GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

EGFR gene polymorphisms -216G>T and -191C>A are risk markers for gastric cancer in Mexican population

J.H. Torres-Jasso[1,2], M.E. Marín[3], E. Santiago-Luna[4], J.C. Leoner[5], J. Torres[6], M.T. Magaña-Torres[1,2], F.J. Perea[1,2], B. Ibarra[1,2] and J.Y. Sánchez-López[1,2]

[1]División de Genética, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Jalisco, México
[2]Doctorado en Genética Humana, Universidad de Guadalajara, Guadalajara, Jalisco, México
[3]Servicio de Endoscopías, Unidad Médica de Atención Ambulatoria No. 52, IMSS, Guadalajara, Jalisco, México
[4]Servicio de Endoscopías, Hospital de Especialidades, Centro Médico Nacional de Occidente, IMSS, Guadalajara, Jalisco, México
[5]Servicio de Oncología, Hospital de Especialidades, Centro Médico Nacional de Occidente, IMSS, Guadalajara, Jalisco, México
[6]Unidad de Investigación Médica de Enfermedades Infecciosas y Parasitarias, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, IMSS, México

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

Genet. Mol. Res. 14 (1): 1802-1807 (2015)
Received February 26, 2014
Accepted July 2, 2014
Published March 13, 2015
DOI http://dx.doi.org/10.4238/2015.March.13.8

ABSTRACT. Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein with tyrosine-kinase activity that plays an important role in multiple cellular functions. EGFR overexpression has been observed in several types of tumors and it is significantly associated with disease stage, survival, prognosis, and progression of cancer. The polymorphisms -216G>T, -191C>A, and $(CA)_n$ first intervening sequence

(IVS1) have been related to EGFR overexpression and have been studied in several types of cancer, but not in gastric cancer (GC). The aim of this study was to determine the association of these 3 polymorphisms and GC. Genomic DNA from 68 GC patients and 102 healthy blood donors were analyzed. Polymorphisms were identified by DNA-sequencing (-216G>T and -191C>A) and GeneScan $(CA)_n$ IVS1. The results showed that the distribution of the -216G>T and -191C>A genotypes differed between groups ($P < 0.05$). The odds ratio for the -216TT genotype was 4.59 (95% confidence interval = 1.55-13.54, $P < 0.05$) and 10.71 (95% confidence interval = 2.31-49.59, $P < 0.05$) for the -191AA genotype, both in a recessive model. The genotype and allele distributions of the $(CA)_n$ IVS1 repeat was similar in both groups. In conclusion, the -216TT and -191AA genotypes and GA haplotype of the *EGFR* gene were found to be associated with an increased risk of gastric cancer in a Mexican population.

Key words: *EGFR*; *ERBB1*; Gastric cancer; Polymorphism

INTRODUCTION

Gastric cancer (GC) has the third highest mortality rate worldwide; in addition, it is the fifth most common cancer, with approximately 952,000 cases diagnosed each year. In Mexico, stomach tumors have an incidence of 6.8 cases per 100,000 (Ferlay et al., 2012). A number of genes are involved in the development of GC, such as those coding for oncogenes, tumor suppressors, cell cycle regulators, DNA repair factors, cell adhesion molecules, and cell growth factors and their receptors, among others (Galysh and Powell, 2009).

The epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein that has tyrosine kinase activity and plays an important role in multiple cellular functions, including cell adhesion, differentiation, migration, survival, and proliferation (www.omim.org/entry/131550). The receptor is encoded by the *EGFR* gene, which is located on chromosome 7p12.1-12.3, and is composed of 28 exons that span approximately 200 kb (www.ncbi.nlm.nih.gov/gene?term=1956).

EGFR overexpression has been observed in several types of tumors and it is significantly associated with disease stage, response to chemotherapy, survival, prognosis, development, and progression of cancer (Mitsudomi and Yatabe, 2010). Three *EGFR* gene polymorphisms, rs712829 (-216G>T), rs712830 (-191C>A), and $(CA)_n$ first intervening sequence (IVS1), have been associated with EGFR overexpression (Gebhard et al., 1999; Liu et al., 2005), as well as with glioblastoma (Carpentier et al., 2006), breast cancer (Brandt et al., 2004), and lung cancer (Zhang et al., 2007).

In this study, we analyzed the association between the -216G>T, 191C>A, and $(CA)_n$ IVS1 polymorphisms in the *EGFR* gene and gastric cancer.

MATERIAL AND METHODS

We studied 68 GC patients from different hospitals of the Instituto Mexicano del Seguro Social (IMSS) from Guadalajara and Mexico City. GC was diagnosed based on histo-

pathological examination according to Lauren classification. The average age of the patients was 60 years (range, 29-86 years), with a 2:1 ratio of males to females (46:22). In the control group, 102 Mexicans who were unrelated volunteers over 18 years had samples collected at the blood bank of the Centro Medico Nacional de Occidente, IMSS. Informed consent was obtained from all patients and controls in accordance with institutional guidelines. The study was conducted according to the Declaration of Helsinki.

Genomic DNA was extracted using the salting out method (Miller et al., 1988). The -216G>T and -191C>A polymorphisms were identified by polymerase chain reaction (PCR) and DNA sequencing. A 446-bp fragment was amplified with the following primers: forward 5'-CCT CCT CGC ATT CTC CTC CTC CT-3' and reverse 5'-TTT TCC TCC AGA GCC CGA CTC GC-3'. PCR was performed according to the AccuPrime GC-Rich Polymerase (Invitrogen, Carlsbad, CA, USA) protocol in a thermal cycler (Techne, Cambridge, UK) with the following program: initial denaturing at 95°C for 4 min; 30 cycles at 95°C for 30 s, 67°C for 30 s, and 72°C for 30 s; a final elongation at 72°C for 10 min. The sequencing reaction was performed using the Big Dye Terminator v3.1 kit. The $(CA)_n$ IVS1 repeat was analyzed using Genescan. The reaction was performed using previously described primers (Kharrat et al., 2007). After the initial denaturation at 94°C for 5 min, the samples were run for 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, and a final elongation at 72°C for 3 min. The resulting products ranged from 190-212 bp, which corresponded to 12-23 CA repeats. A total of 0.5 µL of each product was mixed with 9.35 µL formamide and 0.15 µL GeneScan-500 LIZ size standard. For sequencing and Genescan, we used ABI-PRISM 310 (Applied Biosystems, Foster City, CA, USA).

The allele and genotype frequencies were determined by direct counting. The chi-square test and Hardy-Weinberg equilibrium were estimated using the Epi-Info v3.1 statistical program. Odds ratio (OR) analysis was performed using the Cochran Armitage test under classical inheritance patterns. Haplotypes were established and compared between both studied groups. Linkage disequilibrium was evaluated using haplotype data and the allele correlation test (r^2). Linkage disequilibrium was statistically significant when $r^2 \geq 0.33$ (Shifman et al., 2003) and the Arlequin 3.11 software (Excoffier et al., 2005) was used for analysis. Values were statistically significant when $P < 0.05$.

RESULTS

EGFR -216G>T and -191C>A genotype and allele frequencies are shown in Table 1. The distribution of genotypes between GC patients and controls was significantly different. The -216GT genotype was more frequent in the control group than in GC patients (36.3 *vs* 7.4%), while the -216TT genotype was significantly higher in GC cases (19.1 *vs* 4.9%). The allele frequencies were not statistically significant. The risk analysis for genotype -216TT showed an OR of 4.59 (95%CI = 1.55-13.54, $P < 0.05$) under a recessive pattern of allelic interaction (Table 1).

For the -191C>A polymorphism, the AA genotype and allele A were significantly more frequent in GC patients than in controls (17.6 *vs* 2.0% and 25 *vs* 15.7%, respectively). We observed an OR of 10.71 (95%CI = 2.31-49.59, $P < 0.05$) for the -191AA genotype (Table 1).

Twenty different genotypes for the $(CA)_n$ IVS1 polymorphism were observed in GC cases and 22 in controls. The more frequent combinations in GC cases were $(CA)_{16}/(CA)_{20}$ and $(CA)_{18}/(CA)_{18}$, with frequencies of 19.1 and 8.8%, respectively; for controls $(CA)_{16}/(CA)_{20}$ and $(CA)_{18}/(CA)_{20}$, frequencies were 16.2 and 12.1%, respectively. In both groups, the genotype distribution was statistically similar ($P > 0.05$).

Table 1. Genotype distribution and allele frequencies of the -216G>T and -191C>A polymorphisms.

	Gastric cancer cases [N = 68 (%)]	Controls [N = 102 (%)]	Odds ratio (95%CI)
-216G>T			
Genotypes*			
GG	50 (73.5)	60 (58.8)	TT*
GT	5 (7.4)	37 (36.3)	4.59
TT	13 (19.1)	5 (4.9)	(1.55-13.54)
Alleles**			
G	105 (77.2)	157 (76.9)	
T	31 (22.8)	47 (23.1)	
-191C>A			
Genotypes*			
CC	46 (67.6)	72 (70.5)	AA*
CA	10 (14.7)	28 (27.5)	10.71
AA	12 (17.6)	2 (2.0)	(2.31-49.59)
Alleles*			
C	102 (75.0)	172 (84.3)	
A	34 (25.0)	32 (15.7)	
Haplotypes -216 and -191*			
	Chr (%)	Chr (%)	
GC	47 (53.4)	115 (61.2)	GA*
GA	22 (25.0)	28 (14.9)	10.2
TC	17 (19.3)	45 (23.9)	(2.07-50.46)
TA	2 (2.3)	0 (0)	

*P < 0.05; **P > 0.05; Chr = chromosomes. The polymorphisms were in Hardy-Weinberg equilibrium in controls (P > 0.05).

The allele frequencies of polymorphism between GC cases and controls were similar (P > 0.05; Table 2), and the most common alleles were $(CA)_{20}$, $(CA)_{18}$, and $(CA)_{16}$. We calculated ORs of alleles with repeats below $(CA)_{\leq 16}$ but the results were not significant (P > 0.05).

Table 2. Allele frequencies of the $(CA)_n$ IVS1 observed in gastric cancer patients and controls.

CA repeat (bp)	GC cases		Controls*	
	No. alleles = 136	%	No. alleles = 198	%
12 (190)	1	0.73	0	0
14 (194)	1	0.73	0	0
15 (196)	2	1.47	2	1.0
16 (198)	32	23.53	58	29.3
17 (200)	4	2.94	7	3.5
18 (202)	29	21.32	31	15.7
19 (204)	18	13.24	17	8.6
20 (206)	38	27.95	57	28.8
21 (208)	9	6.62	22	11.1
22 (210)	2	1.47	3	1.5
23 (212)	0	0	1	0.5
		P = NS		

*$(CA)_n$ polymorphism was in Hardy-Weinberg equilibrium.

Haplotypes were established with -216G>T and -191C>A polymorphisms, in 44 GC patients and 94 controls. In 92.5% of the entire sample, the combination was identified directly, while in the remaining 7.5%, with both sites in heterozygote state (2 GC patients and 8 controls), the phase was inferred using the Bayesian algorithm. The 4 expected haplotypes were observed in GC cases and 3 in controls, and the most common was GC, with values of 54 and

61.2%, respectively (Table 1). The distribution of the haplotypes was statistically significant different between groups ($P < 0.05$). The GA haplotype showed an OR of 10.2 (95%CI = 2.07-50.46, $P < 0.05$). The 2 polymorphisms were in linkage equilibrium in both groups ($r^2 < 0.33$).

DISCUSSION

In this study, we found a positive association between the -216G>T and -191C>A *EGFR* polymorphisms and GC. For the -216G>T polymorphism, the carriers of genotype -216TT show an approximately 4-fold higher risk of developing GC than carriers of -216GT or GG genotypes (OR = 4.59, 95%CI = 1.55-13.54). This observation is consistent with the results of Carpentier et al. (2006), who found that the -216T allele increased glioblastoma risk (OR = 1.67, 95%CI = 1.24-2.25).

For the -191C>A polymorphism, the AA genotype increased GC risk by more than 10-fold compared to the CA or CC genotypes (OR = 10.71, 95%CI = 2.31-49.59, $P < 0.05$). This association may not have been detected in previous studies because of the low frequency of polymorphism in Asians, Africans, and Caucasians (<10%) (http://www.ncbi.nlm.nih.gov/SNP/).

The -216G>T and -191C>A polymorphisms are located in a region that is critical for *EGFR* promoter activity. The first is in 1 of the 4 binding sites of the Sp1 protein, which is a transcription factor containing a zinc finger protein motif that allows for direct binding to DNA, enhancing its transcription; in fact, the -216T allele was reported to increase expression by up to 30% (Liu et al., 2005). The -191C>A polymorphism is situated 5 bp away from 1 of the multiple transcriptional initiation sites (Liu et al., 2005), suggesting a role in expression. Therefore, these polymorphisms may be associated with GC through overexpression of the *EGFR* gene.

However, the $(CA)_n$ IVS1 repeat was not associated with GC in our study. This lack of association was also observed in patients with lung cancer (Lee et al., 2007).

In the haplotype analysis, we found that the GA combination increases GC risk by approximately 10-fold. Based on our results, the TA haplotype was associated with an increased risk of GC development; however, because of its low frequency (2/88 in cases *vs* 0/188 controls), this association was not significant.

Although we examined a small number of samples in this study, our results are consistent with those other studies that studied the relationship between these polymorphisms and development of cancer. In summary, the -216 TT and -191 AA genotypes and the GA haplotype of the *EGFR* gene are associated with an increased risk of GC in a Mexican population.

ACKNOWLEDGMENTS

Research supported by the National Council of Science and Technology (CONACyT), Mexico (Grant #68669; SALUD-2007). We would like to thank Ing. Rogelio Troyo Sanroman for his assistance with the statistical analysis.

REFERENCES

Brandt B, Hermann S, Straif K, Tidow N, et al. (2004). Modification of breast cancer risk in young women by a polymorphic sequence in the EGFR gene. *Cancer Res*. 64: 7-12.

Carpentier C, Laigle-Donadey F, Marie Y, Auger N, et al. (2006). Polymorphism in Sp1 recognition site of the EGF receptor gene promoter and risk of glioblastoma. *Neurology* 67: 872-874.

Excoffier L, Laval G and Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47-50.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, et al. (2012). GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2013. Avaliable at [http://globocan.iarc.fr]. Accessed February 14, 2014.

Galysh R and Powell SM (2009). Cancer genetics of human gastric adenocarcinoma. In: The Biology of Gastric Cancers (Wang T, Fox J, Giraud A, eds.). Springer, Berlin, 251-284.

Gebhard F, Zanker KS and Brandt B (1999). Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. *J. Biol. Chem.* 274: 13176-13180.

Kharrat N, Al'Fadhli S, Rebai M, Rebai M, et al. (2007). (AC) dinucleotide repeat polymorphism in intron 1 of human EGFR shows ethnic specificities and high evidence for association with breast cancer. *Int. J. Biol. Markers* 22: 258-264.

Lee SJ, Kim KM, Chae MH, Lee WK, et al. (2007). No association between dinucleotide repeat polymorphism in intron 1 of the epidermal growth factor receptor gene EGFR and risk of lung cancer. *Cancer Genet. Cytogenet.* 172: 29-32.

Liu W, Innocenti F, Wu MH, Desai AA, et al. (2005). A functional common polymorphism in a Sp1 recognition site of the epidermal growth factor receptor gene promoter. *Cancer Res*. 65: 46-53.

Miller SA, Dykes DD and Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16: 1215-1216.

Mitsudomi T and Yatabe Y (2010). Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *FEBS J*. 277: 301-308.

Shifman S, Kuypers J, Kokoris M, Yakir B, et al. (2003). Linkage disequilibrium patterns of the human genome across populations. *Hum. Mol. Genet.* 12: 771-776.

Zhang W, Weissfeld JL, Romkes M, Land SR, et al. (2007). Association of the EGFR intron 1 CA repeat length with lung cancer risk. *Mol. Carcinog.* 46: 372-380.