World Journal of

Gastrointestinal Oncology

World Journal of Gastrointestinal Oncology

Contents

Contents

LETTER TO THE EDITOR

Contents

ABOUT COVER

Editorial Board Member of *World Journal of Gastrointestinal Oncology*, Luigi Marano, MD, PhD, Associate Professor, Department of Medicine, Surgery, and Neurosciences, University of Siena, Siena 53100, Italy. luigi.marano@unisi.it

AIMS AND SCOPE

The primary aim of *World Journal of Gastrointestinal Oncology* (*WJGO*, *World J Gastrointest Oncol*) is to provide scholars and readers from various fields of gastrointestinal oncology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

 WJGO mainly publishes articles reporting research results and findings obtained in the field of gastrointestinal oncology and covering a wide range of topics including liver cell adenoma, gastric neoplasms, appendiceal neoplasms, biliary tract neoplasms, hepatocellular carcinoma, pancreatic carcinoma, cecal neoplasms, colonic neoplasms, colorectal neoplasms, duodenal neoplasms, esophageal neoplasms, gallbladder neoplasms, *etc*.

INDEXING/ABSTRACTING

The *WJGO* is now abstracted and indexed in PubMed, PubMed Central, Science Citation Index Expanded (SCIE, also known as SciSearch®), Journal Citation Reports/Science Edition, Scopus, Reference Citation Analysis, China National Knowledge Infrastructure, China Science and Technology Journal Database, and Superstar Journals Database. The 2022 edition of Journal Citation Reports® cites the 2021 impact factor (IF) for *WJGO* as 3.404; IF without journal self cites: 3.357; 5-year IF: 3.250; Journal Citation Indicator: 0.53; Ranking: 162 among 245 journals in oncology; Quartile category: Q3; Ranking: 59 among 93 journals in gastroenterology and hepatology; and Quartile category: Q3. The *WJGO*'s CiteScore for 2021 is 3.6 and Scopus CiteScore rank 2021: Gastroenterology is 72/149; Oncology is 203/360.

RESPONSIBLE EDITORS FOR THIS ISSUE

EVIDENCE-BASED MEDICINE

Prediction of gastric cancer risk by a polygenic risk score of *Helicobacter pylori*

Xiao-Yu Wang, Li-Li Wang, Shu-Zhen Liang, Chao Yang, Lin Xu, Meng-Chao Yu, Yi-Xuan Wang, Quan-Jiang Dong

Xiao-Yu Wang, Li-Li Wang, Shu-Zhen Liang, Lin Xu, Meng-Chao Yu, Yi-Xuan Wang, Quan-Jiang Dong, Central Laboratories and Department of Gastroenterology, Qingdao Municipal Hospital, Qingdao University, Qingdao 266071, Shandong Province, China

Chao Yang, The Center for Microbes, Development and Health, CAS Key Laboratory of Molecular Virology and Immunology, Institute Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200000, China

Corresponding author: Quan-Jiang Dong, MD, PhD, Chief Doctor, Professor, Central Laboratories and Department of Gastroenterology, Qingdao Municipal Hospital, Qingdao University, No. 5 Donghai Middle Road, Qingdao 266071, Shandong Province, China. allyking114@126.com

Abstract

BACKGROUND
Genetic variants of *Helicobacter pylori* (*H. pylori*) are involved in gastric cancer occurrence. Single nucleotide polymorphisms (SNPs) of *H. pylori* that are associated with gastric cancer have been reported. The combined effect of *H. pylori* SNPs on the risk of gastric cancer remains unclear.

AIM
To assess the performance of a polygenic risk score (PRS) based on *H. pylori* SNPs in predicting the risk of gastric cancer.

METHODS
A total of 15 gastric cancer-associated *H. pylori* SNPs were selected. The associations between these SNPs and gastric cancer were further validated in 1022 global strains with publicly available genome sequences. The PRS model was established based on the validated SNPs. The performance of the PRS for predicting the risk of gastric cancer was assessed in global strains using quintiles and random forest (RF) methods. The variation in the performance of the PRS among different populations of *H. pylori* was further examined.

RESULTS
Analyses of the association between selected SNPs and gastric cancer in the global dataset revealed that the risk allele frequencies of six SNPs were significantly higher in gastric cancer cases than non-gastric cancer cases. The PRS model constructed subsequently with these validated SNPs produced significantly

higher scores in gastric cancer. The odds ratio (OR) value for gastric cancer gradually increased from the first to the fifth quintile of PRS, with the fifth quintile having an OR value as high as 9.76 (95% confidence interval: 5.84-16.29). The results of RF analyses indicated that the area under the curve (AUC) value for classifying gastric cancer and non-gastric cancer was 0.75, suggesting that the PRS based on *H. pylori* SNPs was capable of predicting the risk of gastric cancer. Assessing the performance of the PRS among different *H. pylori* populations demonstrated that it had good predictive power for cancer risk for hpEurope strains, with an AUC value of 0.78.

CONCLUSION
The PRS model based on *H. pylori* SNPs had a good performance for assessment of gastric cancer risk. It would be useful in the prediction of final consequences of the *H. pylori* infection and beneficial for the management of the infection in clinical settings.

Key Words: Polygenic risk scores; *Helicobacter pylori*; Gastric cancer; Single nucleotide polymorphism

Core Tip: Prediction of cancer risk is of importance in the clinical management of populations with a high risk of gastric cancer. This study constructed a polygenic risk score (PRS) model based on *Helicobacter pylori* (*H. pylori*) single nucleotide polymorphisms (SNPs) to predict the risk of gastric cancer. Associations between previously reported *H. pylori* SNPs and gastric cancer were validated in global strains. A PRS model constructed with validated SNPs had a high predictive power for gastric cancer at a global level and for individuals infected with hpEurope strains. It has potential for clinical use in the management of the *H. pylori* infection.

INTRODUCTION

Helicobacter pylori (*H. pylori*) infection affects more than half of the world's population[1,2]. The outcomes of *H. pylori* infection vary among individuals. The consequences of most infections are benign. However, a minority of infected individuals may eventually develop gastric cancer[3,4]. Predicting the outcomes of *H. pylori* infection is a major concern in the management of the infection. Substantial genetic variation has been found in the pathogen. Mutations cause increased virulence in certain strains, enhancing their carcinogenic potential[5,6]. It has been demonstrated that typing *H. pylori* strains based on the genetic variations of virulent genes has the potential to predict the risk of gastric cancer[7,8].

Two studies have been recently conducted to investigate the association between *H. pylori* genomic variations and gastric cancer within the hpEurope and hpEastAsia populations, respectively[9,10]. The first study contained 173 hpEurope strains and found 11 cancer risk-associated variants, including gene loss variants and single nucleotide polymorphisms (SNPs). Risk scores calculated based on the status of the *cag*11, *cag*12 and *cag*20 genes were increased during the progression from inflammation to gastric cancer. The other study identified 11 SNPs and three DNA motifs associated with gastric cancer through examination of 240 hpEastAsia strains. It is unclear whether the association between these variations and gastric cancer exists for all *H. pylori* strains. However, the findings from these studies suggest that SNPs from the *H. pylori* genome have the potential to predict the risk of gastric cancer.

To explore the combined effect of multiple SNPs on disease susceptibility, the polygenic risk score (PRS) model has been developed[11]. A PRS is calculated as a sum of the effects of multiple SNPs on disease. PRS models composed of SNPs from the human genome have been successfully used to predict the risk of cancers such as gastric cancer, colorectal cancer, and breast cancer[12-15]. Few studies, however, have been conducted to explore the capacities of PRS model constructed with SNPs from bacterial genomes in predicting the risk of cancer. Our study aimed to construct a PRS model based on validated risk alleles of *H. pylori* to predict the risk of gastric cancer.

MATERIALS AND METHODS

Strains and SNP selection

A total of 2022 *H. pylori* genome sequences deposited in GenBank at the National Center for Biotechnology Information by December 8, 2021 (https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/169/), and the figshare website (https://figshare.com/s/2174da1fa20ae71c71e0)[10] were downloaded for further analyses. Of them, 1187 *H. pylori* strains had relevant clinical information of patients. Subsequently, duplicate strains and strains isolated from peptic ulcer disease, mucosa-associated lymphoma or stromal tumors were excluded from further analyses. This led to a final dataset of 1022 global strains included in the study. They were divided into gastric cancer (*n* = 253) and non-gastric cancer (*n* = 769) groups. Patients in the latter group were diagnosed with functional dyspepsia (*n* = 46), or chronic gastritis with or without intestinal metaplasia (*n* = 143 and *n* = 580, respectively). A total of 15 *H. pylori* SNPs or genetic variants from the two previous genome-wide association studies (GWASs) were selected for further analyses (Figure 1, Table 1)[9,10]. We cited high-quality articles in *Reference Citation Analysis* (https://www.referencecitationanalysis.com). The selection criteria were as follows: (1) The length of the variants was no longer than five contiguous nucleotides; and (2) The SNP selected was located in a protein-coding region.

Construction of the neighbour-joining tree

Based on the 1022 *H. pylori* genomes, the SNPs in the core genome (present in > 99% isolates) were identified by aligning the assembled genomes against the reference genome (26695-1MET, accession number: CP010436.1) using MUMmer as previously described[16,17]. A neighbour-joining tree was then constructed based on the sequences of concatenated SNPs using TreeBeST software (http://treesoft.sourceforge.net/treebest.shtml) with default parameters.

Statistical analyses

The chi-square test was used to test the difference in the prevalence of risk alleles in strains isolated from gastric cancer and non-gastric cancer. Student's *t* test was used to compare the PRS values between the gastric cancer and non-cancer groups. These tests were performed using SPSS 18.0 software. Odds ratios (ORs) and 95% confidence intervals (CIs) of the selected SNPs were calculated using logistic regression analysis in R (version 3.6.3).

A PRS was created for each strain using the following equation: PRS = $\beta 1 + \beta 2 + \ldots \beta k \ldots + \beta n$. Briefly, in this equation, βk is the value obtained from the regression analysis of the risk allele and disease, and n is the total number of SNPs included in the PRS[18]. Logistic regression analysis was performed to evaluate the association between PRS and gastric cancer risk and by quintiles of the PRS risk distribution, standardized by the controls, and using the 3rd quintile, 40%-60%, as the reference[18].

Random forest algorithm

A random forest (RF) model was built using the AUC-RF algorithm[19]. The input variables were the scores of each of the validated SNPs. A 20-times repeated 10-fold cross-validation of the RF model was performed. The performance of the RF model was demonstrated by receiver operating characteristic curve analysis[20].

RESULTS

Validation of SNPs in the global dataset

Previous studies have identified two sets of *H. pylori* SNPs that are associated with gastric cancer[9,10]. The association between these SNPs and gastric cancer has been verified only in strains from the hpEurope or hpEastAsia populations, respectively. We selected 15 SNPs to validate the association between selected SNPs and gastric cancer in global strains (Table 1). The risk alleles were defined as those with a higher prevalence in strains from gastric cancer. Statistical analyses revealed that the risk alleles of six SNPs showed a significant increase in prevalence in the gastric cancer group compared with the non-gastric cancer group. These SNPs, validated in the global dataset, were used for subsequent analyses.

Establishment of the six-SNP PRS model

To construct a PRS model for predicting the risk of gastric cancer, the logOR values of each validated SNP were calculated (Table 1). A PRS model was subsequently constructed with the sum of the logOR values of six validated SNPs. The mean PRS value was 8.64 ± 1.71 and 6.99 ± 1.27 in the gastric cancer and non-gastric cancer groups, respectively. The PRS value in the gastric cancer group was significantly higher (*P* = 5.6E-36).

Table 1 Selected and validated single nucleotide polymorphisms associated with gastric cancer in the global dataset

SNP	Corresponding locus in the strain 26695	Gene name	Description	Position in gene	Position in chromosome	Risk allele	Amino acid change	Prevalence of risk allele	*P* value	LogOR -value
1[1]	HP0082[1]	*tlpC*[1]	Chemotaxis sensor[1]	163[1]	88029[1]	A[1]	K217E,Q[1]	16.2%[1]	2.88E-15[1]	1.29[1]
2[1]	HP0130[1]	*triH*[1]	BIR, Dps/NapA, RAD21 similarity[1]	345[1]	140797[1]	C[1]	Synonymous[1]	16.2%[1]	2.88E-15[1]	2.26[1]
3[1]	HP0231[1]	*dsbG/K*[1]	Thiol:disulfide interchange protein[1]	433[1]	241625[1]	A[1]	T145A[1]	23.2%[1]	5.34E-15[1]	1.24[1]
4	HP0468		Unknown	729	489762	A	Synonymous	34.6%	0.110	
5	HP0468		Unknown	705-708	489783-489786	CGCC	A236T	1.2%	0.313	
6	HP0709		Adenosyl-chloride synthase	145	762953	A	N49D	14.6%	0.08	
7	HP0709		Adenosyl-chloride synthase	159	762967	A	Synonymous	90.0%	0.274	
8[1]	HP0747[1]	*trmB*[1]	tRNA ([guanine-N(7)-]-methyltrans-ferase[1]	(934-937)[1]	(803467-803470)[1]	GGAA[1]	G312K,G,R+T313A,T,S[1]	38.7%[1]	1.17E-14[1]	1.20[1]
9	HP0797	*hpaA*	Neuraminyllactose-binding hemagglutinin	334	854406	T	S112A	26.8%	0.567	
10[1]	HP0797[1]	*hpaA*[1]	Neuraminyllactose-binding hemagglutinin[1]	325[1]	854415[1]	C[1]	L109F[1]	40.1%[1]	1.12E-14[1]	2.16[1]
11	HP0807	*fecA*-2	Iron importer in outer membrane	2010	861345	C	Synonymous	96.7%	0.158	
12[1]	HP1055[1]		Outer membrane protein[1]	798[1]	1117402[1]	A[1]	Synonymous[1]	34.6%[1]	2.93E-14[1]	2.58[1]
13	HP1250	*csd5*	Cell shape determinant	370	1325727	A	N116H,D	65.8%	1	
14	HP1440	*isp*	Inactive Ser protease	533	1513405	G	G173E	92.0%	0.505	
15	HP1467	*ompA101*	Outer membrane protein of OmpA family	53	1538114	T	I18T	98.4%	0.246	

[1]The single nucleotide polymorphisms validated in this work.

Risk allele: Allele associated with gastric cancer. SNP: Single nucleotide polymorphism; OR: Odds ratio.

Evaluation of the association between PRS and gastric cancer risk

To evaluate the performance of the 6-SNP PRS model for predicting the risk of gastric cancer, the PRS values for each of the selected 1022 strains were grouped according to the quintile method. With the third quintile as the reference, the estimated OR value gradually increased from the first quintile (< 20%) to the fifth quintile (> 80%) (Figure 2, Table 2). The fifth quintile had an OR value as high as 9.76 (95%CI: 5.84-16.29).

To further confirm the combined effect of the validated SNPs for prediction of gastric cancer risk, an RF model was constructed with logOR values from each SNP as input. The classification potentials of the combined logOR values of validated SNPs were then analysed. The importance of each SNP is shown in Figure 3. The AUC value was 0.75 (DeLong 95%CI: 0.71-0.78), suggesting a good classifying capacity of the combined SNPs.

Performance of risk prediction by PRS for different H. pylori populations

Considering the remarkable genomic variations among strains from different *H. pylori* populations, the performance of PRS for predicting the risk of gastric cancer was subsequently assessed in different *H. pylori* populations. The results of the phylogenetic analyses divided the 1022 global strains into five groups, namely, the hpEastAsia, hpAsia2, hpEurope, America-related and Africa-related populations (Figure 4). Due to the small number of gastric cancer cases (2 cases in hpAsia2 and no cases in Africa-related populations), hpAsia2 and Africa-related populations were excluded from subsequent analyses.

Table 2 Associations between polygenic risk score and gastric cancer risk

Quintile	Non-GC	GC	OR (95%CI)	*P* value
1	315 (41.0%)	44 (17.3%)	0.79 (0.46-1.34)	0.405
2	98 (12.7%)	7 (2.8%)	0.40 (0.17-0.97)	0.052
3	141 (18.3%)	25 (9.9%)	1	-
4	141 (18.3%)	49 (19.4%)	1.96 (1.15-3.35)	0.013
5	74 (9.6%)	128 (50.6%)	9.76 (5.84-16.29)	1.25E-14

CI: Confidence interval; OR: Odds ratio; GC: Gastric cancer
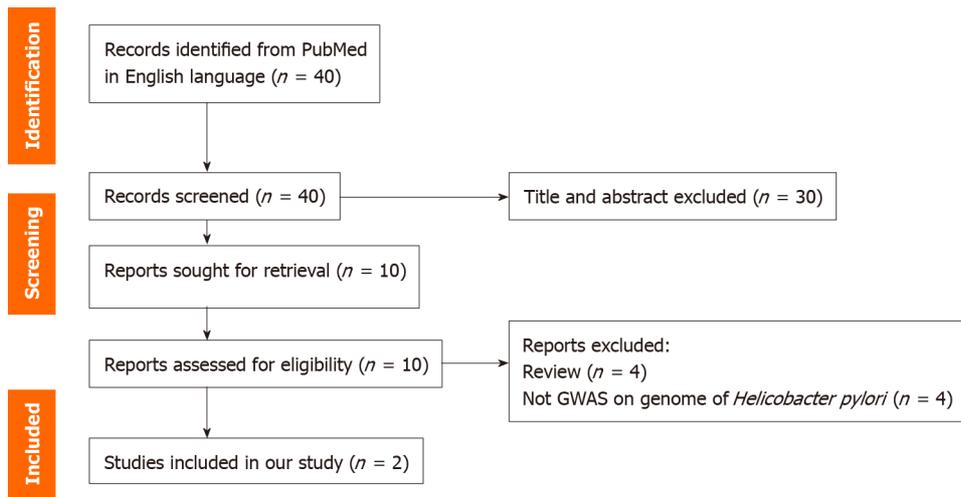
Figure 1 PRISMA flow diagram. GWAS: Genome-wide association studies.

In analysing the performance of the established PRS model in different populations, the PRS value was higher in the gastric cancer group for all populations. Statistical analyses revealed a significant difference in PRS between the gastric cancer and non-gastric cancer groups in the hpEastAsia, hpEurope and America-related populations (Figure 5).

To further verify the combined effects of these SNPs for prediction of gastric cancer risk for different *H. pylori* populations, a RF classification model was built. The results of RF model analyses demonstrated that the AUC value was highest (0.78, DeLong 95%CI: 0.70-0.85) in the hpEurope population, suggesting a good ability of the combined SNPs to predict the risk of gastric cancer (Figure 6). However, the performance of the combined SNPs for risk prediction in other *H. pylori* populations was poor (Figure 6).

DISCUSSION

In this study, we constructed a PRS model based on validated *H. pylori* SNPs to predict the risk of gastric cancer. To our knowledge, our study is the first to evaluate a PRS model for cancer risk prediction constructed with genomic variants of *H. pylori*. *H. pylori* shows substantial genetic variations, resulting in remarkable interstrain differences in carcinogenetic potential[5,21]. The presence/absence or large sequence variation of virulence genes and *H. pylori* SNPs have been shown to promote gastric carcinogenesis. Few studies have been conducted to assess the predictive power of these cancer-related genetic variations for gastric cancer[9,10]. Moreover, the combined effect of multiple variations on the predictive power for cancer risk has not been explored. Findings from this study demonstrate that a PRS model combining six *H. pylori* SNPs had a moderate capacity for prediction of gastric cancer risk. This is similar to the findings in studies on PRS model constructed with cancer-associated SNPs from the human genome[14,15].

To assess the combined effects of SNPs on gastric cancer risk prediction, we first selected 15 cancer-associated *H. pylori* SNPs from two previous GWAS studies. Their association has been validated in strains from specific geographical regions but not in a global strain collection. Our results demonstrated
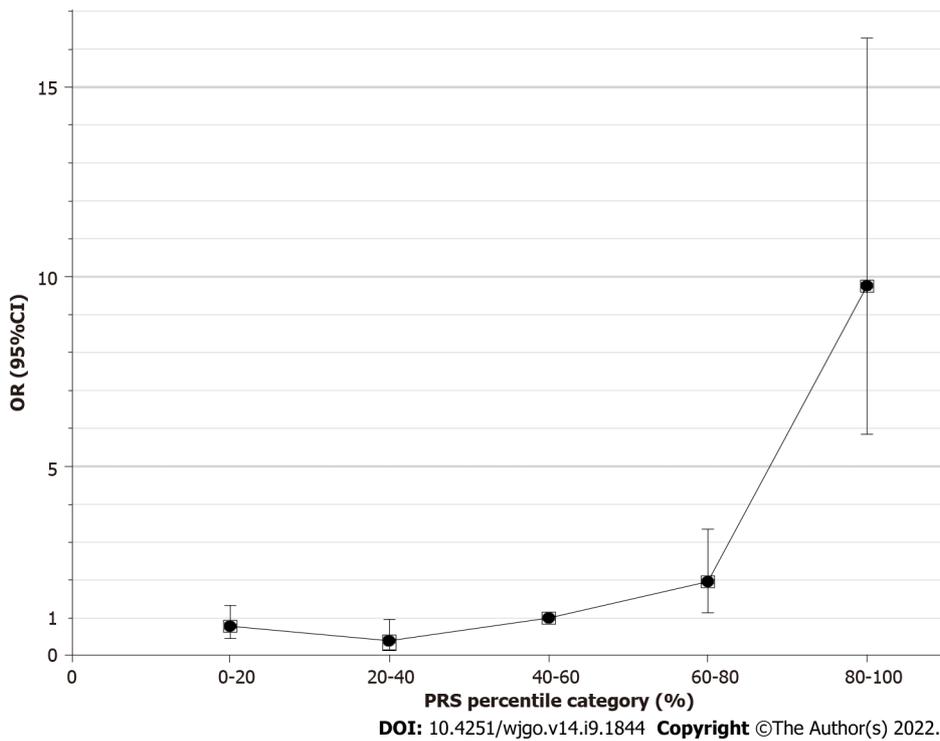
Figure 2 Odds ratio and 95% confidence intervals (error bars) for percentiles of polygenic risk relative to the middle quintile. OR: Odds ratio; CI: Confidence interval; PRS: Polygenic risk score.
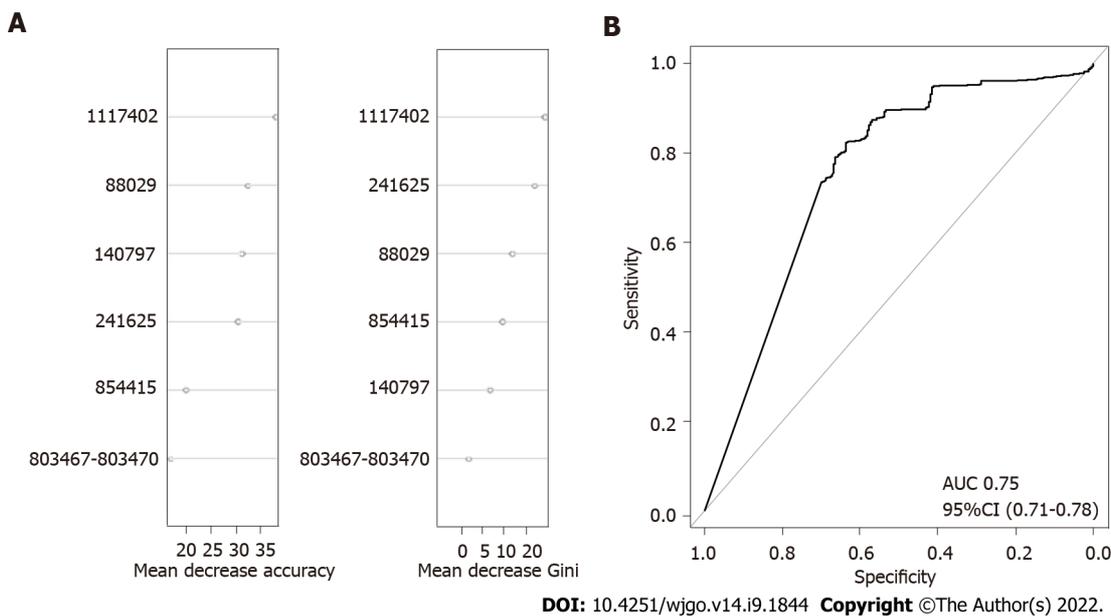
Figure 3 The importance of each validated single nucleotide polymorphisms and receiver operating characteristic curve for the polygenic risk score model on global strains. A: The median decrease in accuracy and median decrease in Gini coefficient of validated single nucleotide polymorphisms; B: Receiver operating characteristic curve of global strains. AUC: Area under the curve; CI: Confidence interval.

that only six of the SNPs showed a close association with gastric cancer in the global dataset. The SNPs at 88029, 241625, 803467 and 854415 in the reference strain 26695 caused nonsynonymous changes in the corresponding amino acid sequence, whereas the SNPs at 140797 and 1117402 in the reference strain 26695 produced synonymous variations. The *hpaA* gene, harbouring the SNP at 854415, encodes an adhesion gene of *H. pylori*[22]. This gene is essential for colonization and is associated with the occurrence of gastric cancer[23-25]. The SNP at 88029 was located on the *tlpC* gene. *TlpC* encodes a chemoreceptor that affects the chemotaxis of strains in the mouse gastric environment. It is associated with the induction of mucosal inflammation of the stomach[26,27]. The SNP at 241625 was located in
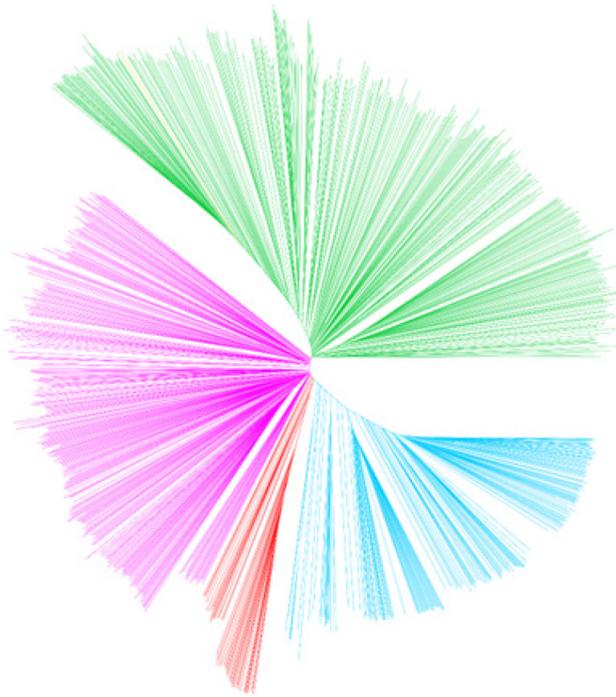
Figure 4 Neighbour-joining tree constructed from concatenated single nucleotide polymorphism sequences. Blue: hpEastAsia; Red: hpAsia2; Pink: hpEurope; Green: America-related populations; Yellow: Africa-related populations (strain hp_151, which was isolated in Morocco, was excluded from further analyses).

Disease	NC	GC	NC	GC	NC	GC
Number of strains	85	133	274	44	365	74
Groups	hpEastAsia		hpEurope		America-related populations	
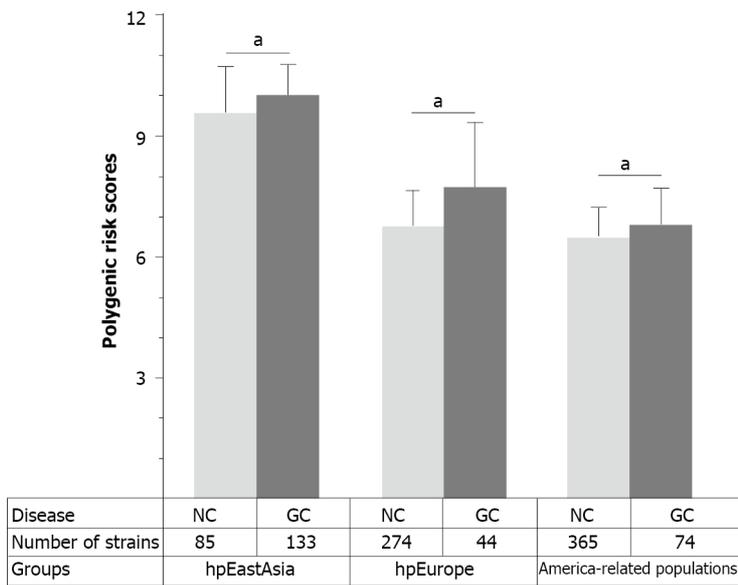
Figure 5 Distribution of polygenic risk score for various groups. NC: Non-gastric cancer; GC: Gastric cancer. $^{a}P < 0.05$.

dsbG/K, which has protein disulfide isomerase activity. *DsbG/K* interacts with a virulence-related factor *in vitro*[28,29]. *In vitro* studies have shown that a lack of *dsbG/K* may cause the loss of T4SS function and inhibit *VacA* secretion, which are considered the main pathogenic factors in *H. pylori*[30].

In this study, we constructed a PRS model with six SNPs validated in a global dataset. Assessments of the performance of the PRS model demonstrated that the PRS value was significantly higher in the gastric cancer group than in the non-gastric cancer group. A significant increase in the risk of gastric cancer was found across the quintiles of the PRS. These findings demonstrate that the six-SNPs PRS model is capable of predicting the risk of gastric cancer. In support of this finding, RF analyses demonstrated that the combination of the six SNPs has a high predictive power for gastric cancer, with an AUC value of 0.75. In a recent report, a PRS model constructed with SNPs from the human genome
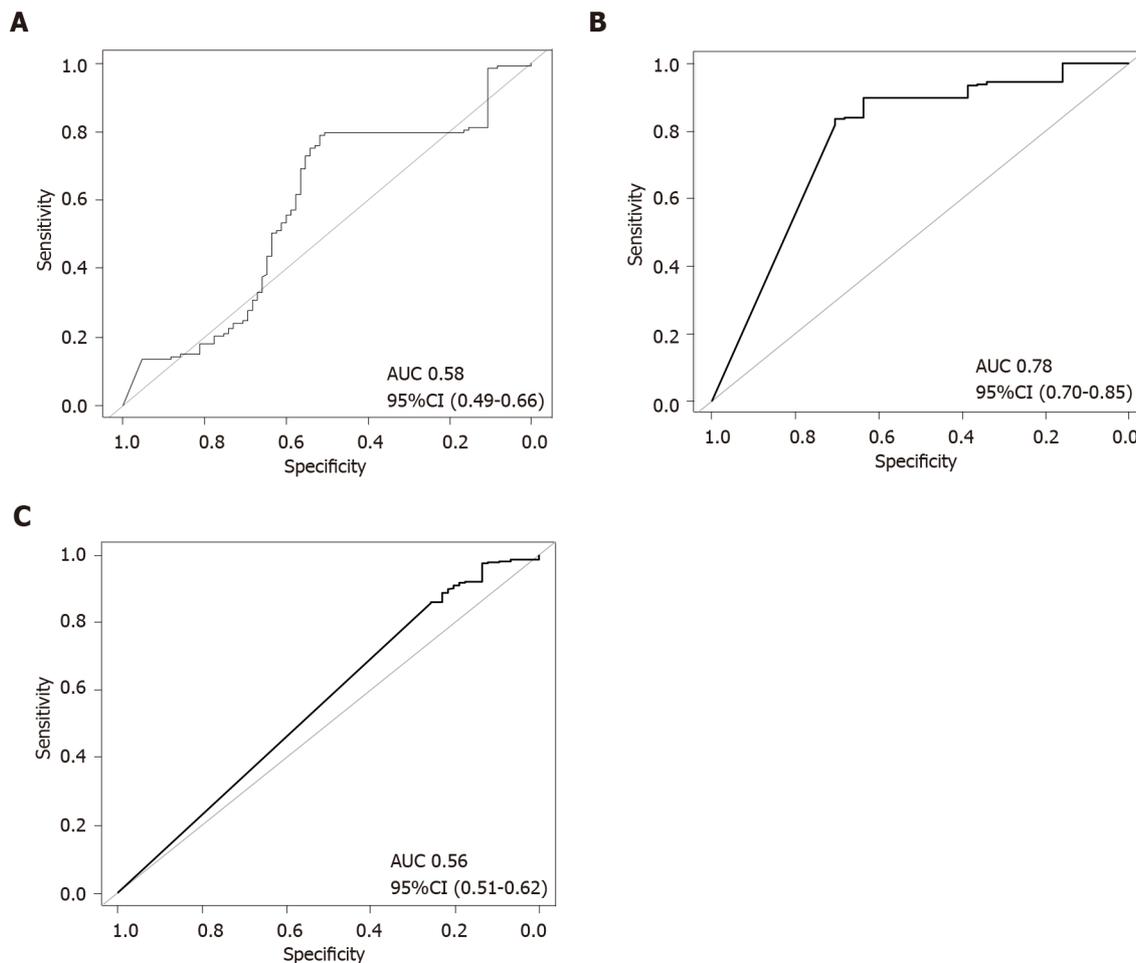
Figure 6 Receiver operating characteristic curve of the polygenic risk score model for various groups. A: Receiver operating characteristic curve (ROC) of hpEastAsia; B: ROC curve of hpEurope; C: ROC curve of America-related populations. AUC: Area under the curve; CI: Confidence interval.

showed unsatisfactory power in classifying gastric cancer from healthy controls, with an AUC value of 0.56[31]. It has been shown that a PRS model derived from 112 SNPs in the human genome and lifestyle factors possesses good predictive capacity for gastric cancer risk[32]. For individuals infected with *H. pylori*, assessment of their gastric cancer risk is of great concern in the clinical settings. Previous reports have demonstrated that certain genetic variants are associated with increased gastric cancer risk[9,10]. Our study, for the first time, demonstrated the combined effect of *H. pylori* genomic variations in the assessment of cancer risk. The PRS model derived from *H. pylori* SNPs would have a high capacity in predicting gastric cancer risk for patients infected with the pathogen. This will benefit the clinical management of the prognosis of the *H. pylori* infection. It is well known that age, gender and lifestyle factors, including alcohol consuming, smoking, diet habits and economic status, are closely associated with gastric cancer[33-35]. In the future, a PRS model constructed with *H. pylori* SNPs and those gastric cancer associated risk factors in this study would have substantially increased power in predicting the risk of gastric cancer. The *H. pylori* genome shows great variations between strains[36,37]. Genetic information differs greatly among *H. pylori* populations, and their carcinogenic potential is also different [5,21]. We thus evaluated the performance of the PRS model across *H. pylori* populations. Our results demonstrated a good predictive power of PRS for hpEurope strains.

A limitation of this study is that the performance of the PRS model was not assessed in hpAsia2 and Africa-related *H. pylori* populations because the number of strains with clinical information available was insufficient. Moreover, we could not consider age, gender, nutrition and other risk factors in the construction of the PRS model, as information on all of these risk factors was not consistently available across databases. A comprehensive risk model enclosing other risk factors of gastric cancer is indicated in future studies. Further *in vitro* and *in vivo* exploration of the roles of the combination of *H. pylori* SNPs identified in this study in gastric cancer would be much helpful in supporting our findings.

CONCLUSION

In summary, we constructed a PRS model based on *H. pylori* SNPs, which showed great potential in the prediction of gastric cancer risk globally, especially for individuals infected with hpEurope strains. Findings from this study demonstrated that the PRS model constructed from bacteria genomic variations, in addition to the PRS model established with human SNPs, can be of great value for disease risk prediction. In clinical practice, it is usually difficult to assess gastric cancer risk in patients infected with *H. pylori*. The model constructed in this study would be beneficial for solving this issue.

ARTICLE HIGHLIGHTS

Research background

Multiple single nucleotide polymorphisms (SNPs) of *Helicobacter pylori* (*H. pylori*) associated with gastric cancer have been identified through bacterial genome-wide association studies. Polygenic risk score (PRS) calculated as a sum of effect of SNPs provides a tool for assessing genetic impact on diseases.

Research motivation

Predicting risk of gastric cancer is a major concern in the management of the *H. pylori* infection.

Research objectives

This study constructed a PRS model based on *H. pylori* SNPs to predict the risk of gastric cancer.

Research methods

Associations between previously reported *H. pylori* SNPs and gastric cancer were validated in global strains. The PRS model based on the validated SNPs was evaluated by quintiles and random forest (RF) methods.

Research results

A PRS model was constructed with six validated SNPs. Quintiles and RF methods demonstrated the combination of six SNPs has a high predictive power for gastric cancer.

Research conclusions

PRS model constructed from bacterial genomic variations can be of great value for gastric cancer risk prediction.

Research perspectives

Comprehensive risk models including personal and genomic information need to be established in future studies.

FOOTNOTES

Lin Xu 0000-0003-3098-8251; Quan-Jiang Dong 0000-0002-5226-7853.

REFERENCES

1   **Everhart JE**. Recent developments in the epidemiology of Helicobacter pylori. *Gastroenterol Clin North Am* 2000; **29**: 559-578 [PMID: 11030073 DOI: 10.1016/s0889-8553(05)70130-8]

2   **Schulz C**, Schütte K, Mayerle J, Malfertheiner P. The role of the gastric bacterial microbiome in gastric cancer: *Helicobacter pylori* and beyond. *Therap Adv Gastroenterol* 2019; **12**: 1756284819894062 [PMID: 31897087 DOI: 10.1177/1756284819894062]

3   **González CA**, Megraud F, Buissonniere A, Lujan Barroso L, Agudo A, Duell EJ, Boutron-Ruault MC, Clavel-Chapelon F, Palli D, Krogh V, Mattiello A, Tumino R, Sacerdote C, Quirós JR, Sanchez-Cantalejo E, Navarro C, Barricarte A, Dorronsoro M, Khaw KT, Wareham N, Allen NE, Tsilidis KK, Bas Bueno-de-Mesquita H, Jeurnink SM, Numans ME, Peeters PHM, Lagiou P, Valanou E, Trichopoulou A, Kaaks R, Lukanova-McGregor A, Bergman MM, Boeing H, Manjer J, Lindkvist B, Stenling R, Hallmans G, Mortensen LM, Overvad K, Olsen A, Tjonneland A, Bakken K, Dumeaux V, Lund E, Jenab M, Romieu I, Michaud D, Mouw T, Carneiro F, Fenge C, Riboli E. Helicobacter pylori infection assessed by ELISA and by immunoblot and noncardia gastric cancer risk in a prospective study: the Eurgast-EPIC project. *Ann Oncol* 2012; **23**: 1320-1324 [PMID: 21917738 DOI: 10.1093/annonc/mdr384]

4   **Kuipers EJ**, Thijs JC, Festen HP. The prevalence of Helicobacter pylori in peptic ulcer disease. *Aliment Pharmacol Ther* 1995; **9** Suppl 2: 59-69 [PMID: 8547530]

5   **Dong QJ**, Zhan SH, Wang LL, Xin YN, Jiang M, Xuan SY. Relatedness of Helicobacter pylori populations to gastric carcinogenesis. *World J Gastroenterol* 2012; **18**: 6571-6576 [PMID: 23236231 DOI: 10.3748/wjg.v18.i45.6571]

6   **Shiota S**, Matsunari O, Watada M, Yamaoka Y. Virulence factors or ancestral origin of Helicobacter pylori: which is a better predictor of gastric cancer risk? *Gut* 2012; **61**: 469-470 [PMID: 21610271 DOI: 10.1136/gutjnl-2011-300317]

7   **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 Helicobacter pylori coevolution shapes the risk of gastric disease. *Proc Natl Acad Sci U S A* 2014; **111**: 1455-1460 [PMID: 24474772 DOI: 10.1073/pnas.1318093111]

8   **Bakhti SZ**, Latifi-Navid S, Safaralizadeh R. Helicobacter pylori-related risk predictors of gastric cancer: The latest models, challenges, and future prospects. *Cancer Med* 2020; **9**: 4808-4822 [PMID: 32363738 DOI: 10.1002/cam4.3068]

9   **Berthenet E**, Yahara K, Thorell K, Pascoe B, Meric G, Mikhail JM, Engstrand L, Enroth H, Burette A, Megraud F, Varon C, Atherton JC, Smith S, Wilkinson TS, Hitchings MD, Falush D, Sheppard SK. A GWAS on Helicobacter pylori strains points to genetic variants associated with gastric cancer risk. *BMC Biol* 2018; **16**: 84 [PMID: 30071832 DOI: 10.1186/s12915-018-0550-3]

10   **Tuan VP**, Yahara K, Dung HDQ, Binh TT, Huu Tung P, Tri TD, Thuan NPM, Khien VV, Trang TTH, Phuc BH, Tshibangu-Kabamba E, Matsumoto T, Akada J, Suzuki R, Okimoto T, Kodama M, Murakami K, Yano H, Fukuyo M, Takahashi N, Kato M, Nishiumi S, Azuma T, Ogura Y, Hayashi T, Toyoda A, Kobayashi I, Yamaoka Y. Genome-wide association study of gastric cancer- and duodenal ulcer-derived *Helicobacter pylori* strains reveals discriminatory genetic variations and novel oncoprotein candidates. *Microb Genom* 2021; **7** [PMID: 34846284 DOI: 10.1099/mgen.0.000680]

11   **Dudbridge F**. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013; **9**: e1003348 [PMID: 23555274 DOI: 10.1371/journal.pgen.1003348]

12   **Seibert TM**, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK, Eeles RA, Easton DF, Kote-Jarai Z, Al Olama AA, Garcia SB, Muir K, Grönberg H, Wiklund F, Aly M, Schleutker J, Sipeky C, Tammela TL, Nordestgaard BG, Nielsen SF, Weischer M, Bisbjerg R, Røder MA, Iversen P, Key TJ, Travis RC, Neal DE, Donovan JL, Hamdy FC, Pharoah P, Pashayan N, Khaw KT, Maier C, Vogel W, Luedeke M, Herkommer K, Kibel AS, Cybulski C, Wokolorczyk D, Kluzniak W, Cannon-Albright L, Brenner H, Cuk K, Saum KU, Park JY, Sellers TA, Slavov C, Kaneva R, Mitev V, Batra J, Clements JA, Spurdle A, Teixeira MR, Paulo P, Maia S, Pandha H, Michael A, Kierzek A, Karow DS, Mills IG, Andreassen OA, Dale AM; PRACTICAL Consortium*. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* 2018; **360**: j5757 [PMID: 29321194 DOI: 10.1136/bmj.j5757]

13   **Lecarpentier J**, Silvestri V, Kuchenbaecker KB, Barrowdale D, Dennis J, McGuffog L, Soucy P, Leslie G, Rizzolo P, Navazio AS, Valentini V, Zelli V, Lee A, Amin Al Olama A, Tyrer JP, Southey M, John EM, Conner TA, Goldgar DE, Buys SS, Janavicius R, Steele L, Ding YC, Neuhausen SL, Hansen TVO, Osorio A, Weitzel JN, Toss A, Medici V, Cortesi L, Zanna I, Palli D, Radice P, Manoukian S, Peissel B, Azzollini J, Viel A, Cini G, Damante G, Tommasi S, Peterlongo P, Fostira F, Hamann U, Evans DG, Henderson A, Brewer C, Eccles D, Cook J, Ong KR, Walker L, Side LE, Porteous ME, Davidson R, Hodgson S, Frost D, Adlard J, Izatt L, Eeles R, Ellis S, Tischkowitz M; EMBRACE, Godwin AK, Meindl A, Gehrig A, Dworniczak B, Sutter C, Engel C, Niederacher D, Steinemann D, Hahnen E, Hauke J, Rhiem K, Kast K, Arnold N, Ditsch N, Wang-Gohrke S, Wappenschmidt B, Wand D, Lasset C, Stoppa-Lyonnet D, Belotti M, Damiola F, Barjhoux L, Mazoyer S; GEMO Study Collaborators, Van Heetvelde M, Poppe B, De Leeneer K, Claes KBM, de la Hoya M, Garcia-Barberan V, Caldes T, Perez Segura P, Kiiski JI, Aittomäki K, Khan S, Nevanlinna H, van Asperen CJ; HEBON, Vaszko T, Kasler M, Olah E, Balmaña J, Gutiérrez-Enríquez S, Diez O, Teulé A, Izquierdo A, Darder E, Brunet J, Del Valle J, Feliubadalo L, Pujana MA, Lazaro C, Arason A, Agnarsson BA, Johannsson OT, Barkardottir RB, Alducci E, Tognazzo S, Montagna M, Teixeira MR, Pinto P, Spurdle AB, Holland H; KConFab Investigators, Lee JW, Lee MH, Lee

J, Kim SW, Kang E, Kim Z, Sharma P, Rebbeck TR, Vijai J, Robson M, Lincoln A, Musinsky J, Gaddam P, Tan YY, Berger A, Singer CF, Loud JT, Greene MH, Mulligan AM, Glendon G, Andrulis IL, Toland AE, Senter L, Bojesen A, Nielsen HR, Skytte AB, Sunde L, Jensen UB, Pedersen IS, Krogh L, Kruse TA, Caligo MA, Yoon SY, Teo SH, von Wachenfeldt A, Huo D, Nielsen SM, Olopade OI, Nathanson KL, Domchek SM, Lorenchick C, Jankowitz RC, Campbell I, James P, Mitchell G, Orr N, Park SK, Thomassen M, Offit K, Couch FJ, Simard J, Easton DF, Chenevix-Trench G, Schmutzler RK, Antoniou AC, Ottini L. Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. *J Clin Oncol* 2017; **35**: 2240-2250 [PMID: 28448241 DOI: 10.1200/JCO.2016.69.4935]

14   **Park B**, Yang S, Lee J, Choi IJ, Kim YI, Kim J. Gastric Cancer Risk Prediction Using an Epidemiological Risk Assessment Model and Polygenic Risk Score. *Cancers (Basel)* 2021; **13** [PMID: 33669642 DOI: 10.3390/cancers13040876]

15   **Duan F**, Song C, Wang P, Ye H, Dai L, Zhang J, Wang K. Polygenic Risk Scores for Prediction of Gastric Cancer Based on Bioinformatics Screening and Validation of Functional lncRNA SNPs. *Clin Transl Gastroenterol* 2021; **12**: e00430 [PMID: 34797779 DOI: 10.14309/ctg.0000000000000430]

16   **Delcher AL**, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* 2003; **Chapter 10**: Unit 10.3 [PMID: 18428693 DOI: 10.1002/0471250953.bi1003s00]

17   **Yang C**, Pei X, Wu Y, Yan L, Yan Y, Song Y, Coyle NM, Martinez-Urtaza J, Quince C, Hu Q, Jiang M, Feil E, Yang D, Zhou D, Yang R, Falush D, Cui Y. Recent mixing of Vibrio parahaemolyticus populations. *ISME J* 2019; **13**: 2578-2588 [PMID: 31235840 DOI: 10.1038/s41396-019-0461-5]

18   **Yiangou K**, Kyriacou K, Kakouri E, Marcou Y, Panayiotidis MI, Loizidou MA, Hadjisavvas A, Michailidou K. Combination of a 15-SNP Polygenic Risk Score and Classical Risk Factors for the Prediction of Breast Cancer Risk in Cypriot Women. *Cancers (Basel)* 2021; **13** [PMID: 34572793 DOI: 10.3390/cancers13184568]

19   **Calle ML**, Urrea V, Boulesteix AL, Malats N. AUC-RF: a new strategy for genomic profiling with random forest. *Hum Hered* 2011; **72**: 121-132 [PMID: 21996641 DOI: 10.1159/000330778]

20   **Hanley JA**, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29-36 [PMID: 7063747 DOI: 10.1148/radiology.143.1.7063747]

21   **Wen S**, Moss SF. Helicobacter pylori virulence factors in gastric carcinogenesis. *Cancer Lett* 2009; **282**: 1-8 [PMID: 19111390 DOI: 10.1016/j.canlet.2008.11.016]

22   **Evans DG**, Karjalainen TK, Evans DJ Jr, Graham DY, Lee CH. Cloning, nucleotide sequence, and expression of a gene encoding an adhesin subunit protein of Helicobacter pylori. *J Bacteriol* 1993; **175**: 674-683 [PMID: 7678592 DOI: 10.1128/jb.175.3.674-683.1993]

23   **Carlsohn E**, Nyström J, Bölin I, Nilsson CL, Svennerholm AM. HpaA is essential for Helicobacter pylori colonization in mice. *Infect Immun* 2006; **74**: 920-926 [PMID: 16428735 DOI: 10.1128/iai.74.2.920-926.2006]

24   **Cai H**, Ye F, Michel A, Murphy G, Sasazuki S, Taylor PR, Qiao YL, Park SK, Yoo KY, Jee SH, Cho ER, Kim J, Chen SC, Abnet CC, Tsugane S, Cai Q, Shu XO, Zheng W, Pawlita M, Epplein M. Helicobacter pylori blood biomarker for gastric cancer risk in East Asia. *Int J Epidemiol* 2016; **45**: 774-781 [PMID: 27170766 DOI: 10.1093/ije/dyw078]

25   **Epplein M**, Zheng W, Xiang YB, Peek RM Jr, Li H, Correa P, Gao J, Michel A, Pawlita M, Cai Q, Shu XO. Prospective study of Helicobacter pylori biomarkers for gastric cancer risk among Chinese men. *Cancer Epidemiol Biomarkers Prev* 2012; **21**: 2185-2192 [PMID: 23035179 DOI: 10.1158/1055-9965.EPI-12-0792-T]

26   **Andermann TM**, Chen YT, Ottemann KM. Two predicted chemoreceptors of Helicobacter pylori promote stomach infection. *Infect Immun* 2002; **70**: 5877-5881 [PMID: 12228322 DOI: 10.1128/iai.70.10.5877-5881.2002]

27   **Williams SM**, Chen YT, Andermann TM, Carter JE, McGee DJ, Ottemann KM. Helicobacter pylori chemotaxis modulates inflammation and bacterium-gastric epithelium interactions in infected mice. *Infect Immun* 2007; **75**: 3747-3757 [PMID: 17517875 DOI: 10.1128/iai.00082-07]

28   **Yoon JY**, Kim J, Lee SJ, Kim HS, Im HN, Yoon HJ, Kim KH, Kim SJ, Han BW, Suh SW. Structural and functional characterization of Helicobacter pylori DsbG. *FEBS Lett* 2011; **585**: 3862-3867 [PMID: 22062156 DOI: 10.1016/j.febslet.2011.10.042]

29   **Lester J**, Kichler S, Oickle B, Fairweather S, Oberc A, Chahal J, Ratnayake D, Creuzenet C. Characterization of Helicobacter pylori HP0231 (DsbK): role in disulfide bond formation, redox homeostasis and production of Helicobacter cystein-rich protein HcpE. *Mol Microbiol* 2015; **96**: 110-133 [PMID: 25582190 DOI: 10.1111/mmi.12923]

30   **Zhong Y**, Anderl F, Kruse T, Schindele F, Jagusztyn-Krynicka EK, Fischer W, Gerhard M, Mejías-Luque R. Helicobacter pylori HP0231 Influences Bacterial Virulence and Is Essential for Gastric Colonization. *PLoS One* 2016; **11**: e0154643 [PMID: 27138472 DOI: 10.1371/journal.pone.0154643]

31   **Choi J**, Jia G, Wen W, Long J, Zheng W. Evaluating polygenic risk scores in assessing risk of nine solid and hematologic cancers in European descendants. *Int J Cancer* 2020; **147**: 3416-3423 [PMID: 32588423 DOI: 10.1002/ijc.33176]

32   **Jin G**, Lv J, Yang M, Wang M, Zhu M, Wang T, Yan C, Yu C, Ding Y, Li G, Ren C, Ni J, Zhang R, Guo Y, Bian Z, Zheng Y, Zhang N, Jiang Y, Chen J, Wang Y, Xu D, Zheng H, Yang L, Chen Y, Walters R, Millwood IY, Dai J, Ma H, Chen K, Chen Z, Hu Z, Wei Q, Shen H, Li L. Genetic risk, incident gastric cancer, and healthy lifestyle: a meta-analysis of genome-wide association studies and prospective cohort study. *Lancet Oncol* 2020; **21**: 1378-1386 [PMID: 33002439 DOI: 10.1016/S1470-2045(20)30460-5]

33   **Mihor A**, Tomsic S, Zagar T, Lokar K, Zadnik V. Socioeconomic inequalities in cancer incidence in Europe: a comprehensive review of population-based epidemiological studies. *Radiol Oncol* 2020; **54**: 1-13 [PMID: 32074075 DOI: 10.2478/raon-2020-0008]

34   **Quach DT**, Hiyama T, Gotoda T. Identifying high-risk individuals for gastric cancer surveillance from western and eastern perspectives: Lessons to learn and possibility to develop an integrated approach for daily practice. *World J Gastroenterol* 2019; **25**: 3546-3562 [PMID: 31367156 DOI: 10.3748/wjg.v25.i27.3546]

35   **Poorolajal J**, Moradi L, Mohammadi Y, Cheraghi Z, Gohari-Ensaf F. Risk factors for stomach cancer: a systematic review and meta-analysis. *Epidemiol Health* 2020; **42**: e2020004 [PMID: 32023777 DOI: 10.4178/epih.e2020004]

36   **Ge Z**, Taylor DE. Contributions of genome sequencing to understanding the biology of Helicobacter pylori. *Annu Rev Microbiol* 1999; **53**: 353-387 [PMID: 10547695 DOI: 10.1146/annurev.micro.53.1.353]

37   **Suerbaum S**, Achtman M. Evolution of Helicobacter pylori: the role of recombination. *Trends Microbiol* 1999; **7**: 182-184 [PMID: 10383222 DOI: 10.1016/s0966-842x(99)01505-x]