Genetic variation analysis in a follow-up study of gastric cancer precursor lesions confirms the association of *MUC2* variants with the evolution of the lesions and identifies a significant association with *NFKB1* and *CD14.*

Osmel Companioni[1], Catalina Bonet[1], Nadia García[1,2], María José Ramírez Lázaro[3], Sergio Lario[3], Jorge Mendoza[4], Mª Magdalena Adrados[5], Elvira Poves[6], Laura Espinosa[6], José Juan Pozo Kreilinger[7], Luís Ortega[8], Luis Bujanda[9], Angel Cosme[10], Angel Ferrández[11], Guillermo Muñoz[12], Miriam Cuatrecasas[13], Ignasi Elizalde[14], Victoria Andreu[15], Mª José Paules[16], Beatriz Madrigal[17], Jesús Barrio[18], María Berdasco[19], Xavier Calvet[3], José Miguel Sanz Anquela[20]*, Javier P. Gisbert[4]*, Carlos A. González[1]* and Núria Sala [1,2§], on behalf of the Study Group[21].

[1]Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO)-IDIBELL, Barcelona, Spain. [2]Translational Research Laboratory, Catalan Institute of Oncology (ICO)-IDIBELL, Barcelona, Spain. [3]Departament of Medicine, Digestive Diseases Service, Institut Universitari Parc Taulí, Sabadell, and Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Spain. [4]Department of Gastroenterology, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria Princesa (IIS-IP) and CIBEREHD, Madrid, Spain. [5]Department of Pathology, Hospital Universitario de la Princesa, IIS-IP, Madrid. [6]Department of Gastroenterology, Hospital Universitario Príncipe de Asturias, Alcalá de Henares, Spain. [7]Department of Pathology, Hospital Universitario La Paz, Madrid. [8]Department of Pathology, Hospital Clínico San Carlos, Madrid, Spain. [9]Department of Pathology and [10]Department of Gastroenterology, Hospital Donostia/Instituto Biodonostia, Universidad del País Vasco (UPV/EHU), and CIBEREHD, San Sebastián, Spain. [11]Department of Gastroenterology and [12]Department of Pathology, Hospital Clínico Universitario Lozano Blesa, Zaragoza, and CIBEREHD, Spain. [13]Department of Pathology, Hospital Clínic de Barcelona, IDIBAPS and CIBEREHD, and Universitat de Barcelona, Spain. [14]Department of Gastroenterology, Hospital Clínic de Barcelona, IDIBAPS and CIBEREHD, Spain. [15]Department of Gastroenterology, Hospital de Viladecans, Spain. [16]Department of Pathology, Hospital Universitari de Bellvitge, L'Hospitalet de Llobregat, Spain. [17]Department of Pathology, Hospital Universitario Río Hortega, Valladolid, Spain. [18]Department of Gastroenterology, Hospital Universitario Río Hortega, Valladolid, Spain. [19]Cancer Epigenetics and Biology Program, IDIBELL Barcelona, Spain. [20]Department of Pathology, Hospital "Principe de Asturias" and University of Alcalá, Alcalá de Henares, Spain. [21]See annex for members of the Study Group.

*These three authors have equally contributed to the granted project design and coordination.

[§] Corresponding author: Núria Sala, Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO)-IDIBELL, Barcelona, Spain. Fax: +34 932607466, email: nsala@iconcologia.net. Tel: +34 932607464.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.31839

Short title: Genetic variation and gastric cancer precursor lesions

Keywords: *MUC2*, *NFKB1*, *CD14*, gastric cancer precursor lesions, genetic association.

Abbreviations used: GC, Gastric cancer; GCPLs, gastric cancer precursor lesions; NAG, non-atrophic gastritis; CAG, Chronic Atrophic Gastritis; IIM, Incomplete Intestinal Metaplasia; CIM, Complete Intestinal Metaplasia; DISP, Dysplasia.

Article Category: Cancer epidemiology/Genetics

Novelty and Impact

Few studies have addressed the association between genetic variability in candidate genes and the evolution (progression, stability or regression) of gastric cancer precursor lesions (GCPLs). By using multinomial logistic regression analysis in an independent, long-term follow-up study, we validate previous results of the association between *MUC2* variants and the evolution of GCPLs, and identify novel significant associations with variants in *NFKB1* and *CD14,* two genes of the *H. pylori* pathway first analyzed in relation to CGPLs.

Precursor lesions play a central role in the development of gastric cancer, yet few studies have examined associations between candidate gene variants and the evolution of gastric cancer precursor lesions (GCPLs). Here, using multinomial logistic regression analysis in a long-term trial of GCPL patients, the authors show that genetic variants in *MUC2* are inversely associated with GCPL regression. Variations in *NFKB1* and *CD14*, genes involved in the immune response against *Helicobacter pylori* infection, a cause of gastric cancer, were also associated with GCPLs. The findings highlight the potential influence of gene variants on GCPL evolution and gastric tumorigenesis.

Abstract

Gastric carcinogenesis proceeds through a series of gastric cancer precursor lesions (GCPLs) leading to gastric cancer (GC) development. Although *Helicobacter pylori* infection initiates this process, genetic factors also play a role. We previously reported that genetic variability in *MUC2* is associated with the evolution of GCPLs. In order to replicate previous results in an independent sample series and to explore whether genetic variability in other candidate genes plays a role in the evolution of GCPL, genomic DNA from 559 patients with GCPLs, recruited from 9 Spanish hospitals and followed for a mean of 12 years, was genotyped for 141 SNPs in 29 genes. After follow-up, 45.5% of the lesions remained stable, 37% regressed and 17.5% progressed to a more severe lesion. Genetic association with the evolution of the lesions (progression or regression) was analyzed by multinomial and binomial logistic regression. After correction for multiple comparisons, the results obtained confirmed the inverse association between *MUC2* variants and the regression of the lesions. A significant association was also observed between *NFKB1* and *CD14* variants and the evolution of the lesions; interestingly, this association was with both progression and regression in the same direction, which could reflect the dual role of inflammation in cancer. Stratified analyses according to *H. pylori* virulence factors indicated some significant and differential effects but none of them passed the FDR test. These results confirm that genetic variability in *MUC2*, *NFKB1* and *CD14* may have a role in the evolution of the GCPLs along time and in gastric carcinogenesis.

Introduction

Gastric cancer (GC) is the fifth cause of cancer related death worldwide, its relative survival at 5 years is only 23% in Europe [1]. GC is the outcome of a multistep sequence (Correa´s cascade) of GCPLs starting as a non-atrophic inflammation, which may progress to chronic atrophic gastritis (CAG), complete and incomplete intestinal metaplasia (IM), dysplasia and finally GC [2]. Infection by *H. pylori* is the main initiating factor for this process; however only 1 to 4% of infected people will develop GC [3]. This fact suggests that other factors, among which host genetic susceptibility, could influence the evolution (progression to a more advanced lesion or regression to a less severe one) of GCPLs over time.

To our knowledge, only four studies [4–7] have evaluated the genetic association between SNPs in candidate genes and the evolution of GCPLs along time. In one of them, performed in a Spanish population from the province of Soria, our group found a significant association between *MUC2* genetic variability and the evolution of GCPLs [4], which at that moment could not be replicated in an independent population. Apart from *MUC2*, in that previous study we also found a weak association between variants in *CD14*, *NFKB1*, *CDH1*, *TFF1*, *TFF2* and *IL1A* genes and the evolution of GCPLs (unpublished results). The other studies explored the association with genes involved in the DNA repair pathway [5], production of inflammatory cytokines [6] and inflammatory prostaglandins [7].

However, there are several gastric carcinogenesis genes whose association with the evolution of GCPLs has not yet been evaluated. Among these, there are genes coding for proteins that interact with *H. pylori*, genes involved in the immune response

against infection, and others with important functions in the gastrointestinal physiology. A significant association between genetic polymorphisms of these genes and the evolution of GCPLs would be an indication of the role of their proteins on GC development. Depending on the results obtained, it could also contribute to the characterization of biomarkers for the identification and surveillance of patients with precursor lesions at higher risk of progression to GC. In intermediate risk populations such as Spain, this approach might help to reducing advanced GC occurrence [8].

The objectives of the present study were therefore to validate the genetic associations with the evolution of GCPLs identified in our previous study [4],and to assess the genetic association with other novel candidate genes, in an independent, long-term follow-up multicenter study performed in 9 Spanish hospitals [9].

Materials and Methods

Patients and samples

Patients and biological samples were obtained from a previously reported observational longitudinal study of GCPLs [9]. Briefly, during 2012-2013, all patients with histological diagnoses of CAG, IM or dysplasia between 1995 and 2004 were identified from the Pathological Anatomy Department's files of 9 public hospitals of Spain. Inclusion criteria were 25-69 years of age, absence of peptic ulcer, Barrett's esophagus, GC, other cancer or gastric resection, and available clinical and demographic information. Patients following the inclusion criteria were invited to a new gastroscopy with gastric biopsy and saliva or blood collection for DNA extraction, and answered a clinical and lifestyle information questionnaire. All patients signed an informed consent to participate in the study, whose protocol was approved by the Ethical Committee of each hospital. The final sample size was 559 patients, followed for a mean of 12 years, with information on their GCPLs and available DNA.

Histopathology

As previously reported [9], for the recruitment biopsy each hospital followed their standard procedures at the time of the first gastric biopsy, although only patients with biopsies that fulfilled a minimum quality criteria (a paraffin block including at least 3 fragments, each one of 1.5mm minimum size and in which both antrum and oxintic gastric mucosa were recognized) were selected and collected for histological evaluation. For the end of follow-up gastroscopy, five specimens were obtained according to the Sydney recommendations [10] (one from the incisura angularis, two from the antrum, and two from the corpus of the stomach). The gastric endoscopy was performed by the same expert endoscopist of each participating hospital, following

standard clinical procedures of each department. When this study was performed, there were no established recommendations on the best endoscopic procedures [11]; therefore, no image enhancement techniques were used to properly identify lesions.

All biopsies (baseline and final), were formalin fixed and paraffin-embedded. Available sections of the baseline biopsy and those obtained from the final biopsy were used for histological diagnosis, following the usual procedures in each hospital. Diagnoses were made with one slide stained with hematoxylin and eosin (HE), although some centers also used Alcian blue- periodic acid Shiff (pH 2.5). The same pathologist from each hospital reviewed the initial and final biopsies following a common standard protocol for the diagnoses and classification of cases.

Patients' lesions were classified according to the most advanced lesion in the Correa's cascade [2] diagnosed in any of the biopsy fragments collected at recruitment as well as at the end of the follow-up [9]. The lesions were scored as follows: 1 = normal, 2 = non-atrophic gastritis (NAG), 3 = non-metaplastic multifocal atrophic gastritis (CAG), 4 = complete intestinal metaplasia including predominant complete (CIM), 5 = incomplete IM including predominant incomplete (IIM), 6 = dysplasia (DYS) and 7 = gastric cancer (GC). We have used subtyping of IM to classify the risk, a method that is considered valid, simple, cheap and useful for clinical practice [12]. The histological diagnosis at recruitment was: CAG, CIM, IIM and DYS. The classification at the end of follow-up was: normal mucosa, NAG, CAG, CIM, IIM, DYS and GC. As in our previous study [4], based on the diagnosis at recruitment and at the end of follow-up, it was considered that the lesions had progressed or regressed if they had respectively advanced or

regressed at least one point in the overall score (1 to 7). The lesions were considered to be stable if they maintained the same score.

***H. pylori* infection and genotyping**

H. pylori status at recruitment was determined by *cagA* and *vacA* genotyping. DNA extraction from the FFPE block at recruitment was performed by phenol-chloroform method. After checking for the integrity of FFPE DNA by means of ACTB gene amplification, the *ure*A gene and the *vac*As, *vac*Am and *cag*A virulence factor genes *of H.pylori* were genotyped by PCR amplification in a 7500Fast Real-Time PCR System (Applied Biosystems), using custom locked nucleic acids primers (LNA, Exiqon, Denmark) and SensiMixdU SYBR® Green Kit (Quantace Ltd, UK). The lengths of the amplified products were estimated in an Agilent 2100 Bioanalyzer using a DNA1000 lab chip (Agilent Technologies Inc., Palo Alto, CA, USA). Details on cagA and vacA genotyping are described in the online Supporting Methods.

Infection by *H. pylori* was considered positive if at least two PCR were positives among *ure*A, *vac*As, *vac*Am and *cag*A. When only one or none of the PCR was positive, infection by *H. pylori* was considered negative.

H. pylori status was also determined by Giemsa staining and from the medical records (rapid urease and urea breath tests) and interviews about previous *H.pylori* infection. All patients having at least one positive diagnostic in any of these sources of information were considered positive.

Sample collection and DNA extraction for SNP genotyping

At collection, between 1 to 5 ml of saliva were mixed with an equal volume (1:1) of a stabilizer buffer from patent Appl. No 12/338 848 and kept at 4°C until processing.

DNA extraction for SNP genotyping was performed by phenol-chloroform method for saliva and FFPE samples, and by the salting-out method for blood samples. Only in a minority of cases (5%) DNA was obtained from FFPE gastric mucosa or from whole blood. DNA was dissolved in Tris-EDTA 1X and diluted to 50 ng/µl in dH_2O for genotyping.

Selection of candidate genes and SNPs

We selected human candidate genes influencing gastric carcinogenesis from studies using mouse models of GC [13], as well as from the *H.pylori* infection pathway [14]. After this gene selection, the SNP Genotype Data of European population (CEU) for each gene plus 5 and 3 Kb upstream and downstream respectively, was downloaded from the HapMap Genome Browser release #28 (Phases 1, 2 & 3 - merged genotypes & frequencies, dbSNP b126). Genotype data was analyzed with Haploview v.4.2 [15], setting MAF≥0.05 and using the LD blocks definition of confidence intervals [16]. tagSNPs were selected using the Tagger Aggressive option (2 and 3 marker haplotypes) with r^2=0.8. Additionally, we also included SNPs associated with GC from GWAS studies[17–20] as well as SNPs associated with the evolution of GCPLs from a previous study of our group [4].

Altogether, we selected 148 SNPs from 29 human genes: 10 human orthologues from mouse models of gastric cancer, 11 genes of the *H. pylori* pathways triggered by the infection, 6 genes previously associated with gastric cancer in GWAS, and 7 genes with SNPs previously associated with the evolution of GCPLs, 5 of which also belonged to other groups. Supporting Table 1 shows the analyzed genes and genotyping data of the SNPs available for analysis after quality control filtering.

Genotyping

Genotyping was performed with the MassArray iPlex-Gold technology (Sequenom) at the National Genotyping Center of Spain (CEGEN) in a customized assay design. To assess the genotyping error rate, a trio of Coriell samples and 5% of test samples duplicated at random were included for quality control. No genotyping assay could be designed for three tagSNPs of *MAP3K14* and one of *IL1B*.

Statistical Analysis

Patients were categorized into groups of progression, regression and stables depending on whether the severity of their histological GCPLs increased, decreased or remained stable between recruitment and the end of follow-up. Chi^2 test for contingency tables was performed to evaluate differences between the above mentioned groups in population characteristics (sex, age, NSAID consumption, family history of GC, smoking, *H. pylori* infection and histological diagnosis).

A chi^2 goodness of fit test was used to check if genotypes were in Hardy-Weinberg equilibrium ($p>0.001$) in all the samples. Association between the SNP genotypes and the evolution of the lesions was analyzed by multinomial logistic regression comparing genotypes in those lesions that progressed or regressed against those in the lesions that remained stable, as the reference group. This analysis was performed using the nnet package implemented in the R software[21]. Association was also analyzed by binomial logistic regression, comparing progression with stables and regression with stables. For this analysis the SNPassoc library of the R software was used. In all analyses ORs were calculated under codominant, dominant, recessive and log-additive (per allele) inheritance models after adjustment by sex and age. The Likelihood-Ratio

Test (LRT) was used to assess the significance of the association. Results were considered significant if p-value<0.05. The False Discovery Rate (FDR) test[22] was applied to correct for multiple comparisons in both analyses. An association was considered significant if the p-value of the FDR test was p≤0.05.

A sub-analysis of association according to the presence of *H. pylori* virulence factors was also carried out by multinomial logistic regression comparing: *i) cag*A+ (N=241) vs *cag*A- (N=119) and *ii)* High Virulence (HV) *vac*As1m1-*cag*A+ (haplotype s1m1-1, N=164) vs Low-Intermediate Virulence (LIV, N=149), that includes all *cag*A-*vac*A haplotypes different from s1m1-1, among which there are 39 low virulence *vac*As2m2-*cag*A- strains. For some samples where it was not possible to obtain a reliable haplotype, this partial haplotypes were classified as 'missing' and excluded from analysis. This was performed because people infected with *cag*A+ and *vac*A s1/m1 strains are at increased risk of developing GCPLs and GC [23] as well as because virulence factors have been found associated with the progression of GCPLs [24]. The Wald test was used to check for heterogeneity.

All analyses were done using the same reference allele for each SNP, which was the commonest one in the general population (Supporting Table 1).

Haplotype frequencies were inferred using the EM (expectation-maximization) algorithm implemented in haplo.stats package of R software [21]. Associations between progression or regression and each haplotype with a frequency ≥0.01 were estimated by binomial logistic regression using the most frequent haplotype among the stable lesion's group as reference.

Results

The available sample for statistical analyses was 141 SNP successfully genotyped in 559 patients' samples. Based on the evolution of the lesions the patients were grouped as stables (N=259, 46.3%), regression (N=204, 36.5%) or progression (N=96, 17.2%). Their main characteristics (sex, age, and family history of GC, smoking, NSAID consumption, *H.pylori* infection and histological diagnosis of the analyzed population at recruitment) are shown in Table 1. When the distribution of these variables in the evolution groups was compared, significant differences ($p<0.0001$, Table 1) were only observed for histological diagnosis at recruitment.

All genotyping results are indicated in Supporting Table 1. Seven SNPs initially selected for genotyping [rs11079008 (*GAST*), rs6090575 (*SRC*), rs4072037 (*MUC1*), rs2236851 and rs6688452 (*RUNX3*), rs28362491 (*NFKB1*), rs6890699 (*CDX1*)] were excluded from statistical analysis because they generated continuous or overlapping clusters. Genotype agreement in the Coriell samples was 100% and in the duplicated samples was of 99%. All SNPs were in Hardy-Weinberg Equilibrium ($p>0.001$).

Table 2 reports all significant associations (p-value<0.05) between the analyzed SNPs and the evolution (progression or regression) of the GCPLs when compared to the stables group, in the multinomial logistic regression analysis. When a SNP was found significant under different inheritance models, the reported result is the one obtained under the most significant model.

It can be observed that SNPs in *CD14*, *CDH1*, *MAP3K14*, *MAPK3*, *MUC2*, *NFKB1*, *NFKBIA*, *NOD1*, *PTGS2*, *SRC* and *TFF2* were associated with the evolution of GCPLs in this multinomial analysis, although the association with *CDH1*, *MAP3K14, NOD1* and

TFF2 variants was weak (Table 2). When the ORs for progression or regression were analyzed it was observed that variants in *CD14*, *NFKB1* and *NFKBIA* were associated with both progression and regression in the same direction. However, other genes were associated only with progression (*MAPK3* and *PTGS2)* or with regression (*MUC2* and *SRC*). *TFF2* rs1079380 showed a slightly significant association with the evolution of the lesions, with a positive effect on regression and an inverse effect on progression that did not reach statistical significance.

After applying the *FDR* test to correct for multiple comparisons, the unique SNPs that remained significant were rs11167532, rs2569190 and rs1583005 in *CD14*, rs10902073 in *MUC2*and rs1598861 and rs7674640 in *NFKB1*. Surprisingly, all these SNPs with the exception of *MUC2* rs10902073 were associated in the same direction with both progression and regression of the lesions, although the effect was higher on progression. In the case of *MUC2*, as previously described [4], both SNPs (rs10902073 and rs10794281) were inversely associated with regression, although only rs10902073 passed the FDR test (Table 2).

An additional binomial association analysis was also performed to compare progression vs stables, and regression vs stables (Supporting Table 2). *CD14* SNPs rs11167352, rs2569190 and rs1583005, and *NFKB1* rs1598861 and rs7674640 were also found to be associated with both progression and regression of the lesions in the same direction, although after adjustment for multiple comparisons, the *CD14* SNPs remained significant only for the association with progression while *NFKB1* rs1598861 maintained significance only with regression; rs7674640 was significantly associated with both progression and regression after FDR test. *MUC2* rs10902073 and

rs10794281 were also inversely associated only with regression, although these associations did not reach significance after FDR (p=0.06 for *MUC2* rs10902073, Supporting Table 2). This binomial analysis revealed some additional associations with variants in the *IL1B*, *IL1RN*, *PTGES* and *TFF1* genes, but in all cases the associations were weak and none of them passed the FDR test.

In general, the haplotype analysis (Table 3) confirmed the associations observed at single SNP level. Thus, the *MUC2* haplotype containing the A and C alleles of rs10902073 and rs10794281, respectively, showed a significant inverse association with the regression of the GCPLs. The *CD14* haplotypes GTGT and GCGC, both containing the more common G 'risk' alleles of rs11167532 and rs2569190, were positively associated with progression as well as with regression. Similarly, the *NFKB1* haplotypes GCAGAG and GCAGGA, both containing the G and A 'risk' alleles of rs980455 and rs1598861, respectively, were also positively associated with progression as well as with regression (Table 3). Haplotype analysis from the other genes was also in good agreement with the single SNP analysis (Table 2 and Supporting table 2), with the exception of the *PTPN11* CAGGAGC haplotype which was inversely associated with regression when none of the SNPs were.

In order to analyze the effect of the genetic variants according to the virulence of *H.pylori* strains, we repeated the analyses stratifying by *cag*A status (positive and negative) as well as by *cag*A and *vac*A haplotype, grouping into high virulence strains (HV, *cag*A+ and *vac*As1m1) and low-intermediate ones (LIV, *vac*A no s1m1, independently of *cag*A status). The results obtained (Supporting Tables 3 and 4) indicated some significant associations not previously observed in the general analysis

and a good general agreement between results stratified according to *cag*A presence/absence and those stratified according to *cag*A and *vac*A haplotypes (high or low virulence). Nevertheless, the heterogeneity analysis revealed that only a few of these SNPs had a significant differential effect according to *H.pylori* virulence subgroups and none of the associations observed was significant after the FDR test was applied.

Discussion

The identification and replication of human genetic variation statistically associated with the evolution of GCPLs may be relevant to reveal genes influencing gastric tumorigenesis. Only four association studies [4-7] have evaluated the genetic variability with respect to GCPLs evolution. In one of these studies performed by our group, *MUC2* variants rs10902073 and rs10794281 were found inversely associated with regression of GCPLs [4], rs2569190 of *CD14* and rs1079380 in *TFF2* were inversely associated with progression and *NFKB1* rs28362491, a SNP not analyzed in this study, was associated with both progression and regression of the lesions (unpublished). In the present study we have evaluated the replicability of these and other associations, and have analyzed for the first time other genes not previously considered and having key roles in gastrointestinal physiology and in the interaction with *H. pylori*, influencing inflammation and, potentially, the evolution of GCPLs. Of all analyzed genes, two of the *H.pylori* pathway (*NFKB1* and *CD14*) and one of the epithelial mucous barrier (*MUC2*) exhibited the most relevant associations. Although several other significant associations have been observed, most of them in genes of the *H.pylori* infection pathway, only variants in *MUC2*, *CD14* and *NFKB1* remained significant after adjustment for multiple comparisons. Furthermore, when the haplotype analyses were performed, the results obtained agreed with those of the single SNP analyses and confirmed the association of common haplotypes in these genes with the evolution of the lesions. When the analyses were stratified according to *H.pylori* virulence factors *cag*A and *vac*As/m, none of the *MUC2*, *CD14* and *NFKB1* variants that were significant after the FRD test in the general analyses had a differential effect on the evolution of the lesions, according to *H.pylori* virulence. This could be explained by previous

findings from our group indicating that *CD14* variation is associated with GC of the cardia, a cancer location not associated with *H.pylori* infection [14]. Although in the same study *NFKB1* variants showed a nominal significant association with distal GC, which is the cancer location associated with *H.pylori* infection [14], *cag*A and *vac*A have been shown to be dispensable for direct NF-kB activation [25]. On the other hand, some SNPs in genes (particularly *NOD1* and *TFF1*) not significant in the general analyses had a significant differential effect in the evolution of the lesions according to either *cag*A or *cag*A and *vac*A status, although none of the associations passed the FDR test, probably because the stratification of patients reduces sample size and statistical power. Additionally, the virulence could not be determined in 166 patients (29%), probably because *H.pylori* had already disappeared at recruitment. It is known that *H.pylori* does not colonize gastric areas with extensive atrophy or intestinal metaplasia [2], which were inclusion criteria for the study.

Mucins are the major components of gastric mucus which protects gastric mucosa against bacterial infection, proteolytic enzymes and hydrochloric acid. They also limit the activation of inflammatory responses at the interface with the environment. Their over-expression, aberrant intracellular localization and changes in glycosylation pattern have been associated with various cancers[26]. MUC2 is a highly O-glycosylated secreted mucin expressed in both histological subtypes of intestinal metaplasia, but not in healthy stomach [2]. MUC2 expression is increased in carcinomas of the gastrointestinal tract; however, $Muc2^{-/-}$ mice develop intestinal adenocarcinomas [27] and it has been suggested that its overexpression in gastric carcinomas might reflect the origin of these tumors from intestinal metaplasia, rather than a role for this mucin in the tumoral process. Nevertheless, despite its tumor suppressor role in the

intestinal tract, attributed to protection of the intestinal epithelium against infection and inflammation, an oncogenic role of MUC2 overexpression has also been suggested as a result of a decreased activation of innate and adaptive immune responses [26]. Therefore, it is biologically plausible that the genetic association of *MUC2* variation with the evolution of GCPLs reflects changes in MUC2 expression in intestinal metaplasia or in its immunomodulatory role, although functional studies would be necessary to be conclusive. Although there is no information in the GTEx portal (https://gtexportal.org/home/) on *MUC2* eQTLs influencing *MUC2* expression in stomach, SNPs rs10902073and rs10794281 tag a strong LD block in the 5'-upstream region of the gene, suggesting that variation in this region may influence *MUC2* expression. It is also possible that the associated SNPs are in linkage disequilibrium with a functional variant in this gene. Other *MUC2* variants significantly associated with the evolution of GCPLs in our previous study [4] have not reached significance in the present study but the association results obtained are in the same direction as in the previous study. Conflicting findings of this type could be attributed to differences in sample size, population structure and allele frequencies between the analyzed populations [28].

CD14 is expressed by monocytes/macrophages and functions as a receptor for bacterial lipopolysaccharide and a mediator of the inflammatory response to *H.pylori* [29]. Although an inverse association with progression of CD14 rs2569190 had also been observed in our previous follow-up study of GCPLs [4], in that study the most common and reference allele was A instead of G. On the other hand, it is noteworthy that the same three SNPs associated with GCPLs evolution in this study had been previously associated, in the same direction, with risk of cardia GC in the EPIC cohort [14]. Finally,

the rs2569190 variant has been shown to be associated with CD14 protein expression in monocytes [30], altogether being in agreement with its genetic association with the evolution of the GCPLs.

NFKB1 (nuclear factor kappa-β, subunit 1) is the other relevant gene associated with evolution of GCPLs. It encodes for a transcriptional factor that induces the expression of pro-inflammatory cytokines, chemokines and cell adhesion molecules to elicit an immune response against *H.pylori* infection. The continued expression of these cytokines due to the bacterial infection causes a chronic inflammation that leads to GCPLs that could progress into GC [31]. Therefore, it is also biologically plausible that genetic variants in *NFKB1*, by modifying its expression or function, influence the progression of GCPLs.

It is remarkable to observe that variation in *CD14* and in *NFKB1*, as well as in other genes of the inflammatory response to *H.pylori* infection that did not reach significance after FDR (*NFKB1A, MAPK3, NOD1*), is associated in the same direction with both progression and regression of the GCPLs, suggesting a dual role of these genes in the evolution of the lesions. A similar result was also obtained in our previous study with *NFKB1* rs28362491, which was significantly associated with both the progression [OR(95%CI)=1.55 (1.06-2.28)p=0.024] and the regression [OR(95%CI)=1.53 (1.04-2.24)p=0.028] of the lesions (unpublished results). These apparently contradictory results, which exclude these genetic variants as biomarkers of the progression of the GCPLs, could be explained by the fact that these are SNPs in genes of mediators of the inmunoinflammatory process [32,33] and that inflammation has an increasingly recognized dual role as antitumorogenic and protumorogenic in

carcinogenesis [34-36]. Furthermore, this dual role will very much depend on the tumor microenvironment and the stage of the inflammatory reaction and tumorigenic process [34,36], which is unknown for our samples. For example, the TGFβ signaling pathway exerts tumor suppressor effects in normal cells and early tumors; however as these early tumors develop, the protective effects of TGFβ are lost and its activation changes to promote cancer progression [35]. The expression of pro-inflammatory cytokines protects gastric tissue by recruitment of effector cells during *H. pylori* acute inflammation. However, when this process becomes chronic, the same pro-inflammatory cytokines produce excessive cell proliferation and free radical production, conditioning tumor transformation, reminiscent of the effects of acute and chronic inflammation [36,37]. Such a dual role on both progression and regression of the lesions has not previously been reported by other genetic association studies. Nevertheless, it is to note that these previous studies were on other pathway genes such as DNA repair genes [5], or they only reported the association with progression [6,7].

The main strengths of our study are that we followed-up the evolution of GCPLs of a relatively large cohort of patients from Spain for a mean of 12 years. Besides, it has a larger sample size and we have analyzed a larger number of candidate genes than previous studies [4-7]. In contrast to the majority of studies that analyze either the host or the bacterial genetic factors, the present study explores both of them in the same patient allowing for stratified analyses according to *H.pylori* virulence factors. In addition, the *H.pylori* status and the virulence factors were determined in the recruitment biopsy sample. This is relevant because some biopsy specimens harbor multiple strains with different virulence that may be selectively lost during microbiological culture. Our main limitations are the still small sample size for a

genetic association study with two different outcomes, the failure in genotyping some tagSNPs, which reduces the coverage of their gene variability, as well as a possible bias in the classification of the gastric lesions or insufficient gastric mucosa sampling. Regarding sample size, this is something difficult to increase given the study and patients' recruitment characteristics [9]; nevertheless, although a greater sample size could have increased the power of our study to identify significant associations, there are reasons to believe that the significance of the reported associations is real: 1) they remain significant after adjustment for multiple comparisons, 2) they validate some previous findings and 3) they have biological plausibility. Regarding possible bias, in fact there is debate on whether the progression or regression of GCPLs observed from gastric biopsies taken at different time points are real or may be biased by mucosa sampling [2.] Recent clinical management guidelines for *H.pylori* infection agreed in that intestinal metaplasia cannot be reversed by bacterial eradication, although its progression is arrested in a subset of patients [38,39]. Nevertheless, recent studies indicate that IM can regress after *H.pylori* eradication treatment [40,41] and another study on the molecular determinants of progression of gastric cancer identifies patterns associated with IM regression [42].

Recent findings indicate that *H. pylori eradication treatment has an effect on the progression of the lesions* [40,41]. Analysis of a subset of patients (n=287) of our study which were positive for H. pylori infection and had information on eradication treatment, did not reveal significant differences on regression or progression of the lesions (Chi^2 test) according to treatment. It is to mention that since our study was not designed to evaluate the effect of *H. pylori* treatment on the evolution of the lesions,

infection at follow-up was not evaluated from *H. pylori* genotyping but from medical records and patients' questionnaires.

In conclusion, our study confirms previous results according to which genetic variability in *MUC2* is associated with the evolution of GCPLs and identifies novel significant associations with variants in *NFKB1* and *CD14,* two genes of the *H.pylori* pathway whose association with the evolution of GCPLs has not been previously reported. Although these associations have a biological plausibility, they need validation in larger and independent populations. Moreover, since a tagSNP strategy was used to optimize the coverage of gene variability, linkage disequilibrium analysis, as well as expression and functional studies, will be required to identify the causal variants and to explain their mechanism of action [43].

Funding and Acknowledgements

Funding

This work was supported by: Instituto de Salud Carlos III (Ministry of Economy, Industry and Competitiveness-MINECO, Spain), cofounded by European Regional Development Funds (ERDF/FEDER) '*A way to build Europe*', grant numbers: Exp. PI10/01089 to CAG, PI10/01203 to XC, PI10/02267 to MB, and PI10-01031/02654/01881/02612/0232/02934/01179 to the different participating hospitals PIs, as well as PI12/01187 and PI16/00599 to NS; RETICC RD12/0036/0018 to CAG. Agency for Management of University and Research Grants *(AGAUR)*, Secretariat of Universities and Research, Ministry of Enterprise and Knowledge of the Government of Catalonia, AGAUR 2014SGR726. Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), for PhD scholarship to OC, and CERCA Program/Generalitat de Catalunya for institutional support.

Acknowledgements

We thank the participating patients and their families for their collaboration; Leila Luján Barroso for her contribution as data manager; Dr Maria Torres, from the Spanish National Genotyping Center (CEGEN), Santiago node; Dr.Ville Pimenoff for advice with the tagSNP selection, and Angela Expósito for her administrative support.

Annex. Members of the Study Group.

Study Group: Department of Gastroenterology, Hospital Universitario de La Princesa, IIS-IP and CIBEREHD, Madrid: Adrian G. McNicholl, Aicia C Marín, Fernando Casals. Department of Pathology, Hospital Universitario de La Princesa, IIS-IP and CIBEREHD,

Madrid: Consuelo López. Department of Gastroenterology, Hospital Universitario La Paz, Madrid: Mª Dolores Martín Arranz. Department of Pathology, Hospital Donostia/Instituto Biodonostia,UPV/EHU, and CIBEREHD, San Sebastián: Maddi Garmendia. Department of Pathology, Hospital Universitario Rio Hortega, Valladolid: Mª Angeles Torres. Department of Gastroenterology, Hospital de Viladecans, Barcelona, Spain: Mercè Barenys and Beatriz García Zafra. Pathology Department, Complejo Hospitalario de Soria, Soria, Spain: Mª Luisa Pardo. Pathology department, Hospital Bidasoa, Hondarribia, Guipúzcoa, Spain: José Mª Arrinda.

Conflict of Interest Statement: None declared

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
2. Correa P, Piazuelo MB. The gastric precancerous cascade. *J Dig Dis* 2012;13:2–9.
3. Ghoshal UC, Chaturvedi R, Correa P. The enigma of Helicobacter pylori infection and gastric cancer. *Indian J Gastroenterol* 2011;29:95–100.
4. Marín F, Bonet C, Muñoz X, García N, Pardo ML, Ruiz-Liso JM, Alonso P, Capellà G, Sanz-Anquela JM, González CA, Sala N. Genetic variation in MUC1, MUC2 and MUC6 genes and evolution of gastric cancer precursor lesions in a long-term follow-up in a high-risk area in Spain. *Carcinogenesis* 2012;33:1072–80.
5. Li W-Q, Zhang L, Ma J-L, Zhang Y, Li J-Y, Pan K-F, You W-C. Association between genetic polymorphisms of DNA base excision repair genes and evolution of precancerous gastric lesions in a Chinese population. *Carcinogenesis* 2009;30:500–5.
6. Li Z-W, Wu Y, Sun Y, Liu L-Y, Tian M-M, Feng G-S, You W-C, Li J-Y. Inflammatory cytokine gene polymorphisms increase the risk of atrophic gastritis and intestinal metaplasia. *World J Gastroenterol* 2010;16:1788–94.
7. Pereira C, Sousa H, Ferreira P, Fragoso M, Moreira-Dias L, Lopes C, Medeiros R, Dinis-Riberio M. -765G > C COX-2 polymorphism may be a susceptibility marker for gastric adenocarcinoma in patients with atrophy or intestinal metaplasia. *World J Gastroenterol* 2006;12:5473–8.
8. de Vries AC, van Grieken NCT, Looman CWN, Casparie MK, de Vries E, Meijer G A, Kuipers EJ. Gastric Cancer Risk in Patients With Premalignant Gastric Lesions: A Nationwide Cohort Study in the Netherlands. *Gastroenterology* 2008;134:945–52.
9. González CA, Sanz Anquela JM, Companioni O, Bonet C, Berdasco M, López C, Mendoza J, Martín-Arranz MD, Rey E, Poves E, Espinosa L, Barrio J, Torres MÁ, Cuatrecasas M, Elizalde I, Bujanda L, Garmendia M, Ferrández Á, Muñoz G, Andreu V, Paules MJ, Lario S, Ramírez MJ, Study group, Gisbert JP.. Incomplete type of intestinal metaplasia has the highest risk to progress to gastric cancer: results of the Spanish follow-up multicenter study. *J Gastroenterol Hepatol* 2016;31:953–8.
10. Dixon MF, Genta RM, Yardley JH, Correa P. Classification and Grading of Gastritis. *Am J Surg Pathol* 1996;20:1161–81.
11. East JE, Vleugels JL, Roelandt P, Bhandari P, Bisschops R, Dekker E, Hassan C, Horgan G, Kiesslich R, Longcroft-Wheaton G, Wilson A, Dumonceau JM. Advanced endoscopic imaging: European Society of Gastrointestinal Endoscopy (ESGE) Technology Review. *Endoscopy* 2016; 48(11):1029-1045.
12. González CA, Sanz-Anquela JM, Gisbert JP, Correa P. Utility of subtyping intestinal metaplasia as marker of gastric cancer risk. A review of the evidence. *Int J Cancer* 2013; 133:1023-32
13. Hayakawa Y, Fox JG, Gonda T, Worthley DL, Muthupalani S, Wang TC. Mouse models of gastric cancer. *Cancers (Basel)* 2013;5:92–130.
14. Companioni O, Bonet C, Muñoz X, Weiderpass E, Panico S, Tumino R, Palli D, Agnoli C, Vineis P, Boutron-Ruault MC, Racine A, Clavel-Chapelon F,Travis RC, Khaw KT, Riboli E, Murphy N, Vergnaud AC, Trichopoulou A, Benetou V,

Trichopoulos D, Lund E, Johansen D, Lindkvist B, Johansson M, Sund M, Ardanaz E, Sánchez-Cantalejo E, Huerta JM, Dorronsoro M, Ramón Quirós J, Tjonneland A, Mortensen LM, Overvad K, Chang-Claude J, Rizzato C, Boeing H, Bueno-de-Mesquita HB, Siersema P, Peeters PH, Numans ME, Carneiro F, Licaj I, Freisling H, Sala N, González CA. . Polymorphisms of Helicobacter pylori signaling pathway genes and gastric cancer risk in the European prospective investigation into cancer-eurgast cohort. *Int J Cancer* 2014;134:92–101.

15. Smith AV. Manipulating HapMap Data Using HaploView. *Cold Spring Harb Protoc* 2008;2008:pdb.prot5025-prot5025.
16. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C,Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–9.
17. Jin G, Ma H, Wu C, Dai J, Zhang R, Shi Y, Lu J, Miao X, Wang M, Zhou Y, Chen J, Li H, Pan S, Chu M, Lu F, Yu D, Jiang Y, Dong J, Hu L, Chen Y, Xu L, Shu Y, Pan S, Tan W, Zhou B, Lu D, Wu T, Zhang Z, Chen F, Wang X, Hu Z, Lin D, Shen H. Genetic variants at 6p21.1 and 7p15.3 are associated with risk of multiple cancers in han chinese. *Am J Hum Genet* 2012;91:928–34.
18. Saeki N, Saito A, Choi IJ, Matsuo K, Ohnami S, Totsuka H, Chiku S, Kuchiba A, Lee Y, Yoon K, Kook M, Park SR, Kim YW, Tanaka H, Tajima K, Hirose H, Tanioka F, Matsuno Y, Sugimura H, Kato S, Nakamura T, Nishina T, Yasui W, Aoyagi K, Sasaki H, Yanagihara K, Katai H, Shimoda T, Yoshida T, Nakamura Y, Hirohashi S, Sakamoto H. A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology* 2011;140:892–902.
19. Sakamoto H, Yoshimura K, Saeki N, Katai H, Shimoda T, Matsuno Y, Saito D, Sugimura H, Tanioka F, Kato S, Matsukura N, Matsuda N, Nakamura T, Hyodo I, Nishina T, Yasui W, Hirose H, Hayashi M, Toshiro E, Ohnami S, Sekine A, Sato Y, Totsuka H, Ando M, Takemura R, Takahashi Y, Ohdaira M, Aoki K, Honmyo I, Chiku S, Aoyagi K, Sasaki H, Ohnami S, Yanagihara K, Yoon KA, Kook MC, Lee YS, Park SR, Kim CG, Choi IJ, Yoshida T, Nakamura Y, Hirohashi S . Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* 2008;40:730–40.
20. Shi Y, Hu Z, Wu C, Dai J, Li H, Dong J, Wang M, Miao X, Zhou Y, Lu F, Zhang H, Hu L, Jiang Y, Li Z, Chu M, Ma H, Chen J, Jin G, Tan W, Wu T, Zhang Z, Lin D, Shen H. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat Genet* 2011;43:1215–8.
21. R Core Team (2013). R Foundation for Statistical Computing, Vienna A. R: A language and environment for statistical computing. http://www.r-project.org/ [Internet]. Available from: http://www.r-project.org/
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
23. Ferreira RM, Machado JC, Figueiredo C. Clinical relevance of Helicobacter pylori vacA and cagA genotypes in gastric carcinoma. *Best Pract Res Clin Gastroenterol* 2014;28:1003–15.
24. González CA, Figueiredo C, Lic CB, Ferreira RM, Pardo ML, Ruiz Liso JM, Alonso P, Sala N, Capella G, Sanz-Anquela JM. Helicobacter pylori cagA and vacA

genotypes as predictors of progression of gastric preneoplastic lesions: a long-term follow-up in a high-risk area in Spain. *Am J Gastroenterol* 2011;106:867–74.

25. Naumann M, Sokolova O. NF-κB Signaling in Gastric Cancer. *Toxins (Basel)* 2017;9:119.
26. Kufe DW. Mucins in cancer: function, prognosis and therapy. *Nat Rev Cancer* 2009;9:874–85.
27. Leteurtre E, Zerimech F, Piessen G, Wacrenier A, Leroy X, Copin MC, Mariette C, Aubert JP, Porchet N, Buisine MP. Relationships between mucinous gastric carcinoma, MUC2 expression and survival. *World J Gastroenterol* 2006;12:3324–31.
28. Pasche B, Yi N. Candidate gene association studies: successes and failures. *Curr Opin Genet Dev* 2010;20:257–61.
29. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. *Int Rev Immunol* 2011;30:16–34.
30. de Aguiar BB, Girardi I, Paskulin DD, de Franca E, Dornelles C, Dias FS, Bonorino C, Alho CS. CD14 expression in the first 24h of sepsis: effect of -260C>T CD14 SNP. *Immunol Invest* 2008;37:752–69.
31. Smith M-G, Hold G-L, Tahara E, El-Omar E-M. Cellular and molecular aspects of gastric cancer. *World J Gastroenterol* 2006;12:2979–90.
32. Taniguchi K, Karin M. NF-κB, inflammation, immunity and cancer: coming of age. *Nature Reviews Immunol* 2018; 18: 309-324
33. Cui J, Chen Y, Wang HY, Wang RF. Mechanisms and pathways of innate immune activation and regulation in health and cancer. *Hum Vaccin Immunother* 2014;10(11):3270-85
34. Lebrun J-J. The Dual Role of TGFβ in Human Cancer: From Tumor Suppression to Cancer Metastasis. *ISRN Mol Biol* 2012;2012:1–28.
35. Shrihari T. Dual role of inflammatory mediators in cancer. *Ecancermedicalscience* 2017;11:721.
36. Parisi L, Gini E, Baci D, Tremolati M, Fanuli M, Bassani B, Farronato G, Bruno A, Mortara L. Macrophage Polarization in Chronic Inflammatory Diseases: Killers or Builders? *J Immunol Res* 2018 Jan 14;2018:8917804. doi: 10.1155/2018/8917804. eCollection 2018.
37. Merchant JL. Inflammation, atrophy, gastric cancer: Connecting the molecular dots. *Gastroenterology* 2005;129:1079–82.
38. Dinis-Ribeiro M, Areia M, de Vries AC, Marcos-Pinto R, Monteiro-Soares M, O'Connor A, Pereira C, Pimentel-Nunes P, Correia R, Ensari A, Dumonceau JM, Machado JC, Macedo G, Malfertheiner P, Matysiak-Budnik T, Megraud F, Miki K, O'Morain C, Peek RM, Ponchon T, Ristimaki A, Rembacken B, Carneiro F, Kuipers EJ. Management of precancerous conditions and lesions in the stomach (MAPS): guideline from the European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter Study Group (EHSG), European Society of Pathology (ESP), and the Sociedade Portuguesa de Endoscopia Digestiva (SPED). *Endoscopy* 2012 Jan;44:74-94
39. Malfertheiner P, Megraud F, O'Morain CA, Gisbert JP, Kuipers EJ, Axon AT, Bazzoli F, Gasbarrini A, Atherton J, Graham DY, Hunt R, Moayyedi P, Rokkas T, Rugge M, Selgrad M, Suerbaum S, Sugano K, El-Omar EM; European

Helicobacter and Microbiota Study Group and Consensus panel. Management of Helicobacter pylori infection-the Maastricht V/Florence Consensus Report. *Gut* 2017;66:6–30.

40. Mera RM, Bravo LE, Camargo MC, Bravo JC, Delgado AG, Romero-Gallo J, Yepez MC, Realpe JL, Schneider BG, Morgan DR, Peek RM, Correa P, Wilson KT, Piazuelo MB. Dynamics of Helicobacter pylori infection as a determinant of progression of gastric precancerous lesions: 16-year follow-up of an eradication trial. *Gut* 2018;67:1239-1246.
41. Hwang Y-J, Kim N, Lee HS, Lee JB, Choi YJ, Yoon H, Shin CM, Park YS, Lee DH. Reversibility of atrophic gastritis and intestinal metaplasia after Helicobacter pylori eradication - a prospective study for up to 10 years. *Aliment Pharmacol Ther* 2018;47:380-390.
42. Huang KK, Ramnarayanan K, Zhu F, Srivastava S, Xu C, Tan ALK, Lee M, Tay S, Das K, Xing M, Fatehullah A, Alkaff SMF, Lim TKH, Lee J, Ho KY, Rozen SG, Teh BT, Barker N, Chia CK, Khor C, Ooi CJ, Fock KM, So J, Lim WC, Ling KL, Ang TL, Wong A, Rao J, Rajnakova A, Lim LG, Yap WM, Teh M, Yeoh KG, Tan P. Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer. *Cancer Cell* 2018;33:137-150.e5.
43. Gallagher MD, Chen-Plotkin AS. The post-GWAS Era: From Association to Function. *Am J Hum Genet* 2018;102:717-730.

TABLES

Table 1. Characteristics of the analysed population at recruitment.

Variables	Stables [a] N=259		Regression [b] N=204		Progression[c]N =96		Total N=559		p-value [d]
	N	%	N	%	N	%	N	%	
Sex									
Men	137	52.9	108	52.9	57	59.4	302	54.0	0.51
Women	122	47.1	96	47.1	39	40.6	257	46.0	
Age									
≤50 years	103	39.8	93	45.6	32	33.3	228	40.8	0.12
>50 years	156	60.2	111	54.4	64	66.7	331	59.2	
NSAID consumption									
YES	136	52.5	95	46.6	43	44.8	274	49.0	0.58
NO	122	47.1	102	50	46	47.9	270	48.3	
Missing	1	0.4	7	3.4	7	7.3	15	2.7	
Family history of GC									
YES	40	15.4	18	8.8	16	16.7	74	13.2	0.08
NO	215	83.0	174	85.3	74	77.1	463	82.8	
Missing	4	1.6	12	5.9	6	6.2	22	3.9	
Smoking									
Current or ex-smoker	149	57.5	112	54.9	53	55.2	314	56.2	0.88
Never smoker	109	42.1	88	43.1	37	38.5	234	41.9	
Missing	1	0.4	4	2.0	6	6.3	11	1.9	
***H. pylori* infection**									
YES	172	66.4	134	65.7	55	57.3	361	64.6	0.69
NO	78	30.1	58	28.4	30	31.2	166	29.7	
Missing	9	3.5	12	5.9	11	11.5	32	5.7	
cagA+	117	68.0	89	66.4	35	63.6	241	66.7	
cagA-	54	31.4	45	33.6	20	36.4	119	33.0	0.79
Missing	1	0.6	0	0	0	0	1	0.3	
cagA+vacAs1m1-1	83	48.3	53	39.6	28	51.0	164	45.4	
cagA-vacA no s1m1-1	64	37.2	66	49.2	19	34.5	149	41.3	0.09
Missing	25	14.5	15	11.2	8	14.5	48	13.3	
Histological Diagnosis [e]									
CAG	52	20.1	54	26.5	35	36.5	141	25.2	**< 0.0001**
CIM	116	44.8	56	27.4	43	44.8	215	38.5	
IIM	83	32.0	83	40.7	18	18.7	184	32.9	
DISP	8	3.1	11	5.4	0	0	19	3.4	

[a, b, c] Evolution groups of the gastric cancer precursor lesions (GCPLs), as defined in Materials and Methods section. Percentages are indicated with respect to each evolution group except for *H. pylori cag*A and *vac*A virulence groups, where the percentages are calculated with respect to the total of *H. pylori* infected patients. [d]chi^2 test of comparison. NSAID: Non-steroidal anti-inflammatory drug. [e]CAG: Chronic Atrophic Gastritis, IIM: Incomplete Intestinal Metaplasia, CIM: Complete Intestinal Metaplasia, DISP: Dysplasia.

Table 2. OR (95% CI) of SNPs associated with progression and regression of GCPLs after multinomial logistic regression analysis.

Gene	SNP [a]	Alleles [b]	LRT p-value	FDR q-value	OR (CI 95%) Progression	OR (CI 95%) Regression	Genetic model [c]
CD14	**rs11167532**	G/A	0.0005	**0.017**	**0.5 (0.4-0.8)**	**0.7 (0.5-0.9)**	Log-additive
CD14	rs778588	T/C	0.030	0.456	1.3 (0.8-2.1)	**1.6 (1.1-2.4)**	Dominant
CD14	**rs2569190**	G/A	0.0003	**0.017**	**0.5 (0.4-0.8)**	**0.7 (0.5-0.9)**	Log-additive
CD14	rs5744455	C/T	0.014	0.219	1.8 (1.2-2.7)	1.2 (0.9-1.7)	Log-additive
CD14	**rs1583005**	C/T	0.0004	**0.017**	**1.9 (1.3-2.7)**	**1.5 (1.1-1.9)**	Log-additive
CDH1	rs4783573	A/G	0.049	0.373	0.4 (0.1-1.0)	1.5 (0.7-2.0)	Recessive
MAP3K14	rs16939926	T/C	0.043	0.399	1.3 (0.7-2.4)	0.5 (0.3-1.0)	Log-additive
MAPK3	rs11865086	A/C	0.021	0.446	2.1 (1.2-3.9)	1.5 (1.0-2.3)	Dominant
MUC2	**rs10902073**	C/A	0.0016	**0.041**	0.9 (0.7-1.3)	**0.6 (0.5-0.8)**	Log-additive
MUC2	rs10794281	T/C	0.046	0.373	0.8 (0.4-1.5)	0.5 (0.3-0.9)	Recessive
NFKB1	rs980455	A/G	0.018	0.258	1.4 (1.0-1.9)	1.4 (1.1-1.8)	Log-additive
NFKB1	**rs1598861**	A/C	0.0003	**0.014**	**0.3 (0.1-0.7)**	**0.4 (0.2-0.7)**	Recessive
NFKB1	rs4648090	G/A	0.033	0.399	1.7 (1.0-2.8)	1.6 (1.1-2.4)	Dominant
NFKB1	**rs7674640**	T/C	0.000037	**0.003**	**0.3 (0.2-0.6)**	**0.5 (0.3-0.7)**	Recessive
NFKBIA	rs3138054	G/A	0.024	0.301	0.6 (0.4-1.0)	0.6 (0.4-0.9)	Log-additive
NOD1	rs11536450	C/G	0.031	0.399	1.5 (0.9-2.6)	1.5 (0.9-2.3)	Codominant
PTGS2	rs5275	T/C	0.013	0.219	0.6 (0.4-0.9)	0.8 (0.6-1.05)	Log-additive
PTGS2	rs4648276	T/C	0.030	0.342	0.5 (0.3-0.9)	1.0 (0.7-1.4)	Log-additive
SRC	rs6017901	T/C	0.039	0.399	1.2 (0.6-2.3)	0.5 (0.2-0.9)	Log-additive
TFF2	rs1079380	A/G	0.046	0.399	0.8 (0.5-1.1)	1.2 (0.9-1.6)	Log-additive

[a] SNPs marked in bold are significant after the FDR test. [b] Common allele (reference)/Variant allele. [c] When the same SNP is significant for several genetic models, the most significant value is shown.

Table 3. Haplotypes significantly associated with the evolution of GCPLs.

Gene	SNPs that	Haplotype[a]	Progression			Regression		
			Haplotypefrequency	OR (CI 95%) [b]	p-value	Haplotypefrequency	OR (CI 95%) [c]	p-value
CD14	rs111675	ATAC	0.44	1	------	0.43	1	------
		GTGT	0.21	**2.3 (1.5-**	0.00	0.20	**1.5 (1.0-**	0.03
		GCGC	0.27	**1.7 (1.1-**	0.01	0.27	**1.5 (1.1-**	0.01
CDH1	rs718605	AA	0.39	1	------	0.37	1	------
		AG	0.27	1.1	NS	0.05	**2.3 (1.1-**	0.03
MAP3K14	rs169399	TTCT	0.47	1	------	0.49	1	------
		CTCC	0.32	1.0 (0.7-	NS	0.05	**0.5 (0.3-**	0.05
MUC2	rs109020	CT	0.56	1	------	0.60	1	------
		AC	0.38	0.9 (0.7-	NS	0.33	**0.6 (0.5-**	0.00
NFKB1	rs980455, rs131177	ACCGGG	0.37	1 (Referen	------	0.36	1 (Referen	------
		GCAGA	0.25	**1.6 (1.1-**	0.03	0.26	**1.5 (1.1-**	0.02
		GCAGG	0.06	**2.2 (1.1-**	0.02	0.06	**2.1 (1.2-**	0.00
NFKBIA	rs313805	GA	0.70	1	------	0.71	1	------
		AG	0.15	**0.6 (0.3-**	0.04	0.15	**0.7 (0.5-**	0.04
PTGES	rs107606	AAT	0.48	1	------	0.50	1	------
		GAC	0.10	1.0 (0.6-	NS	0.08	**0.5 (0.3-**	0.02
PTGS2	rs5275, rs464827	TTCGAT	0.21	1 (Referen	------	0.20	1 (Referen	------
		CCGGA	0.15	**0.5 (0.3-**	0.04	0.17	0.9 (0.6-	NS
PTPN11	rs17822304,	CAAGAGC	0.43	1 (Referen	------	0.44	1 (Referen	------
		CAGGA	0.19	0.9 (0.6 -	NS	0.17	**0.7 (0.5-**	0.03
TFF1	rs178841	AGTC	0.38	1	------	0.41	1	------
		AACC	0.11	**2.0 (1.2-**	0.01	0.11	1.2 (0.8-	NS

[a]The underlined alleles indicate which SNPs are associated with evolution of GCPLs. [b, c]Odd ratios (CI 95%) and p-values of haplotypes significantly associated with the evolution of GCPLs. ORs were calculated according to the binomial logistic regression log-additive model with the reference haplotype being the commonest one. NS