer. J Gastrointest Surg. 2007;11:314-7.
6. Medina-Franco H, Medina AR, Vizcaino G, Medina-Franco JL. Single nucleotide polymorphisms in the promoter region of the E-cadherin gene in gastric cancer: case-control study in a young Mexican population. Ann Surg Oncol. 2007;14:2246-9.
7. Ramos de la Medina A, More H, Medina-Franco H, Humar B, Gamboa A, Ortiz LJ, et al. Single nucleotide polymorphisms (SNPs) at *CDH1* promoter region in familial gastric cancer. Rev Esp Enferm Dig. 2006;98:36-41.
8. Slavin T, Neuhausen SL, Rybak C, Solomon I, Nehoray B, Blazer K, et al. Genetic gastric cancer susceptibility in the international clinical cancer genomics community research network. Cancer Genet. 2017;216-217:111-9.
9. Martínez Valenzuela C, Castelán-Maldonado EE, Carvajal-Zarrabal O, Calderón-Garcidueñas AL. First report of a Mexican family with mutation in the *CDH1* gene. Mol Genet Genomic Med. 2020;8:e1208.
10. van der Post R, Vogelaar I, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, et al. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH-1mutation carriers. J Med Gen. 2015;52:361-74.
11. Zhou J, Troyanskaya O. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Met. 2015; 12:931-4.
12. Flores-Saiffe Farías A, López EJ, Vázquez CM, Li W, Prado-Montes de Oca E. Predicting functional regulatory SNPs in the human antimicrobial peptide genes DEFB1 and CAMP in tuberculosis and HIV/AIDS. Comput Biol Chem. 2015;59:117-25.
13. Zerbino D, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:D754-61.
14. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945-59.
15. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003;164:1567-87.
16. Aguilar-Velázquez JA, Martínez-Cortés G, Inclán-Sánchez A, Favela-Mendoza AF, Velarde-Félix JS, Rangel-Villalobos H. Forensic parameters and admixture in Mestizos from five geographic regions of Mexico based on 20 autosomal STRs (powerplex 21 system). Int J Legal Med. 2018;132:1293-6.
17. Hill CR, Duewer DL, Kline MC, Coble MD, Butler JM. U.S. population data for 29 autosomal STR loci. Forensic Sci Int Genet. 2013;7:82-3.
18. Aguilar-Velázquez JA, Locia-Aguilar G, López-Saucedo B, Deheza-Bautista S, Favela-Mendoza AF, Rangel-Villalobos H. Forensic parameters and admixture in seven geographical regions of the Guerrero state (South, Mexico) based on STRs of the Globalfiler® kit. Ann Hum Biol. 2019;45:524-30.
19. Rubí-Castellanos R, Martínez-Cortes G, Muñoz-Valle JF, González-Martin A, Cerda-Flores RM, Anaya-Palafox M, et al. Pre-hispanic Mesoamerican demography approximates the

present-day ancestry of Mestizos throughout the territory of Mexico. Am J Phys Anthropol. 2009;139:284-94.
20. Cho C, Jung J, Jiang L, Lee EJ, Kim DS, Kim BS, et al. Combinatory RNA-sequencing analyses reveal a dual mode of gene regulation by ADAR1 in gastric cancer. Dig Dis Sci. 2018; 63:1835-50.
21. Oliveira C, Suriano G, Ferreira P, Canedo P, Kaurah P, Mateus R, et al. Genetic screening for familial gastric cancer. Hered Cancer Clin Pract. 2004;2:51-64.
22. Leng S, Bernauer AM, Zhai R, Tellez CS, Su L, Burki EA, et al. Discovery of common SNPs in the miR-205/200 family-regulated epithelial to mesenchymal transition pathway and their association with risk for non-small cell lung cancer. Int J Mol Epidemiol Genet. 2011;2:145-55.
23. Hozyasz KK, Mostowska A, Wójcicki P, Lasota A, Offert B, Balcerek A, et al. Nucleotide variants of the cancer predisposing gene CDH1 and the risk of non-syndromic cleft lip with or without cleft palate. Fam Cancer. 2014;13:415-21.
24. Kang S, Li Y, Li B, Wang N, Zhou RM, Zhao XW. Genetic variation of the E-cadherin gene is associated with primary infertility in patients with ovarian endometriosis. Fertil Steril. 2014; 102:1149-54.
25. Fernández-Rozadilla C, de Castro L, Clofent J, Brea-Fernández A, Bessa X, Abulí A, et al. Single nucleotide polymorphisms in the Wnt and BMP pathways and colorectal cancer risk in a Spanish cohort. PLoS One. 2010;5:e12673.
26. Memni H, Macherki Y, Klayech Z, Ben-Haj-Ayed A, Farhat K, Remadi Y, et al. E-cadherin genetic variants predict survival outcome in breast cancer patients. J Transl Med. 2016;14:320.
27. Creyghton M, Cheng A, Welstead G, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107:21931-6.
28. Ma L, Yuan L, An J, Borton MC, Zhang Q, Liu Z. Histone H3 lysine 23 acetylation is associated with oncogene TRIM24 expression and a poor prognosis in breast cancer. Tumour Biol. 2016; 37:14803-12.
29. Klein B, Jang S, Lachance C, Mi W, Lyu J, Sakuraba S, et al. Histone H3K23-specific acetylation by MORF is coupled to H3K14 acylation. Nat Commun. 2019;10:4724.
30. Li G, He Q, Yang L, Wang SB, Yu DD, He YQ, et al. Clinical significance of myeloid zinc finger 1 expression in the progression of gastric tumourigenesis. Cell Physiol Biochem. 2017;44: 1242-50.
31. Qu X, Yan X, Kong C, Zhu Y, Li H, Pan D, et al. c-Myb promotes growth and metastasis of colorectal cancer through c-fos-induced epithelial-mesenchymal transition. Cancer Sci. 2019; 110:3183-96.
32. Andersen L, Mørk N, Reinert L, Kofod-Olsen E, Narita R, Jørgensen SE, et al. Functional IRF3 deficiency in a patient with herpes simplex encephalitis. J Exp Med. 2015;212:1371-9.
33. Morales-Sanchez A, Fuentes-Panana E. Epstein-Barr virus-associated gastric cancer and potential mechanisms of oncogenesis. Curr Cancer Drug Targets. 2017;17:534-54.
34. Baghbanian M, Mousa SH, Doosti M, Moghimi M. Association between gastric pathology and Hepatitis B virus infection in patients with or without Helicobacter pylori. Asian Pac J Cancer Prev. 2019;20:2177-80.
35. Schierhout G, McGregor S, Gessain A, Einsiedel L, Martinello M, Kaldor J. Association between HTLV-1 infection and adverse health outcomes: a systematic review and meta-analysis of epidemiological studies. Lancet Infect Dis. 2020;20:133-43.
36. Jonsson BA, Adami HO, Hägglund M, Bergh A, Göransoon I, Stattin P, et al. -160C/A polymorphism in the E-cadherin gene promoter and risk of hereditary, familial and sporadic prostate cancer. Int J Cancer. 2004;109:348-52.
37. Suske G. The Sp-family of transcription factors. Gene. 1999; 238:291-300.
38. Ratajewski M, Walczak-Drzewiecka A, Gorzkiewicz M, Salkowska A, Dastych J. Expression of human gene coding RORγT receptor depends on the Sp2 transcription factor. J Leukoc Biol. 2016;100:1213-23.
39. Yago M, Ohki R, Hatakeyama S, Fujita T, Ishikawa F. Variant forms of upstream stimulatory factors (USFs) control the promoter activity of hTERT, the human gene encoding the catalytic subunit of telomerase. FEBS Lett. 2002;520:40-6.
40. Liu N, Guo X, Liu J, Cong YS. Role of telomerase in the tumour microenvironment. Clin Exp Pharmacol Physiol. 2019;47: 357-64.
41. Jutzi J, Bogeska R, Nikoloski G, Schmid CA, Seeger TS, Stegelmann F, et al. MPN patients harbor recurrent truncating mutations in transcription factor NF-E2. J Exp Med. 2013;210: 1003-19.
42. Ma X, Ma Y, Zhou H, Zhang JH, Sun MJ. Identification of the lncRNA miRNA mRNA network associated with gastric cancer via integrated bioinformatics analysis. Oncol Lett. 2019;18: 5769-84.
43. Lev Maor G, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. Trends Genet. 2015;31:274-80.
44. Mun DG, Bhin J, Kim S, Kim H, Jung JH, Jung Y, et al. Proteogenomic characterization of human early-onset gastric cancer. Cancer Cell. 2019;35:111-24.
45. Guindalini R, Cormedi M, Maistro S, Pasini FS, Branas PC, Dos Santos L, et al. Frequency of CDH1 germline variants and contribution of dietary habits in early age onset gastric cancer patients in Brazil. Gastric Cancer. 2019;22:920-31.