Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Hum Genet. Author manuscript; available in PMC 2017 August 01.

Published in final edited form as:
Hum Genet. 2016 August ; 135(8): 895–906. doi:10.1007/s00439-016-1687-1.

Epigenetic and genetic variation in *GATA5* is associated with gastric disease risk

Rafal S. Sobota[#1], **Nuri Kodaman**[#1,2,3], **Robertino Mera**[4], **M. Blanca Piazuelo**[4], **Luis E. Bravo**[5], **Alvaro Pazos**[6], **Jovanny Zabaleta**[7], **Alberto G. Delgado**[4], **Wael El-Rifai**[8,9], **Douglas R. Morgan**[4], **Keith T. Wilson**[4,9], **Pelayo Correa**[4], **Scott M. Williams**[2,3], and **Barbara G. Schneider**[4]

[1] Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

[2] Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH, USA

[3] Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA

[4] Division of Gastroenterology, Department of Medicine, Vanderbilt University Medical Center, 2215 Garland Avenue, Nashville, TN 37232, USA

[5] Department of Pathology, School of Medicine, Universidad del Valle, Cali 760043, Colombia

[6] Department of Biology, University of Nariño, Pasto 520002, Colombia

[7] Stanley S. Scott Cancer Center, Louisiana State University Health Sciences Center, New Orleans, LA, USA

[8] Department of Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

[9] Department of Veterans Affairs, Veterans Affairs Tennessee Valley Healthcare System and Office of Medical Research, Nashville, TN, USA

[#] These authors contributed equally to this work.

Abstract

Gastric cancer incidence varies considerably among populations, even those with comparable rates of *Helicobacter pylori* infection. To test the hypothesis that genetic variation plays a role in gastric disease, we assessed the relationship between genotypes and gastric histopathology in a Colombian study population, using a genotyping array of immune-related single nucleotide polymorphisms (SNPs). Two synonymous SNPs (rs6061243 and rs6587239) were associated with progression of premalignant gastric lesions in a dominant-effects model after correction for multiple comparisons ($p = 2.63\text{E–}07$ and $p = 7.97\text{E–}07$, respectively); effect sizes were $\beta = -0.863$ and $\beta = -0.815$, respectively, where β is an estimate of effect on histopathology scores, which

Barbara G. Schneider, barbara.schneider@vanderbilt.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-016-1687-1) contains supplementary material, which is available to authorized users.

Compliance with ethical standards
Conflict of interest The authors have no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ranged from 1 (normal) to 5 (dysplasia). In our replication cohort, a second Colombian population, both SNPs were associated with histopathology when additively modeled ($\beta = -0.256$, 95 % CI = −0.47, −0.039; and $\beta = -0.239$, 95 % CI = −0.45, −0.024), and rs6587239 was significantly associated in a dominant-effects model ($\beta = -0.330$, 95 % CI = −0.66, 0.00). Because promoter methylation of *GATA5* has previously been associated with gastric cancer, we also tested for the association of methylation status with more advanced histopathology scores in our samples and found a significant relationship ($p = 0.001$). A multivariate regression model revealed that the effects of both the promoter methylation and the exonic SNPs in *GATA5* were independent. A SNP-by-methylation interaction term was also significant. This interaction between *GATA5* variants and *GATA5* promoter methylation indicates that the association of either factor with gastric disease progression is modified by the other.

Introduction

Gastric cancer is the third most common cause of cancer-related mortality worldwide, accounting for an estimated of 723,000 deaths annually (Ferlay et al. 2014). Lack of symptoms early in the course of disease often leads to late diagnoses and fatal outcomes. Most gastric cancers are adenocarcinomas, which may be classified into intestinal and diffuse subtypes. Intestinal-type gastric adenocarcinoma, the most common subtype in high-risk regions, is associated with a well-defined histopathological cascade: chronic superficial gastritis, atrophic gastritis, intestinal metaplasia, dysplasia and adenocarcinoma (Correa 1988). Typically, these lesions advance over decades, affording opportunity for early detection and interventions.

Most cases of gastric cancer can be attributed to infection with *Helicobacter pylori*, a class I carcinogen (IARC Working Group 1994). Gastric disease progression to noncardia gastric cancer depends in part on genetic variation in the infecting *H. pylori* strain. In some strains, insertion of a 40-kb *cag* pathogenicity island containing the *cagA* virulence factor causes a severe inflammatory response and increases the risk of gastric cancer (Blaser et al. 1995; Covacci et al. 1993). Variants of another *H. pylori* gene, *vacA*, also increase gastric cancer risk by inducing vacuole formation, stimulating epithelial-cell apoptosis, and suppressing host T cell responses to the bacterium (Atherton et al. 1995).

Genetic variation in the host also influences the clinical trajectory of gastric disease. Polymorphisms known to modulate levels of cytokines IL-1β, TNF-α, and IL-10 have been associated with elevated risk of non-cardia gastric cancer (El-Omar et al. 2000, 2003). Additionally, genome-wide association studies (GWAS) in Asian populations have identified associations between gastric cancer and variants in *PSCA* and *MUC1* (Abnet et al. 2010; Sakamoto et al. 2008; Shi et al. 2011) and near *PRKAA1* and *PTGER4* (Shi et al. 2011; Song et al. 2013). The interaction of genetic variants in the host and pathogen also affects disease progression (Kodaman et al. 2014a, b).

In addition to host genetic factors, epigenetic variation, specifically changes in DNA methylation, have been associated with the development of gastric cancer. Such epigenetic changes often involve DNA promoter methylation and subsequent inactivation of tumor suppressor genes (The Cancer Genome Atlas Research Network 2014). Many changes in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DNA methylation occur non-specifically, as a function of aging or environmental factors (Teschendorff et al. 2010). Some changes may occur in premalignant lesions, making methylation changes potentially useful as biomarkers of gastric cancer risk (Schneider et al. 2015; Watanabe et al. 2009).

In the present study, we tested for association between human genetic variation and gastric disease severity, using gastric biopsies from patients residing in Colombian regions with high and low risk for gastric cancer. Genes relevant to immunological function were examined. Identifying such variants may improve our understanding of gastric cancer etiology, guide therapy, and provide genetic markers of disease progression risk.

Materials and methods

Study participants

Adult patients presenting with dyspeptic symptoms that merited upper gastrointestinal endoscopy were recruited from two regions in the state of Nariño, Colombia: Tumaco on the Pacific coast and the Pasto region in the Andes Mountains that, respectively, have low and high incidence rates for gastric cancer (Bravo et al. 2002). Patients with chronic conditions such as diabetes, heart disease, or a prior gastrectomy were excluded, as were those who had been treated with H_2-receptor antagonists, proton pump inhibitors, or antimicrobials during the month prior to the endoscopy. Human subjects approvals were obtained from the Institutional Review Boards of the participating hospitals, the Universidad del Valle in Cali, Colombia, and Vanderbilt University. Informed consent was obtained from each subject. The discovery cohort included 130 patients, and 159 patients from an independent set comprised a replication cohort. Samples from 189 patients from both cohorts were assessed for percent DNA methylation (Methylation group).

Biopsies and histopathology scoring

Gastric mucosa biopsy samples were taken from three sites: the antrum (greater curvature, within 3 cm of the pylorus), incisura angularis (lesser curvature), and corpus (middle anterior wall). After staining with hematoxylin and eosin, tissue sections were evaluated independently by two pathologists blinded to sample origin. Diagnostic categories for histopathology scoring employed the updated Sydney system for gastritis (Dixon et al. 1996) and the Padova International Classification for Dysplasia (Rugge et al. 2000) as previously described (Kodaman et al. 2014a). Cases with discordant diagnoses were reviewed until a consensus was reached. Ordinal values for diagnostic categories were as follows: 1 = normal, 2 = non-atrophic gastritis (NAG), 3 = multifocal atrophic gastritis without intestinal metaplasia (MAG), 4 = intestinal metaplasia (IM), and 5 = dysplasia. The most severe lesion observed in all biopsies from an individual represented that person's final diagnosis in our study.

Helicobacter pylori culture and characterization

For the discovery cohort, virulence determinants *cagA* and *vacA* s were characterized as previously described (Kodaman et al. 2014a). In the replication cohort and methylation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

group, for 104 subjects for whom single colony cultures were not available, the *vacA* s region was characterized from biopsy DNA as previously described (Schneider et al. 2015).

Genotyping in the discovery cohort

Human DNA from the study participants was extracted from blood samples using Puregene kits (Qiagen, Valencia, CA, USA). DNA was genotyped using the Illumina platform Immunochip array, which assays 196,524 SNPs in genes primarily involved in immune response (Cortes and Brown 2011). If fewer than 95 % of samples could be genotyped for a given SNP, that SNP was removed from further analyses. Study participants for whom fewer than 90 % of SNPs could be genotyped were excluded from analyses. A priori power analyses suggested that a minor allele frequency threshold of 0.2 would allow us to detect a beta of 0.5 with power of 0.90, given our sample size (Table S1). SNPs in perfect linkage disequilibrium (LD) and those failing the Hardy–Weinberg test at the 0.01 level were also excluded; 39,802 markers passed quality control and were used in the final analysis (Table S2).

Characterizing human ancestry in the discovery cohort

Estimation of human ancestry was performed as previously described (Kodaman et al. 2014a). Briefly, all Immunochip SNPs in LD were removed, using an r^2 threshold >0.1 prior to analysis (performed in PLINK (Purcell et al. 2007)) and the remaining SNPs were used to estimate ancestry. Using an admixture model in STRUCTURE, (Falush et al. 2003; Pritchard et al. 2000) model probability was maximized using $K = 3$; the inferred ancestral groups matched the expected distribution of Amerindian, African, and European ancestry (Wang et al. 2007, 2008). To validate this assumption, we obtained genotype data for the European CEU and African YRI populations from Phase 2 of HapMap (International Hapmap Consortium 2005) the Spanish Iberian (IBS) population from the 1000 Genomes Project (Abecasis et al. 2012), and the H952 subset of the Karitania, Surui, and Colombian populations in the Human Genome Diversity Project (Cavalli-Sforza 2005). After merging these data with those from our own population, the final dataset contained 5947 SNPs from 514 individuals. The admixture model of STRUCTURE was run 10 times, with 50,000 iterations after a burn-in of 50,000 iterations. Replicate runs were collated in CLUMPP (Jakobsson and Rosenberg 2007) and means of individual ancestry were calculated.

Genotyping in the replication cohort

SNP analyses were performed using Pyrosequencing®, as follows: 20–100 ng of DNA was amplified using forward and reverse primers (rs6061243 forward primer: biotin-CTTACCTGGGGGGCCATG; reverse primer: CTCTTCGCTTCCAGGATCCAC; sequencing primer: GCTCAGCAGCCACTT; rs6587239 forward primer: CTTCCGCTTCCGTGTCTG; reverse primer: biotin-GGAATGAGGGGGTGCTGT; sequencing primer: TCCGTGTCTGGATGC; rs1046587 forward primer: biotin-GGGAGTGGAGTGGAAGAAGA; reverse primer: CACTCGGCCCTCACTTTG; sequencing primer: GGCCCTCACTTTGCT). Biotinylated strands of PCR products were purified on streptavidin-sepharose beads. Sequencing primers were annealed, following the manufacturer's recommendations (Qiagen, Valencia, CA, USA). Assignment of alleles was performed using a Pyromark MD Pyrosequencing instrument, in accordance with the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

manufacturer's instructions. Assignment of alleles was confirmed by comparing the Immunochip array with Pyrosequencing results, using 36 samples that were analyzed by both methods. Concordance in the 36 samples was 100 % between the two methods. Alleles were designated ancestral if identical with those in chimpanzee, or derived if different.

GATA5 methylation analysis in the methylation group

We performed a *GATA5* methylation analysis on samples from both cohorts. DNA (up to 2 ug) was bisulfite-modified using a DNA Methylation Direct kit (Zymo Research Corp., Irvine, CA). Levels of DNA methylation were quantitated (in 30–60 ng modified DNA per reaction) by Pyrosequencing®. *GATA5* Pyrosequencing primers were F: GGGATTAGTATAAATTTGGAGAAGTT; R: biotin-CCACCTAACCCTAACAAACCCTACT; Sequencing primer: GGTTAGTTAGTTTAGTTGTATT. This assay quantitates methylation across chr20:61,051,250 to 61,051,265 (hg19). The initial PCR cycling program consisted of 95 °C for 15 min, then 46 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s, followed by a 10 min incubation at 72 °C. The initial PCR reagents included two units of Amplitaq Gold Hot Start DNA polymerase (Life Technologies, Carlsbad, CA) per reaction. The biotinylated strand of each PCR product was isolated from 15 μl of each PCR product, using streptavidin beads (GE Healthcare Biosciences, Piscataway, NJ, USA). Pyrosequencing reactions were performed using a PyroMark MD Pyrosequencing instrument, as recommended by the manufacturer (Qiagen, Valencia, CA, USA). Bisulfite-modified methylated HeLa DNA (New England Biolabs, Ipswich, MA, USA) was employed as a positive control with each experiment; bisulfite-modified pooled normal human blood DNA (Promega Corp., Madison, WI, USA) and no-template buffer were used as negative controls, at least one of which was used with each experiment. Results were reported as percentage of DNA molecules that are methylated.

Statistical analyses of the discovery and replication cohorts

In the discovery cohort, univariate linear regression analyses modeling histopathology scores were first performed with each available variable: participant age, region of origin, sex, Amerindian ancestry, European ancestry, African ancestry, and *H. pylori* virulence factors *cagA* and *vacA* s. Significant variables were used in the final model. Patient genotypes were assessed using both additive and dominant models. Recessive models would have been underpowered and were not performed. A Bonferroni correction was used to determine the appropriate level of statistical significance using the number of SNPs passing quality controls (prior to any analyses) as the number of independent tests performed (0.05/(39,802) $= 1.26 \times 10^{-6}$).

The SNPs significant in the discovery cohort were tested for association in the replication cohort using both additive and dominant models. Ancestry estimates could not be ascertained for the replication cohort because genome-wide data were not available; therefore, analyses were adjusted for region of origin as proxy for ancestry. Analyses were also adjusted for the *vacA* s virulence factor.

Author Manuscript

Genetic association analyses in the discovery and replication cohorts were performed using PLINK (Purcell et al. 2007) and STATA v11.1, (Statacorp, College Station, TX, USA), respectively. LD of significant variants was assessed using Haploview (Barrett et al. 2005). The qqman package in R (Turner 2014) was used to create QQ and Manhattan plots. A 50 kb region surrounding the SNP with the strongest association was plotted with Locus Zoom (Pruim et al. 2010). Variants in regions of interest (within 0.5 megabase) were imputed with IMPUTE2 (v2.3.1), using one multipopulation phased reference panel from the 1000 Genomes project (Abecasis et al. 2012; Howie et al. 2009). Linear regression on histopathology scores in the discovery cohort was also performed using imputed variants, adjusting for the same covariates as in the original analysis.

Statistical analysis of the methylation group

Linear regression was used to test the methylation level of the *GATA5* promoter for association with gastric histopathology, using STATA v11.1. Histopathology was modeled in univariate analyses with participant age, region of origin, sex, *vacA* s status, dominant rs6587239 genotype, and *GATA5* methylation as a continuous variable. Variables significant in univariate analyses were included in the final model. To assess whether (dominant) genotype or methylation had non-linear effects on histopathology, an interaction term (genotype-by-methylation) was, then added to the multivariate model. Similar analyses were performed using ordinal regression.

Results

Discovery cohort

The discovery cohort consisted of 130 patients (mean age 48.7 ± 7.0), of whom 78 lived in the low risk coastal region and 52 in the high-risk mountain region (Table 1). Human ancestry and *H. pylori* virulence factors were both significantly associated with histopathology score and advanced gastric disease in this cohort (Table 2). As expected, the presence of *cagA* in *H. pylori* was associated with higher histopathology scores ($\beta = 0.55$, $p = 0.037$), while *vacA* s2 alleles were associated with less severe lesions ($\beta = -0.76$, $p = 0.001$). In Colombia, continental human ancestry is known to be primarily African, Amerindian, and European (Kodaman et al. 2014a); in our study population, we found that human African ancestry was associated with less severe lesions ($\beta = -0.71$, $p = 0.005$) and European and Amerindian ancestry with more severe lesions ($\beta = 1.65$ and 0.65, $p = 0.004$ and 0.046, respectively). To adjust for all three ancestries in subsequent regression models, we included two of them, which by default defined the third. Neither region of origin, sex, nor age was independently associated with histopathology scores (Table 2). The independent variables in the final model were therefore genotype, African ancestry, Amerindian ancestry, and *cagA*/*vacA* s. The genomic inflation factor was 1.02 in these analyses, indicating that population stratification was adequately controlled (Figure S1).

In tests of dominant effects, two SNPs in the *GATA5* gene (rs6061243 and rs6587239) were significantly associated with histopathology scores after correction for multiple testing ($p = 2.63 \times 10^{-7}$ and 7.97×10^{-7}, respectively) (Table 3; Fig. 1). Ordinal regression modeling of histopathology scores produced similar associations (Table S3). Both rs6061243 (C/G) and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

rs6587239 (C/T) are synonymous exonic SNPs in strong linkage disequilibrium with each other ($r^2 = 0.98$), consistent with a single association at this locus (Figure S2A). These variants correspond to c.981G>C, Ser327Ser in the coding strand in exon 5 and c.852G>A, Lys284Lys in the coding strand in exon 4, respectively. In both cases, presence of the derived allele (G for rs6061243 and T for rs6587239) was associated with lower histopathology scores ($\beta = -0.86$ and $\beta = -0.82$, respectively, in the dominant model) where β estimates the change in histopathology score with presence of the minor allele. The effect sizes in additive-effects models, where β estimates the change in histopathology score with additional copies of the minor allele, were similar, but not as strong ($\beta = -0.41$, $p = 4.92 \times 10^{-4}$; and $\beta = -0.40$, $p = 7.35 \times 10^{-4}$, respectively).

Linear regression on histopathology scores in the discovery cohort was also performed using imputed variants, adjusting for the same covariates as in the original analysis (Table S4; Fig. 1). No additional imputed SNPs were identified as more significantly associated than rs6587239.

Four other SNPs were significant at $p < 10^{-4}$ (Table 3). The minor allele (A) of rs1046587 in the 3′ UTR region of *LTB4R* was associated with higher histopathology scores ($\beta = 0.71$) in the dominant-effects model. An intergenic locus rs1113665 was also significant ($p = 9.64 \times 10^{-5}$). Two SNPs near *PAM* on chromosome 5 (rs158250 and rs464720) were significant at this threshold in tests using an additive genetic model ($\beta = -0.57$ and -0.52, respectively). The SNPs near PAM were in strong linkage disequilibrium ($r^2 = 0.93$) (Figure S2B).

Replication cohort

The replication cohort consisted of 159 patients (see Table 1 for demographic data). Distributions of sex and age did not differ significantly from those of the discovery cohort ($p = 0.565$ and 0.109, respectively), but the mean histopathology score was significantly higher ($p = 1.98 \times 10^{-5}$) (Tables 1, S5), and a greater percentage of individuals were residents of the mountain region ($p = 1.58 \times 10^{-14}$). The *vacA* s2 allele frequency among *H. pylori* samples was also greater in this cohort, although not statistically significant ($p = 0.051$). All models were subsequently adjusted for *vacA* s and region of origin, which were significant in univariate analyses of gastric histopathology scores. We replicated the association between rs6587239 and gastric histopathology scores in tests of both additive and dominant effects ($\beta = -0.26$, $p = 0.021$; $\beta = -0.33$, $p = 0.050$; respectively; Table 4), with the derived allele (T) being associated with lower histopathology scores. While rs6061243 was significantly associated with histopathology only in the additive-effects model ($\beta = -0.24$, $p = 0.029$), the direction of effect was consistent in the dominant model ($\beta = -0.30$, $p = 0.076$). Ordinal regression modeling of histopathology scores produced similar results (Table S6). The p value of the association of SNP rs1046587 on chromosome 14 (Figure S3) and its proximity to *LTB4R,* a leukotriene receptor, prompted us to test this SNP in the Replication cohort. However, it was not significant ($p = 0.427$).

Methylation group

Because promoter methylation in *GATA5* has been previously associated with gastric cancer, we assessed whether methylation was associated with severity of gastric lesions in our study.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We quantitated the levels of *GATA5* promoter methylation in 189 individuals (Table 1), and found a significant positive association between the degree of methylation and histopathology scores in an unadjusted model ($\beta = 0.028$, $p = 0.001$) (Table 5). Homozygosity of the rs6587239-C allele was correlated with significantly higher methylation (Spearman $\rho = -0.157$, $p = 0.031$).

Multivariate tests of genetic and epigenetic variation in *GATA5*

Given that SNPs in *GATA5* and DNA methylation patterns in its promoter were both significantly associated with disease severity, we tested both variables in a single regression model. Region of origin, age, and *vacA* s were included as covariates, as they were all associated with histopathology scores in univariate tests ($p = 0.008$, 0.009, and 0.002, respectively). Associations with histopathology scores were significant for rs6587239 and *GATA5* promoter methylation in the multivariate test ($p = 0.039$ and $p = 0.018$, respectively; Table S7), indicating independent effects.

To assess whether the effects of the derived allele of rs6587239 on histopathology score varied non-linearly with the level of methylation, we added a methylation-by-genotype interaction term to the above model. The interaction was significant ($p = 0.033$), as were age, *vacA* s, and rs6587239. However, methylation was not significant in the model that included the interaction term (Table 5). These results demonstrated that the combined effect of genotype and methylation on histopathology scores was non-linear; at low levels of methylation, individuals who were homozygous for the C allele at rs6587239 presented with more advanced lesions (>1 unit, on average, corresponding to progression from multifocal atrophic gastritis to intestinal metaplasia) compared to individuals who had at least one T allele. This effect on progression was attenuated at higher levels of methylation, where those with CC genotypes had scores less than a half unit more advanced, on average (Fig. 1d). Analyses of rs6061243 produced similar results to that of rs6587239 with the ancestral allele (C) associating with higher scores (Table S8). Ordinal regression analyses showed a similar interaction effect and loss of significance of methylation (Supplemental Table S9).

Validation of previous GWAS results in the discovery cohort

In our analyses of the discovery cohort, we replicated several genes previously reported to associate with intestinal gastric cancer in published GWAS (Shi et al. 2011; Song et al. 2013). Three intronic SNPs in *ZBTB20*, (rs4399857, rs6438214, and rs9860965) had p values below 0.05 when modeled both additively ($p = 8.57 \times 10^{-3}$, 0.0160 and 6.19×10^{-4}, respectively) and dominantly ($p = 6.97 \times 10^{-3}$, 0.014, and 4.23×10^{-4}, respectively) (Table S10). The SNPs rs4399857 and rs6438214 were in strong LD ($r^2 = 0.94$) with each other, but in weaker LD with rs9860965 ($r^2 = 0.46$ and 0.44, respectively) (Figure S2C). Additionally, rs7725052, a SNP in the previously associated 5p13.1 region that includes genes *PTGER4* and *PRKAA1* (Shi et al. 2011; Song et al. 2013), was significant in the dominant-effects model ($p = 0.037$). Another gene previously associated with gastric cancer, *MUC1* (Saeki et al. 2013), was not well tagged by the SNPs in this analysis, but intronic SNPs in genes on either side of *MUC1* (rs11264341 in *TRIM46* and rs2075571 in *THBS3*), were significant in additive-effects models ($p = 0.031$ and 0.023, respectively); rs11264341 is ~10 kb upstream of *MUC1* while rs2075571 is ~12 kb downstream, and they are in strong

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

LD with each other (r^2 = 0.95) (Figure S2D). In a meta-analysis of gene association studies for various subtypes of gastric cancer in different populations, Mocellin et al. (2015) reported other genes in addition to those listed above. Concordant with their results, we found these variants significant in dominant models: rs7635274 near *TGFBR2* (p = 0.027), rs7743193 near *TNF* (p = 0.039), and rs10882332 and rs4304697 near *PCLE1* (p = 0.044 and 0.022, respectively). We were unable to replicate the association of advanced disease with rs2294008 in *PSCA* (Table S10) (Saeki et al. 2013; Sakamoto et al. 2008; Sala et al. 2012). We detected no significant associations of variants near *MTX1*, which is related to risk of diffuse gastric cancer; or to variants near *CASP8* or to those near *GSTP1*, which was found significantly related to gastric cancer risk in Asian but not Caucasian populations (Mocellin et al. 2015).

Discussion

We observed and replicated a novel association between severity of gastric lesions and exonic SNPs in *GATA5*. Although the effects of these synonymous variants are unclear, it is possible that they induce differential gene expression through changes in transcript stability, editing, and binding affinity of protein to DNA (Fernandez-Calero et al. 2014). *GATA5* is a gene involved in cell and tissue differentiation and gastrointestinal development. It encodes a lineage-restricted transcription factor, expressed in the stem cell compartment in the gut, and this transcription factor is upregulated as these cells undergo terminal differentiation (Gao et al. 1998; Laverriere et al. 1994). Importantly, *GATA5* is expressed in gastric epithelial cells, as indicated by in situ hybridization and immunohistochemistry studies (Laverriere et al. 1994; Wen et al. 2010); data from the Genotype-Tissue Expression (GTEx) study identify stomach as one of the tissues with the highest expression of GATA5 among the 53 tested (Mele et al. 2015). Consistent with the role of promoting differentiation, GATA5 activates the promoter of pepsinogen A, a digestive enzyme produced by gastric chief cells (Sakamoto et al. 2000). Other genes regulated by GATA5 include trefoil factors 1 and 2 (*TFF1, TFF2)*, and inhibin α (*INHA)* (Akiyama et al. 2003).

In its capacity as a transcription factor, GATA5 is highly pleiotropic, and thus, polymorphisms that modulate its expression may influence gastric disease in multiple ways. One function of possible relevance for gastric carcinogenesis is that GATA5 activates transcription of TFF1 and TFF2 (Akiyama et al. 2003). TFF1 in particular is a well-established tumor suppressor in gastric tissue: TFF1 knockout mice develop adenocarcinomas of the gastric antrum, and human gastric biopsies show a decrease in TFF1 expression as gastric lesions progress along the multi-step carcinogenic cascade (Soutto et al. 2011). This decrease in TFF1 expression is accompanied by an increase in NFκB activation, which promotes the inflammatory response to *H. pylori* (Soutto et al. 2015). Furthermore, TFF1 induces p53 protein expression by downregulating its negative regulator, miR-504 (Soutto et al. 2014).

In vitro evidence indicates that GATA5 acts as a tumor suppressor in gut and in other cell types. Over-expression of GATA5 reduces colony formation, cell proliferation, migration, invasion and anchorage-independent growth in colorectal cancer cells (Hellebrekers et al. 2009); decreased expression has been implicated in carcinogenesis in both gastric and renal

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cancer cell lines as well as in colorectal and invasive ductal breast cancer tissue specimens (Hellebrekers et al. 2009; Peters et al. 2014; Wang et al. 2014). Inactivation of the likely tumor suppressor *GATA5* by promoter methylation was previously reported in gastric and colorectal cell lines, gastric cancers and dysplasias (Akiyama et al. 2003; Wang et al. 2011; Wen et al. 2010). Methylation of CpG sites within enhancer sequences near *GATA5* alters a transcription factor USF1 binding site, downregulating GATA5 expression in the gut (Chen et al. 2012). Wen et al. (2010) found that *GATA5* promoter methylation was associated with loss of GATA5 protein in human gastric adenocarcinomas. These observations are consistent with our methylation analyses in which increased methylation was associated with higher histopathology scores, as tissues progressed in the premalignant cascade. The mechanism by which *GATA5* promoter methylation is associated with gastric disease progression may operate directly or indirectly via *TFF1, TFF2* or other target genes.

A multivariate test adjusted for relevant covariates indicated that *GATA5* promoter methylation and rs6587239 genotype CC were independently associated with higher histopathology scores. We then added a methylation-by-genotype interaction term, as a significant interaction would indicate that the effect of one variable was modified by the levels of the other. The interaction was significant, which may indicate that the association of the CC genotype with more advanced lesions is attenuated when cells with this *GATA5* genotype become inactivated by promoter methylation. Alternatively, the relationship between SNP status and methylation may signify that one or both of our associating SNPs is a methylation quantitative trait locus (meQTL), a variant associated with quantitative differences in methylation levels (Kerkel et al. 2008). Indeed, the association between the CC genotype and increased methylation supports this possibility. Future studies examining meQTLs in stomach and other portions of the gastrointestinal tract would be of interest, especially because there is a relatively high incidence of methylation defects in tumors of the GI tract, as shown in Figure S4.6 of the Supplement to The Cancer Genome Atlas paper on gastric adenocarcinoma (The Cancer Genome Atlas Research Network 2014).

In conclusion, we showed genetic and epigenetic data that together provide compelling evidence to support *GATA5* promoter methylation and exonic variants as risk factors in the development of gastric disease. The associations we presented for each of these data types were also affected by an interaction between them. This observed interaction between genotype and methylation status together with the negative correlation between genotype and methylation indicate a complex relationship between genotype and epigenetics that likely contributes to risk of gastric disease progression.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, the National Center for Advancing Translational Sciences, the National Institutes of Health, or Vanderbilt University Medical Center.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Acknowledgments

We wish to acknowledge Ms. Jacki Sands (Qiagen) for assistance in assay design for Pyrosequencing. This study was supported by the National Center for Advancing Translational Sciences, Grant 5UL1TR000445; and Grant 2P01CA028842 from the National Cancer Institute. This work was supported by Public Health Service award 5T32 GM007347 from the National Institute of General Medical Sciences for the Vanderbilt Medical-Scientist Training Program. SMW, RSS and NK were partially supported by National Institutes of Health Grant 5P20 GM103534.

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. doi:10.1038/nature11632. [PubMed: 23128226]

Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO, Yuan JM, Zheng W, Dawsey SM, Dong LM, Lee MP, Ding T, Qiao YL, Gao YT, Koh WP, Xiang YB, Tang ZZ, Fan JH, Wang C, Wheeler W, Gail MH, Yeager M, Yuenger J, Hutchinson A, Jacobs KB, Giffen CA, Burdett L, Fraumeni JF Jr, Tucker MA, Chow WH, Goldstein AM, Chanock SJ, Taylor PR. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet. 2010; 42:764–767. doi:10.1038/ng.649. [PubMed: 20729852]

Akiyama Y, Watkins N, Suzuki H, Jair KW, van Engeland M, Esteller M, Sakai H, Ren CY, Yuasa Y, Herman JG, Baylin SB. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. Mol Cell Biol. 2003; 23:8429–8439. [PubMed: 14612389]

Atherton JC, Cao P, Peek RM. Mosaicism in vacuolating cytotoxin alleles of [i]Helicobacter pylori[/i]. Association of specific vacA types with cytotoxin production and peptic ulceration. J Biol Chem. 1995; 270:17771–17777. [PubMed: 7629077]

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. doi:10.1093/bioinformatics/bth457. [PubMed: 15297300]

Blaser MJ, Perez-Perez GI, Kleanthou TL, Cover RM, Peek RM, Chyou PH, Stemmermann GN, Nomura A. Infection with [i]Helicobacter pylori[/i] strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res. 1995; 55:2111–2115. [PubMed: 7743510]

Bravo LE, van Doorn LJ, Realpe JL, Correa P. Virulence-associated genotypes of [i]Helicobacter pylori[/i]: do they explain the African enigma? Am J Gastroenterol. 2002; 97:2839–2842. [PubMed: 12425557]

Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. Nat Rev Genet. 2005; 6:333–340. doi:10.1038/nrg1596. [PubMed: 15803201]

Chen B, Hsu R, Li Z, Kogut PC, Du Q, Rouser K, Camoretti-Mercado B, Solway J. Upstream stimulatory factor 1 activates GATA5 expression through an E-box motif. Biochem J. 2012; 446:89–98. doi:10.1042/bj20111942. [PubMed: 22625849]

Correa P. A human model of gastric carcinogenesis. Cancer Res. 1988; 48:3554–3560. [PubMed: 3288329]

Cortes A, Brown MA. Promise and pitfalls of the immunochip. Arthritis Res Ther. 2011; 13:101. doi: 10.1186/ar3204. [PubMed: 21345260]

Covacci A, Censini S, Bugnoli M, Petracca R, Burroni D, Macchia G, Massone A, Papini E, Ziang Z, Figura N, Rappuoli R. Molecular characterization of the 128-kDa immunodominant antigen of [i]Helicobacter pylori[/i] associated with cytotoxicity and duodenal ulcer. Proc Nat Acad Sci USA. 1993; 90:5791–5795. [PubMed: 8516329]

Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol. 1996; 20:1161–1181. [PubMed: 8827022]

El-Omar EM, Carrington M, Chow WH, McColl KE, Bream JH, Young HA, Herrera JE, Lissowska J, Yuan CC, Rothman N, Lanyon G, Martin M, Fraumeni JF Jr, Rabkin CS. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature. 2000; 404:398–402. [PubMed: 10746728]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

El-Omar EM, Rabkin CS, Gammon MD, Vaughan TL, Risch HA, Schoenberg JB, Stanford JL, Mayne ST, Goedert J, Blot WJ, Fraumeni JF Jr, Chow WH. Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. Gastroenterology. 2003; 124:1193–1201. [PubMed: 12730860]

Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164:1567–1587. [PubMed: 12930761]

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide. IARC CancerBase. 2014; 11 [15 Jan 2014] http://globocan.iarc.fr.

Fernandez-Calero T, Astrada S, Alberti A, Horjales S, Arnal JF, Rovira C, Bollati-Fogolin M, Flouriot G, Marin M. The transcriptional activities and cellular localization of the human estrogen receptor alpha are affected by the synonymous Ala87 mutation. J Steroid Biochem Mol Biol. 2014; 143:99–104. doi:10.1016/j.jsbmb.2014.02.016. [PubMed: 24607813]

Gao X, Sedgwick T, Shi YB, Evans T. Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. Mol Cell Biol. 1998; 18:2901–2911. [PubMed: 9566909]

Hellebrekers DM, Lentjes MH, van den Bosch SM, Melotte V, Wouters KA, Daenen KL, Smits KM, Akiyama Y, Yuasa Y, Sanduleanu S, Khalid-de Bakker CA, Jonkers D, Weijenberg MP, Louwagie J, van Criekinge W, Carvalho B, Meijer GA, Baylin SB, Herman JG, de Bruine AP, van Engeland M. GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. Clin Cancer Res. 2009; 15:3990–3997. doi:10.1158/1078-0432.ccr-09-0055. [PubMed: 19509152]

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. doi:10.1371/journal.pgen.1000529. [PubMed: 19543373]

IARC Working Group. IARC monograph on the evaluation of carcinogenic risks to humans: schistosomes, liver flukes and [i]Helicobacter pylori[/i]. International Agency for Research on Cancer; Lyon: 1994.

International Hapmap Consortium. A haplotype map of the human genome. Nature. 2005; 437:1299–1320. doi:10.1038/nature04226. [PubMed: 16255080]

Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007; 23:1801–1806. doi:10.1093/bioinformatics/btm233. [PubMed: 17485429]

Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet. 2008; 40:904–908. doi:10.1038/ng.174. [PubMed: 18568024]

Kodaman N, Pazos A, Schneider BG, Piazuelo MB, Mera R, Sobota RS, Sicinschi LA, Shaffer CL, Romero-Gallo J, de Sablet T, Harder RH, Bravo LE, Peek RM Jr, Wilson KT, Cover TL, Williams SM, Correa P. Human and [i]Helicobacter pylori[/i] coevolution shapes the risk of gastric disease. Proc Natl Acad Sci USA. 2014a; 111:1455–1460. doi:10.1073/pnas.1318093111. [PubMed: 24474772]

Kodaman N, Sobota RS, Mera R, Schneider BG, Williams SM. Disrupted human-pathogen co-evolution: a model for disease. Front Genet. 2014b; 5:290. doi:10.3389/fgene.2014.00290. [PubMed: 25202324]

Laverriere AC, MacNeill C, Mueller C, Poelmann RE, Burch JB, Evans T. GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. J Biol Chem. 1994; 269:23177–23184. [PubMed: 8083222]

Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segre AV, Djebali S, Niarchou A, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigo R. Human genomics. The human transcriptome across tissues and individuals. Science. 2015; 348:660–665. doi:10.1126/science.aaa0355. [PubMed: 25954002]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Mocellin S, Verdi D, Pooley KA, Nitti D. Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. Gut. 2015; 64:1209–1219. doi:10.1136/gutjnl-2015-309168. [PubMed: 25731870]

Peters I, Dubrowinskaja N, Kogosov M, Abbas M, Hennenlotter J, von Klot C, Merseburger AS, Stenzl A, Scherer R, Kuczyk MA, Serth J. Decreased GATA5 mRNA expression associates with CpG island methylation and shortened recurrence-free survival in clear cell renal cell carcinoma. BMC Cancer. 2014; 14:101. doi:10.1186/1471-2407-14-101. [PubMed: 24533449]

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–2337. doi:10.1093/bioinformatics/btq419. [PubMed: 20634204]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. doi:10.1086/519795. [PubMed: 17701901]

Rugge M, Correa P, Dixon MF, Hattori T, Leandro G, Lewin K, Riddell RH, Sipponen P, Watanabe H. Gastric dysplasia—the Padova international classification. Am J Surg Pathol. 2000; 24:167–176. [PubMed: 10680883]

Saeki N, Ono H, Sakamoto H, Yoshida T. Genetic factors related to gastric cancer susceptibility identified using a genome-wide association study. Cancer Sci. 2013; 104:1–8. doi:10.1111/cas.12042. [PubMed: 23057512]

Sakamoto N, Fukuda K, Watanuki K, Sakai D, Komano T, Scotting PJ, Yasugi S. Role for cGATA-5 in transcriptional regulation of the embryonic chicken pepsinogen gene by epithelial-mesenchymal interactions in the developing chicken stomach. Dev Biol. 2000; 223:103–113. doi:10.1006/dbio.2000.9731. [PubMed: 10864464]

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; 40:730–740. doi:10.1038/ng.152. [PubMed: 18488030]

Sala N, Munoz X, Travier N, Agudo A, Duell EJ, Moreno V, Overvad K, Tjonneland A, Boutron-Ruault MC, Clavel-Chapelon F, Canzian F, Kaaks R, Boeing H, Meidtner K, Trichopoulos A, Tsiotas K, Zylis D, Vineis P, Panico S, Palli D, Krogh V, Tumino R, Lund E, Bueno-de-Mesquita HB, Numans ME, Peeters PH, Quiros JR, Sanchez MJ, Navarro C, Ardanaz E, Dorronsoro M, Hall-mans G, Stenling R, Manjer J, Allen NE, Travis RC, Khaw KT, Jenab M, Offerhaus GJ, Riboli E, Gonzalez CA. Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: results from the EPIC-EURGAST study. Int J Cancer. 2012; 130:2417–2427. doi:10.1002/ijc.26243. [PubMed: 21681742]

Schneider BG, Mera R, Piazuelo MB, Bravo JC, Zabaleta J, Delgado AG, Bravo LE, Wilson KT, El-Rifai W, Peek RM Jr, Correa P. DNA methylation predicts progression of human gastric lesions. Cancer Epidemiol Biomark Prev. 2015; 24:1607–1613. doi:10.1158/1055-9965.epi-15-0388.

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; 43:1215–1218. doi:10.1038/ng.978. [PubMed: 22037551]

Song HR, Kim HN, Kweon SS, Choi JS, Shim HJ, Cho SH, Chung IJ, Park YK, Kim SH, Choi YD, Joo KW, Shin MH. Genetic variations in the PRKAA1 and ZBTB20 genes and gastric cancer susceptibility in a Korean population. Mol Carcinog. 2013; 52(Suppl 1):E155–E160. doi:10.1002/mc.22063. [PubMed: 23861218]

Soutto M, Belkhiri A, Piazuelo MB, Schneider BG, Peng D, Jiang A, Washington MK, Kokoye Y, Crowe SE, Zaika A, Correa P, Peek RM Jr, El-Rifai W. Loss of TFF1 is associated with activation of NF-kappaB-mediated inflammation and gastric neoplasia in mice and humans. J Clin Invest. 2011; 121:1753–1767. doi:10.1172/jci43922. [PubMed: 21490402]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Soutto M, Chen Z, Saleh MA, Katsha A, Zhu S, Zaika A, Belkhiri A, El-Rifai W. TFF1 activates p53 through down-regulation of miR-504 in gastric cancer. Oncotarget. 2014; 5:5663–5673. doi: 10.18632/oncotarget.2156. [PubMed: 25015107]

Soutto M, Chen Z, Katsha AM, Romero-Gallo J, Krishna US, Piazuelo MB, Washington MK, Peek RM Jr, Belkhiri A, El-Rifai WM. Trefoil factor 1 expression suppresses [i]Helicobacter pylori[/i]-induced inflammation in gastric carcinogenesis. Cancer. 2015 doi:10.1002/cncr.29644.

Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, Beck S, Wagner W, Laird PW, Jacobs IJ, Widschwendter M. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res. 2010; 20:440–446. doi:10.1101/gr.103606.109. [PubMed: 20219944]

The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014; 513:202–209. doi:10.1038/nature13480. [PubMed: 25079317]

Turner SD. qqman: an R package for visualizing GWAS results using Q–Q and manhattan plots. bioRxiv beta. 2014:1–2. doi:10.1101/005165.

Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. Genetic variation and population structure in native Americans. PLoS Genet. 2007; 3:e185. doi:10.1371/journal.pgen.0030185. [PubMed: 18039031]

Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet. 2008; 4:e1000037. doi:10.1371/journal.pgen.1000037. [PubMed: 18369456]

Wang X, Kang GH, Campan M, Weisenberger DJ, Long TI, Cozen W, Bernstein L, Wu AH, Siegmund KD, Shibata D, Laird PW. Epigenetic subgroups of esophageal and gastric adenocarcinoma with differential GATA5 DNA methylation associated with clinical and lifestyle factors. PLoS One. 2011; 6:e25985. doi:10.1371/journal.pone.0025985. [PubMed: 22028801]

Wang D, Yang PN, Chen J, Zhou XY, Liu QJ, Li HJ, Li CL. Promoter hypermethylation may be an important mechanism of the transcriptional inactivation of ARRDC3, GATA5, and ELP3 in invasive ductal breast carcinoma. Mol Cell Biochem. 2014; 396:67–77. doi:10.1007/s11010-014-2143-y. [PubMed: 25148870]

Watanabe Y, Kim HS, Castoro RJ, Chung W, Estecio MR, Kondo K, Guo Y, Ahmed SS, Toyota M, Itoh F, Suk KT, Cho MY, Shen L, Jelinek J, Issa JP. Sensitive and specific detection of early gastric cancer with DNA methylation analysis of gastric washes. Gastroenterology. 2009; 136:2149–2158. doi:10.1053/j.gastro.2009.02.085. [PubMed: 19375421]

Wen XZ, Akiyama Y, Pan KF, Liu ZJ, Lu ZM, Zhou J, Gu LK, Dong CX, Zhu BD, Ji JF, You WC, Deng DJ. Methylation of GATA-4 and GATA-5 and development of sporadic gastric carcinomas. World J Gastroenterol. 2010; 16:1201–1208. [PubMed: 20222162]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

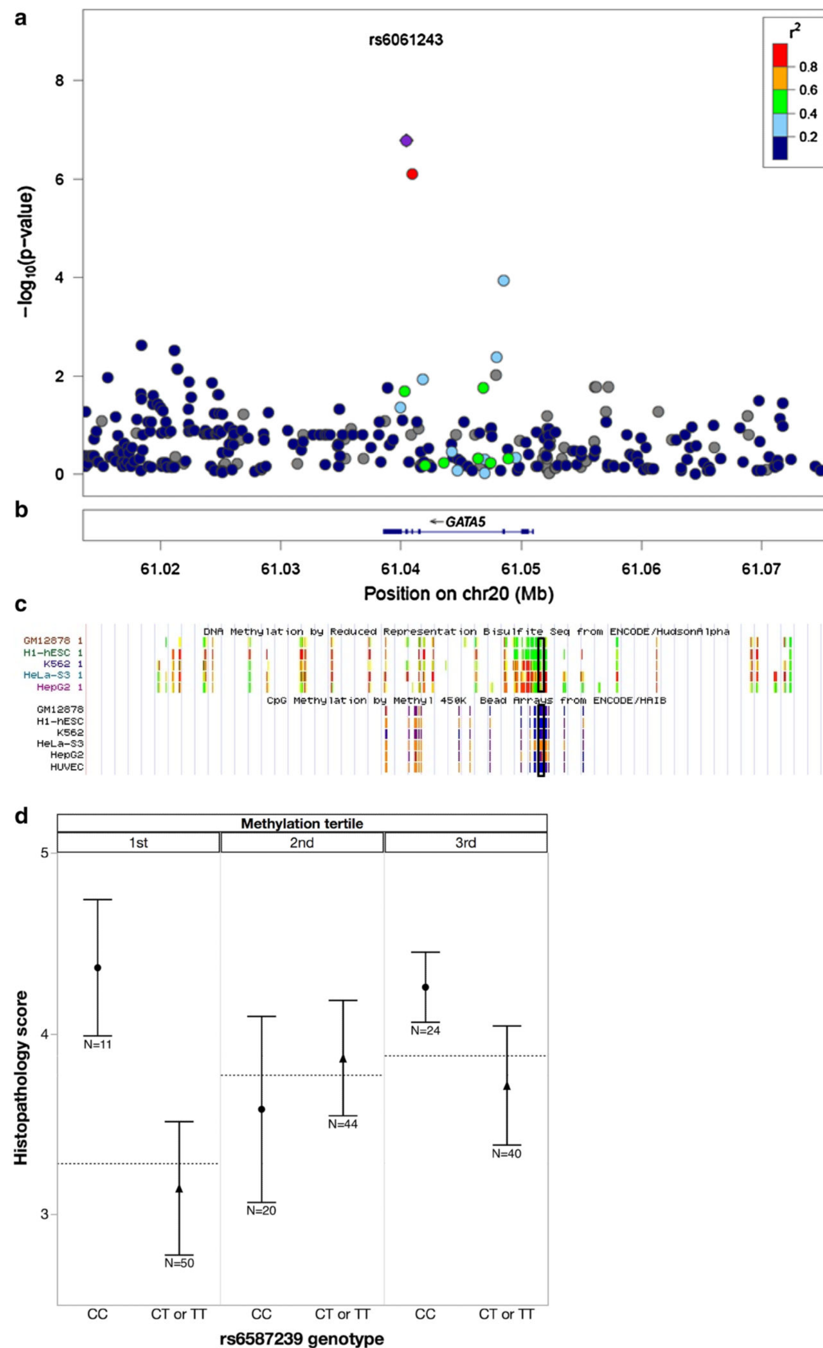

Fig. 1.

a From discovery cohort data, SNPs were imputed within a 50 kb region around *GATA5*, using a dominant model. The SNP with the most significant association, rs6061243, is shown as a *purple square*. Other *colors* represent linkage disequilibrium (measured with r^2) between imputed variants and the index SNP. **b** Annotation of DNA methylation in the region shown is from the UCSC Genome Browser (hg19). **c** Positions of the regions assayed for DNA methylation in the methylation group are located within the *black boxes*. **d** Gastric histopathology scores versus genotypes of rs6587239, in tertiles of *GATA5* promoter

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

methylation. *Dotted lines* indicate mean histopathology scores for each tertile (color figure online)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Characteristics of study participants

Variable	Discovery cohort *N* = 130	Replication cohort *N* = 159	Methylation group *N* = 189
Sex: # males[a]	86 (66 %)	100 (63 %)	132 (70 %)
Age[b]	48.66 ± 6.95	50.02 ± 7.31	49.53 ± 7.27
Age[c]	47 (39, 64)	49 (31, 67)	48 (31, 67)
Region of origin: # from mountains[a]	52 (40 %)	133 (84 %)	126 (67 %)
Histopathology score[b]	2.92 ± 0.92	3.41 ± 0.98	3.37 ± 0.97
GATA5 methylation: %	NA	NA	21.45 ± 8.37
cagA: # positive[a]	115 (89 %)	NA	NA
vacA: # s2 allele[a]	18 (14 %)	11 (7 %)	19 (10 %)

[a]Values are expressed as counts (percent total)

[b]Values are expressed as mean ± standard deviation

[c]Values are expressed as median (min, max)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Association of gastric histopathology scores in the discovery cohort

Variable	Coefficient	95 % confidence interval	*p* value for difference in histopathology score
Sex (male)[a]	0.18	(–0.16, 0.52)	0.291
Region of origin (mountain)[a]	0.30	(–0.023, 0.62)	0.068
Age[a]	–0.0004	(–0.024, 0.023)	0.970
cagA[b]	0.55	(0.033, 1.06)	0.037
vacA s2[c]	–0.76	(–1.21, –0.31)	0.001
European ancestry[a]	1.65	(0.54, 2.75)	0.004
African ancestry[a]	–0.71	(–1.19, –0.22)	0.005
Amerindian ancestry[a]	0.65	(0.013, 1.29)	0.046

[a]130 observations

[b]129 observations

[c]128 observations

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Single nucleotide polymorphisms associating with gastric histopathology scores in the discovery cohort

SNP	Chr	Position[a]	Minor allele	MAF	Model	*n*	*β*	*p* value	Gene
rs6061243[b]	20	61,040,453	G	0.446	Dom	126	–0.86	2.63×10^{-7}	*GATA5*
rs6587239[b]	20	61,040,951	T	0.438	Dom	127	–0.82	7.97×10^{-7}	*GATA5*
rs1046587	14	24,768,060	A	0.216	Dom	127	0.71	1.29×10^{-5}	*LTB4R*
rs158250	5	102,406,779	C	0.215	Add	127	–0.57	4.10×10^{-5}	near *PAM*
rs1046587	14	24,768,060	A	0.216	Add	127	0.51	6.51×10^{-5}	*LTB4R*
rs464720	5	102,406,887	T	0.227	Add	127	–0.52	8.30×10^{-5}	near *PAM*
rs464720	5	102,406,887	T	0.227	Dom	127	–0.61	8.54×10^{-5}	near *PAM*
rs1113665	6	109,579,675	G	0.396	Dom	127	0.62	9.64×10^{-5}	*LOC100996634*

Associations $<1 \times 10^{-4}$ are shown; analyses were adjusted for African and Amerindian ancestry and *H. pylori* virulence factors *cagA* and *vacA s*

MAF minor allele frequency

[a]Position in the hg19 assembly

[b]Passes Bonferroni multiple testing adjustment

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Association of gastric histopathology scores with additive and dominant models of *GATA5* SNPs in the Replication cohort

Independent variables	Minor allele	MAF	*n*	*β*	95 % CI	*p* value
Add rs6061243	G	0.460	149	–0.24	(–0.45, –0.024)	0.029
Region of origin				0.11	(–0.29, 0.50)	0.6
vacA s2				–0.63	(–1.082, –0.18)	0.007
Add rs6587239	T	0.457	150	–0.26	(–0.47, –0.039)	0.021
Region of origin				0.20	(–0.20, 0.59)	0.328
vacA s2				–0.62	(–1.08, –0.16)	0.009
Dom rs6061243	G	0.460	149	–0.30	(–0.62, 0.032)	0.076
Region of origin				0.12	(–0.28, 0.52)	0.564
vacA s2				–0.63	(–1.08, –0.17)	0.007
Dom rs6587239	T	0.457	150	–0.33	(–0.66, 0.00)	0.050
Region of origin				0.22	(–0.18, 0.62)	0.276
vacA s2				–0.62	(–1.08, –0.16)	0.009

Analyses were adjusted for region of origin and the *H. pylori* virulence factor *vacA* s

Add additive model, *Dom* dominant model, *MAF* minor allele frequency

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 5

Variables affecting gastric histopathology scores in the methylation group

Independent variable(s)	β	95% CI	p value
Age	0.026	(0.006, 0.045)	0.009
Region of origin	0.40	(0.10, 0.69)	0.008
Dominant rs6587239	−0.45	(−0.76, −0.15)	0.004
GATA5 methylation	0.028	(0.012, 0.045)	0.001
Sex (male)	−0.28	(−0.58, 0.023)	0.070
vacA s	−0.59	(−0.96, −0.22)	0.002
Age	0.022	(0.005, 0.040)	0.013
Region of origin	0.18	(0.093, 0.46)	0.194
Dominant rs6587239	−1.20	(−2.07, −0.33)	0.007
GATA methylation	−0.012	(−0.044, 0.020)	0.463
vacA s2	−0.73	(−1.08, −0.39)	4.28×10^{-5}
Interaction of Dom rs6587239 * *GATA5* methylation	0.040	(0.003, 0.076)	0.033

n = 189

The top half of the Table shows independent covariates in univariate tests.

The lower half shows covariates in a single regression model.