Gastroenterology 2017;152:2011–2021

Exome Array Analysis Identifies Variants in *SPOCD1* and *BTN3A2* That Affect Risk for Gastric Cancer

Meng Zhu,[1,2,*] **Caiwang Yan**,[1,*] **Chuanli Ren**,[1,3,*] **Xiaodan Huang**,[4,*] Xun Zhu,[1] Haiyong Gu,[5] Meilin Wang,[6] Shouyu Wang,[7] Yong Gao,[8] Yong Ji,[9] Xiaoping Miao,[10] Ming Yang,[11] Jinfei Chen,[12] Jiangbo Du,[1] Tongtong Huang,[1] Yue Jiang,[1] Juncheng Dai,[1] Hongxia Ma,[1] Jianwei Zhou,[7] Zhaoming Wang,[1] Zhibin Hu,[1] Guozhong Ji,[4] Zhengdong Zhang,[6] Hongbing Shen,[1] Yongyong Shi,[13] and Guangfu Jin[1,2]

[1]*Department of Epidemiology and Biostatistics, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, School of Public Health, Nanjing Medical University, Nanjing, China;* [2]*State Key Laboratory of Reproductive Medicine, International Joint Research Center for Environment and Human Health, Nanjing Medical University, Nanjing, China;* [3]*Clinical Medical Testing Laboratory, Northern Jiangsu People's Hospital and Clinical Medical College of Yangzhou University, Yangzhou, China;* [4]*Institute of Digestive Endoscopy and Medical Center for Digestive Diseases, Second Affiliated Hospital of Nanjing Medical University, Nanjing, China;* [5]*Department of Cardiothoracic Surgery, Affiliated People's Hospital of Jiangsu University, Zhenjiang, China;* [6]*Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China;* [7]*Department of Molecular Cell Biology and Toxicology, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, School of Public Health, Nanjing Medical University, Nanjing, China;* [8]*Department of Medical Oncology, The Affiliated Huai'an First People's Hospital of Nanjing Medical University, Huai'an, China;* [9]*Department of Cardiothoracic Surgery, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China;* [10]*MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, China;* [11]*Shandong Provincial Key Laboratory of Radiation Oncology, Cancer Research Center, Shandong Cancer Hospital Affiliated to Shandong University, Shandong Academy of Medical Sciences, Jinan, China;* [12]*Department of Oncology, Nanjing First Hospital, Nanjing Medical University, Nanjing, China;* [13]*Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Bio-X Institutes, Shanghai Jiao Tong University, Shanghai, China*

BACKGROUND & AIMS: Several genetic variants have been associated with gastric cancer risk, although these account for only a fraction of cases of gastric cancer. We aimed to identify low-frequency and other genetic variants that determine gastric cancer susceptibility. **METHODS:** We performed exome array analysis of DNA in blood samples from 1113 patients with gastric cancer, collected at hospitals from 2006 to 2010 in China, and 1848 individuals without cancer (controls) undergoing physical examinations. Among 71,290 variants analyzed (including 25,784 common variants), 24 variants were selected and replicated in an analysis of DNA in blood samples from 4687 additional cases of gastric cancer and 5780 controls. We compared expression of candidate genes in tumor vs normal gastric tissues using data from TCGA and performed functional annotation analyses. An immortalized human gastric epithelial cell line (GES1) and 7 human gastric cancer lines were used to express transgenes, knock down gene expression (with small interfering RNAs), disrupt genes (using the CRISPR/Cas9 system), or assess expression of reporter constructs. We measured cell proliferation, colony formation, invasion, and migration, and assessed growth of xenograft tumors in nude mice. **RESULTS:** A low-frequency missense variant rs112754928 in the SPOC domain containing 1 gene (*SPOCD1*; encoding p.Arg71Trp), at 1p35.2, was reproducibly associated with reduced risk of gastric cancer (odds ratio, 0.56; $P = 3.48 \times 10^{-8}$). *SPOCD1* was overexpressed in gastric tumors, and knockout of *SPOCD1* reduced gastric cancer cell proliferation, invasive activity, and migration, as well as growth of xenograft tumors in nude mice. We also associated the variant rs1679709 at 6p22.1 with reduced risk for gastric cancer (odds ratio, 0.80; $P = 1.17 \times 10^{-13}$). The protective allele rs1679709-A correlated with the surrounding haplotype rs2799077-T–rs2799079-C, which reduced the enhancer activity of this site to decrease expression of the butyrophilin subfamily 3 member A2 gene (*BTN3A2*). *BTN3A2* is overexpressed in gastric tumors, and deletion of *BTN3A2* inhibited proliferation, migration, and invasion of gastric cancer cells. **CONCLUSIONS:** We have associated variants at 1p35.2 and 6p22.1 with gastric cancer risk, indicating a role for *SPOCD1* and *BTN3A2* in gastric carcinogenesis.

Keywords: Stomach Cancer; Genetics; Mutation; Gene Regulation.

In China, gastric cancer is the second most frequently diagnosed cancer and the second leading cause of cancer-related death, with an estimated of 679,100 new

***Authors share co-first authorship.**

Abbreviations used in this paper: **BTN, butyrophilin; ENCODE, Encyclopedia of DNA Elements; GTEx, genotype-tissue expression; GWAS, genome-wide association studies; GWAS-BJ, the Beijing samples of our previous GWAS; GWAS-NCI, gastric cancer GWAS publicly available from dbGAP; LD, linkage disequilibrium; MAF, minor allele frequency; OR, odds ratios; SEM, standard error of the mean; sgRNA, single-guide RNA; TCGA, The Cancer Genome Atlas.**

Most current article

0016-5085/$36.00
http://dx.doi.org/10.1053/j.gastro.2017.02.017

EDITOR'S NOTES

BACKGROUND AND CONTEXT

Genome-wide association studies have identified some common genetic variants associated with gastric cancer risk, but the role of low-frequency variants in gastric cancer needs further study

NEW FINDINGS

Low-frequency missense variants in *SPOCD1* and a haplotype distantly regulating *BTN3A2* were shown to affect gastric cancer risk.

LIMITATIONS

This study did not evaluate low-frequency variants outside of coding regions.

IMPACT

This study indicates an important contribution of low-frequency variants to gastric cancer heritability, and may improve our understanding of the pathogenesis of gastric cancer.

cases and 498,000 deaths in 2015.[1] The majority of gastric cancer cases are sporadic and result from complex interplay between genetic and environmental factors. *Helicobacter pylori* (*H. pylori*) infection, cigarette smoking, and an unhealthy diet, such as consumption of pickled and smoked foods, are major environmental risk factors for gastric cancer.[2] In the presence of the exposure to environmental factors, genetic factors determine an individual's predisposition to gastric cancer.[3]

Genetic variations, including single nucleotide polymorphisms, have been determined to contribute to the pathogenesis of gastric cancer.[4] Over the past few years, we and other groups have performed genome-wide association studies (GWAS) of gastric cancer and have identified several susceptibility loci, including 1q22 (*MUC1*),[5] 3q13.31 (*ZBTB20*),[6] 5p13.1 (*PRKAA1*),[6] 5q14.3 (*lnc-POLR3G-4*),[7] 6p21.1 (*UNC5CL*),[8] 8q24 (*PSCA*),[9] and 10q23 (*PLCE1*).[10] However, GWAS have mainly focused on common proxy variants (minor allele frequency [MAF], >5%), and the identified loci explain only a small fraction of gastric cancer heritability. The remaining missing heritability has not yet been elucidated and requires further investigation.

Recently, a whole-genome sequencing-based GWAS of Icelanders revealed that low-frequency (MAF <5%) variants in *ATM* are also associated with the risk of gastric cancer.[11] In contrast with the modest effects of the individual common variants observed in the previous GWAS, these low-frequency *ATM* variants had strong effects on gastric cancer risk (odds ratios [ORs] = 4.27 and 6.87 for p.Gln852* and p.Ser644*, respectively).[11] In addition to gastric cancer, low-frequency variants have also been detected in association with lung,[12, 13] breast,[14] ovarian,[15,16] and prostate cancers.[17] Taken together, these findings support the notion that low-frequency variants, which are usually missed in general GWAS, also account for a portion of the missing heritability of cancers, including gastric cancer.

In the current study, we first screened low-frequency coding variants associated with gastric cancer risk in 1140 cases and 1854 controls using the Illumina Human Exome BeadChip platform (referred to as "exome array" hereafter). Then, we evaluated the promising associations in an additional 4687 cases and 5780 controls from Chinese populations. Consequently, a low-frequency missense variant in the SPOC domain containing 1 gene (*SPOCD1*) and a common variant related to the butyrophilin (BTN) subfamily 3 member A2 gene (*BTN3A2*) expression were identified to be significantly associated with gastric cancer risk. Further functional experiments supported the involvement of these 2 genes (*SPOCD1* and *BTN3A2*) in gastric carcinogenesis.

Materials and Methods

Study Populations

Three-stage case-control analysis was performed in this study. The demographic information for the subjects is summarized in Supplementary Table 1. In the discovery stage, 1140 cases of gastric cancer and 1854 controls were recruited in Jiangsu Province from 2006 to 2010; some of these subjects (402 cases and 649 controls) were also included in our previous gastric cancer GWAS.[6] The first replication stage (replication I) included 1073 cases and 1334 controls from Jiangsu Province, and the second replication stage (replication II) included 3614 cases and 4446 controls from Jiangsu, Hubei, and Shandong. All of the cases were recruited from local hospitals and were histopathologically confirmed as gastric cancer. The cancer-free control subjects were selected from individuals receiving routine physical examination at their local hospitals or those participating in community screening for non-communicable diseases. After signing an informed consent form, 5 mL venous blood was collected for DNA extraction from each participant, and they were then interviewed in person to obtain demographic data (eg, age and sex) and lifestyle information (eg, drinking and smoking status). This study was approved by the institutional review board of Nanjing Medical University.

Genotyping and Quality Control in the Discovery Stage

In the discovery stage, 1140 cases of gastric cancer and 1854 controls were genotyped using an exome array. The cases and controls were genotyped simultaneously by technicians who were blinded to the sample status. Genotype calling was performed using GenomeStudio software (Illumina, San Diego, CA), and cluster plots were manually checked as described by Guo et al.[18] A systematic quality control approach was then applied to filter the genetic variants and samples (Supplementary Figure 1). As a result, 176,570 variants and 30 subjects were excluded from further analysis. All of the remaining individuals had an overall genotyping rate of more than 95%. The principal component analysis did not detect population outliers, and the cases and controls were genetically matched (Supplementary Figure 2). The genomic-control inflation factor (λ) was 1.02 (Supplementary Figure 3), which indicated minimal population stratification among the subjects. To estimate the genotyping reliability of the array, we set 37

repeat samples and found a genotyping concordance rate of 99.98%. Moreover, a total of 6187 variants in 1051 subjects genotyped with the exome array were also genotyped in our previous GWAS using the Affymetrix Genome-Wide Human single nucleotide polymorphism Array 6.0 (Affymetrix, Santa Clara, CA),[6] and the concordance rate was 99.89% between these 2 platforms. Accordingly, we also excluded 3 samples and 10 variants because the concordance rates were <95% between the platforms. Finally, 71,290 variants in 1113 cases and 1848 controls were retained for further genetic association analysis. The information regarding the variants of the array is summarized in Supplementary Table 2.

Selection and Genotyping of Variants in the Replication Stages

The associations were assessed using an additive model by logistic regression analyses (Supplementary Figure 4). Promising variants were then selected for further replication according to the allele frequencies. For low-frequency variants (MAF $\leq$ 0.05), the following criteria were used: (1) an association P value $\leq 5.0 \times 10^{-3}$; (2) clear genotype cluster on visual inspection of the cluster plot; (3) inclusion of only 1 variant if multiple variants were in linkage disequilibrium (LD) ($r^2 \geq$ 0.5); (4) nonsynonymous variant or location in a splice site; (5) at least 6 copies of the minor allele in the study samples; and (6) exclusion of a variant if it was well imputed (INFO > 0.8) but association $P > .05$ in both the Beijing samples of our previous GWAS (referred to as GWAS-BJ)[6] and another gastric cancer GWAS publicly available from dbGAP (phs000361.v1.p1, including 1625 cases of gastric cancer and 2100 controls, referred to as GWAS-NCI)[10] or association direction was inconsistent in either one. A total of 25,784 qualified common variants (MAF >0.05) were also included in the exome array, among which promising variants were also selected. The same criteria (1–3) were used as for the low-frequency variants, as well as 2 additional criteria: (4) independence from the previous GWAS-identified loci; and (5) a consistent association $P <$ .05 in either the GWAS-BJ or GWAS-NCI dataset.

A total of 24 variants (21 low-frequency variants and 3 common variants) were selected and genotyped using the iPLEX Sequenom MassARRAY platform (Sequenom, San Diego, CA) in the replication I stage (Supplementary Table 3). To improve the genotype calling quality for the low-frequency variants, we included 27 positive control samples selected from the discovery stage in each 384-well plate. In the replication II stage, 3 variants showing consistent associations between the discovery and replication I stages were further genotyped by TaqMan allelic discrimination assay with an ABI 7900 system (Applied Biosystems, Carlsbad, CA). Positive and negative controls were also included in each 384-well plate for quality control, and the genotyping was performed by technicians who were blinded to the sample status.

Differential Expression and Genotype-Expression Correlation Analysis

We obtained mRNA expression and genotypic data for the gastric cancer samples from The Cancer Genome Atlas (TCGA) on April 8, 2015. The normalized expectation-maximization read counts were available for 413 samples, including 32 paired samples (tumors with adjacent normal tissues). The independent sample t-test (413 tumor samples vs 32 normal tissues) and paired sample t-test (32 paired samples) were used to examine differences in gene expression between the tumors and adjacent normal tissues. Data from the Genotype-Tissue Expression project (GTEx v6) were used to perform expression quantitative trait loci analysis of stomach tissues. A total of 412 gastric cancer samples from TCGA, for which both genotype and expression data were available, were used to validate the identified expression quantitative trait loci from GTEx. The associations between the variants and expression (log-transformed) of the corresponding genes in TCGA were evaluated using a linear regression model. Genome-wide expression correlation analysis was also performed to identify co-expressed genes in 413 gastric cancer tumor samples from TCGA. After Bonferroni correction, KEGG enrichment analysis of the significantly co-expressed genes ($P < 2.46 \times 10^{-6}$;.05/20,305 genes tested) was conducted using 'clusterProfiler' R package.[19]

RNAi, Plasmid and Transient Transfection

Specific siRNAs targeting *SPOCD1* and *BTN3A2* were custom-designed and provided by Ribobio (Guangzhou, China) (Supplementary Table 4). The pGV144-*SPOCD1* and pGV144-*BTN3A2* plasmids were constructed by Genechem (Guangzhou, China); 1.0×10^5 cells were seeded on 60-mm culture plates and transfected with oligonucleotides or plasmids using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Transfected cells were incubated at 37°C with 5% CO_2 for 48 hours.

CRISPR/Cas9-Mediated Knockout of the Gene and Enhancer Region

The deletion of *SPOCD1*, *BTN3A2*, and enhancer region with CRISPR/Cas9 system was previously described.[20] Briefly, guide RNAs were designed to recognize chr1:32,264,099-32,264,121 (*SPOCD1*), chr6:26,370,560-26,370,582 (*BTN3A2*), chr6:28,234,156-28,234,178 (*Enhancer-sgRNA1*), and chr6:28,236,326-28,236,348 (*Enhancer-sgRNA2*) (hg19) and cloned into PGL3. Constructs were introduced into gastric cancer cell line (BGC823) using Lipofectamine 2000 reagent (Invitrogen) along with a plasmid encoding Cas9 (1.0 μg of single-guide RNA [sgRNA] and 2.0 μg of Cas9) according to the manufacturer's protocol (Invitrogen). After 24 hours, Puromycin (1.0 μg mL^{-1}, Gibco, Carlsbad, CA) and Blasticidin (10.0 μg mL^{-1}, InvivoGene, San Diego, CA) were added to the medium for a 48 hours treatment and subsequently single clones were selected through serial dilution. The knockout of *SPOCD1* and *BTN3A2* genes was confirmed by sequencing and western blotting (Supplementary Figures 5 and 6). The deletion of enhancer region was confirmed by polymerase chain reaction and sequencing results (Supplementary Figure 7). sgRNA sequences and primers used for amplifying the sgRNA target site and for sequencing are listed in Supplementary Tables 4 and 5.

Mouse Xenograft Tumor Model

Animal care and handling procedures were performed in accordance with the National Institutes of Health's Guide for the Care and Use of Laboratory Animals, and were approved by the Committee on the Ethics of Animal Experiments of Nanjing Medical University (Nanjing, China). Athymic nude mice were purchased from the Vital River Laboratory Animal Technology

Co. (Beijing, China) and maintained in laminar flow cabinets under specific pathogen-free conditions. Cells (5.0×10^6) were injected subcutaneously into the bilateral armpit of 5-week-old male BALB/C nude mice. Tumor growth rate was monitored by measuring tumor diameters twice a week. Both maximum (L) and minimum (W) length of the tumor were measured using a slide caliper, and the tumor volume was calculated as $^1/_2$ LW^2. Fourteen days after injection, the mice were euthanized and tumors were collected, weighed, and analyzed.

Construction of Reporter Plasmids, Transient Transfections and the Luciferase Assay

An 1800-bp (chr6:28234211-28236075; hg19) DNA fragment, which included the potential regulatory elements of *BTN3A2* according to H3K27AC ChIP-seq data for multiple cell lines from the Encyclopedia of DNA Elements (ENCODE) project and fetal stomach tissues from Gene Expression Omnibus, were selected for enhancer reporter assays. A fragment containing 2 variants (rs2799077 and rs2799079; $r^2 = 1.00$) that were almost perfectly correlated with rs1679709 ($r^2 = 0.96$) was constructed in pGL3-Promoter Vector (Promega, Madison, WI). The constructed plasmids were sequenced to confirm the accuracy. Cells were seeded in a 24-well plate at 7.5×10^4 cells per well for 24 hours and then co-transfected with 400 ng of each of the plasmid constructs and 8 ng pRL-SV40 plasmid as a normalizing control using Lipofectamine 2000 reagent (Invitrogen). After 48 hours of culturing, the cells were lysed, and 20 μL of the resulting supernatant was used for assessment of luciferase activity using a Dual-Luciferase Reporter Assay System (Promega). These results were expressed as the ratio of firefly to Renilla luciferase activities. Each cell line was used in 3 independent transfection experiments, and each experiment was performed in triplicate.

Statistical Analysis

Genetic association analysis was performed using logistic regression, assuming an additive genetic model, implemented in PLINK.[21] In the discovery stage, the population structure was evaluated using principal component analysis with EIGENSOFT4.2 software[22] based on 4861 autosomal scaffold markers included on the exome array. The top 10 principal components of ancestry and demographic characteristics (including age, gender, smoking and drinking status) were included as covariates in the logistic regression model when estimating the ORs and 95% confidence intervals. The logistic score test[23] and the Firth bias-corrected logistic likelihood ratio test[24] were also used to assess the association results for the low-frequency variants. Variants on X chromosome were tested for males and females separately, and the results were subsequently combined with a fixed-effect meta-analysis.[25] In the replication I stage, we used age, gender, and smoking and drinking status as covariates, while in the replication II stage, age and gender were available for the multivariate adjustment. Then, joint analysis was performed to combine the results for different stages with adjusting for age, gender, and study stage. We performed 3 gene-based tests based on nonsynonymous and splice site variants with a MAF <5% (n = 45,506): a sequence kernel association test,[26] a unweighted combined multivariate collapsing burden test,[27] and a variable-threshold burden test.[28] We defined statistical significance using Bonferroni correction, and set the exome-wide association significance level at 7.01×10^{-7} and 1.95×10^{-6} for single-variant (0.05/71,290 variants tested) and gene-based analysis (0.05/[8535 genes × 3 tests]), respectively. Protein stability was predicted using 3 different methods, MUpro,[29] I-Mutant Suite,[30] and iStable.[31] We created the quantile-quantile plot and Manhattan plot using R 3.2.1, and generated regional plots using LocusZoom.[32] The variants in the exome array were annotated based on GENCODE version 7 coding transcripts,[33] dbNSFP v2.0,[34] or documentation files obtained from the Illumina Product Support Files.

Protocols for other procedures are provided in the Supplementary Materials and Methods.

Results

Identification of Genetic Variants Associated With Gastric Cancer Risk

After quality control, a total of 71,290 variants were evaluated with the association of gastric cancer risk in the discovery stage, including 1113 cases and 1848 controls, and none of the variants reached our pre-defined significance level of 7.01×10^{-7} (Supplementary Figure 4). However, the gastric cancer risk-related loci reported in the previous GWAS, including 1q22 (*MUC1*, rs2049805, OR = 0.71, $P = 3.75 \times 10^{-6}$), 5p13.1 (*PRKAA1*, rs3805495, OR = 1.28, $P = 9.74 \times 10^{-6}$), and 10q23 (*PLCE1*, rs2274223, OR = 1.35, $P = 5.06 \times 10^{-6}$), were validated in the present study with consistent directions. Next, we performed a 2-stage replication study to assess 24 promising variants according to the selection criteria described above. In the replication I stage, which included 1073 cases of gastric cancer and 1334 controls, we identified 3 variants (rs112754928, rs148342903, and rs1679709) that were consistently associated with gastric cancer risk at a $P < .05$ (Supplementary Table 3). In the replication II stage, we further genotyped these 3 variants in an additional 3614 cases of gastric cancer and 4446 controls, and finally confirmed the association of 2 variants (rs112754928 and rs1679709) with gastric cancer risk (Supplementary Table 3). After combining the results of the 3 stages and those from previous GWAS (qualified data obtained for rs1679709 but not for rs112754928), we found that the low-frequency variant rs112754928 at 1p35.2 (OR = 0.56, 95%confidence interval: 0.46-0.69, $P = 3.48 \times 10^{-8}$) and the common variant rs1679709 at 6p22.1 (OR = 0.80, 95% confidence interval: 0.72-0.83, $P = 1.17 \times 10^{-13}$) were significantly associated with gastric cancer risk without heterogeneity between studies (Table 1 and Supplementary Figure 8). However, in the gene-based analysis, we did not find any genes reaching the predefined significance level of 1.95×10^{-6} (Supplementary Table 6).

Functional Confirmation of SPOCD1 as a Gastric Cancer Susceptibility Gene at 1p35.2

After evaluating the flanking region (2 Mb upstream and downstream) of rs112754928 at 1p35.2 (Supplementary Figure 9), we found that rs112754928 and its 2 strong LD variants (rs112752591 and rs112651926) were all

Table 1. Identified Variants Associated With Gastric Cancer Risk

Chr	Variant	Major/ minor allele	Associated gene	Stage	Cases[a]	Controls[a]	MAF Cases	MAF Controls	OR (95%CI)[b]	P[b]
1p35.2	rs112754928	C/T	*SPOCD1*	Discovery	1098/15/0	1793/55/0	0.007	0.015	0.42 (0.23-0.77)	4.53×10^{-3}
				Replication I	1030/26/1	1219/62/0	0.013	0.024	0.47 (0.29-0.77)	2.69×10^{-3}
				Replication II	3372/79/5	4188/185/5	0.013	0.022	0.59 (0.45-0.77)	7.46×10^{-5}
				Combined[c]					0.56 (0.46-0.69)	3.48×10^{-8}
				Meta[d]					0.54 (0.44-0.67)	2.33×10^{-8}
6p22.1	rs1679709	G/A	*BTN3A2*	Discovery	827/261/25	1233/557/58	0.140	0.182	0.73 (0.62-0.85)	4.70×10^{-5}
				Replication I	768/264/28	888/363/55	0.151	0.181	0.82 (0.70-0.96)	1.35×10^{-2}
				Replication II	2525/865/75	2947/1294/136	0.147	0.179	0.80 (0.73-0.87)	1.10×10^{-6}
				GWAS-BJ	337/108/11	827/273/18	0.138	0.142	0.97 (0.72-1.31)	0.857
				GWAS-NCI	1199/400/26	1474/573/53	0.139	0.161	0.83 (0.73-0.95)	6.14×10^{-3}
				Combined[c]					0.80 (0.76-0.85)	1.17×10^{-13}
				Meta[d]					0.80 (0.76-0.85)	1.04×10^{-12}

Chr, chromosome; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio.
[a]Major homozygote/heterozygote/minor homozygote.
[b]Derived from the logistic regression model adjusting for age, gender, smoking and drinking status, and the top 10 principal components, if appropriate, assuming an additive genetic model.
[c]Joint analysis was performed to combine the discovery, replication stages, and existing GWAS datasets (GWAS-BJ and GWAS-NCI for rs1679709) by additional adjusting for study stage.
[d]Fixed-effect meta-analysis was used to combine the results from the discovery, replication stages, and existing GWAS datasets.

localized to the exon of the *SPOCD1* gene. These 3 variants all resulted in amino acid changes (encoding p.Arg71Trp, p.Thr349Ala, and p.Gln355Arg, respectively) and the former 2 could consistently lead to decreased protein stability of SPOCD1 as predicted *in silico* using multiple methods (Supplementary Table 7). TCGA data revealed that *SPOCD1* was significantly overexpressed in gastric cancer tumors (Supplementary Figure 10*A*). *SPOCD1* was also highly expressed in gastric cancer cell lines as compared with GES1 (Supplementary Figure 10*B*). KEGG enrichment analysis showed that the co-expressed genes of *SPOCD1* were significantly enriched in the extracellular matrix receptor interaction pathway and the PI3K-Akt signaling pathway (corrected $P = 6.68 \times 10^{-8}$ and 1.05×10^{-5}, respectively) (Supplementary Table 8). These data indicate that *SPOCD1* may account for the gastric cancer susceptibility signal at 1p35.2.

To further explore the biological significance of *SPOCD1* in gastric cancer, we first knocked down *SPOCD1* using siRNAs in HGC27 and BGC823 cells, which expressed relatively high levels of *SPOCD1*, and observed that knockdown of *SPOCD1* attenuated malignant cellular phenotypes including cell proliferation, colony formation, migration, and invasion (Supplementary Figure 11). We subsequently established a *SPOCD1* knockout BGC823 cell line using CRISPR/Cas9 system (Supplementary Figure 5), and confirmed that deletion of *SPOCD1* significantly inhibited cell proliferation (Figure 1*A* and Supplementary Figure 12*A*), clonogenicity (Figure 1*B*), as well as invasion and migration ability (Figure 1*C*). Then, we rescued the expression of *SPOCD1* in the established *SPOCD1*$^{-/-}$ BGC823 cell line, and found that the malignant cellular phenotypes were recovered as compared with non-rescued BGC823 cells (Figure 1*D*–*F* and Supplementary Figure 12*B*). Furthermore, we performed xenograft tumor assays using *SPOCD1*$^{-/-}$ BGC823 cell line, and found that the knockout of *SPOCD1* reduced growth of xenograft tumors in nude mice (Figure 2). The genetic association results, together with these functional results, suggest that the variant alleles tagged by rs112754928 may reduce the oncogenic role of *SPOCD1* and lead to a decreased gastric cancer risk.

Regulation of BTN3A2 Expression by Functional Variants at 6p22.1

We also evaluated genetic variants in the flanking region (2 Mb upstream and downstream) of rs1679709 at 6p22.1, a missense variant in *NKAPL* (encoding p.Glu398Gly). Three variants (rs9986596, rs11965538, and rs853678) in strong LD with rs16797709 showed similar *P* values ($P < 10^{-4}$) (Supplementary Table 7 and Supplementary Figure 9), and these associations were abolished after conditioning on rs1679709 (Supplementary Table 7). Next, we explored the relationships between the rs1679709 genotype and the expression of flanking genes within 2 Mb in stomach tissue based on GTEx, and found that 2 genes (*BTN3A2* and *ZSCAN26*) were significantly associated with the rs1679709 genotype after Bonferroni correction (corrected $P = 9.24 \times 10^{-3}$ and .023, respectively) (Supplementary Table 9, Figure 3*A*). The expression of *BTN3A2*, but not that of *ZSCAN26*, was also correlated with the rs1679709 genotype in 412 gastric cancer subjects from TCGA ($P = 3.06 \times 10^{-3}$) (Figure 3*B*). These data suggest that the protective allele rs1679709-A of gastric cancer is associated with decreased expression of *BTN3A2*, which may act as a gastric cancer susceptibility gene at 6p22.1.

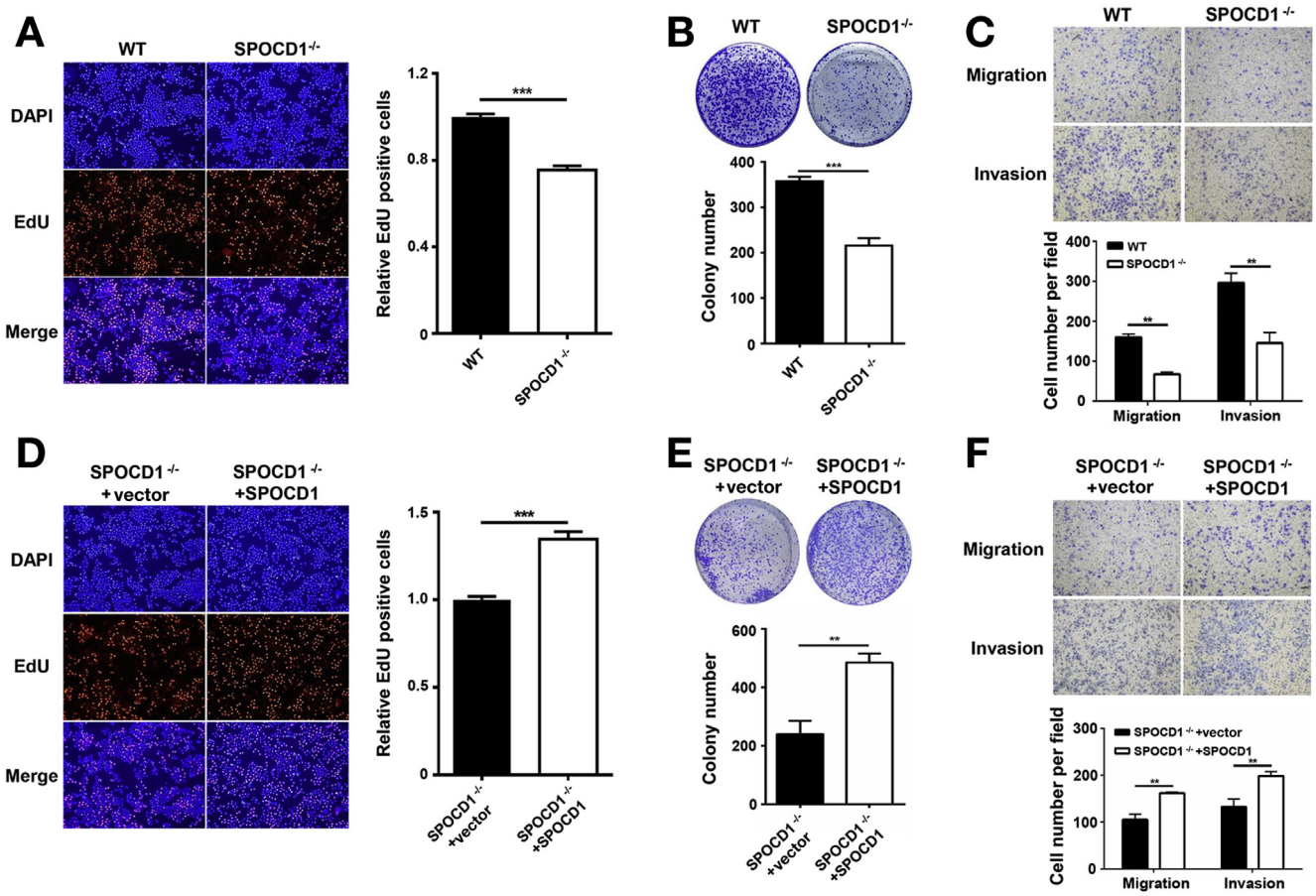

Figure 1. SPOCD1 promotes gastric cancer malignant cellular phenotypes. (*A*) EdU proliferation analysis of the effect of *SPOCD1* on the growth of *SPOCD1* knockout (*SPOCD1*$^{-/-}$) or negative control (WT) in BGC823 cells. (*B*) Colony formation assay of BGC823 cells with knockout of *SPOCD1* or negative control. The numbers of colonies were counted and were presented in a histogram. (*C*) Representative images (*top*) and quantification (*bottom*) of transwell migration and invasion assays in BGC823 cells after knockout of *SPOCD1* or in negative control. (*D–F*) Overexpression of *SPOCD1* in knockout *SPOCD1*$^{-/-}$ BGC823 cells rescued the vitalities of cell proliferation, colony formation, migration, and invasion. Representative images (*top*) and quantification (*bottom*) are shown in colony formation and transwell (migration or invasion) assays. Error bars represent SEM, n = 3. All experiments were performed at least 3 times and data analysis was conducted by 2-sided *t*-test. $^{**}P < .01$, $^{***}P < .001$.

To determine potential functional variants at 6p22.1, we mapped the variants in strong LD ($r^2 > 0.8$) with rs1679709 to regulatory elements based on data from ENCODE, and detected 2 noncoding variants, rs2799077 and rs2799079 ($r^2 = 1.00$ with each other, and $r^2 = 0.96$ with rs1679709), located in an enhancer region and tagged by H3K27AC histone modification (Supplementary Table 10). These findings were further supported by ChIP-seq data for H3K27AC and H3K4ME1 in fetal stomach tissues from Gene Expression Omnibus (GSM1102783 and GSM1102794) (Figure 3*C*). To validate these findings, we conducted luciferase assay to evaluate the regulatory effects of rs2799077 and rs2799079 in vitro. We found that the variant haplotype rs2799077-T– rs2799079-C, which was highly correlated with the protective allele rs1679709-A, significantly reduced the enhancer activity of the reporter gene compared with wild haplotype rs2799077-C– rs2799079-A (Figure 3*D*). To investigate the potential regulating target gene of the enhancer element, we deleted the epigenetically marked enhancer region containing the variants rs2799077 and rs2799079 by CRISPR/Cas9-mediated genome editing. A significant reduction (about 26%) in *BTN3A2* gene expression was observed after Cas9-mediated enhancer deletion (Figure 3*E*), confirming *BTN3A2* as a regulatory gene of the enhancer. Taken together, these results suggest that the variants rs2799077 and rs2799079 may regulate *BTN3A2* expression by modulating enhancer function.

BTN3A2 Promotes Gastric Cancer Cell Proliferation and Invasion

Based on the expression data from TCGA, *BTN3A2* was significantly overexpressed in gastric cancer tumors (Supplementary Figure 10*C*). We also observed high expression levels of *BTN3A2* in gastric cancer cell lines as compared with GES1 (Supplementary Figure 10*D*). KEGG enrichment analysis based on co-expressed genes in gastric cancer tumors showed multiple significant immune-related pathways, including antigen processing and presentation, and natural killer cell-mediated cytotoxicity

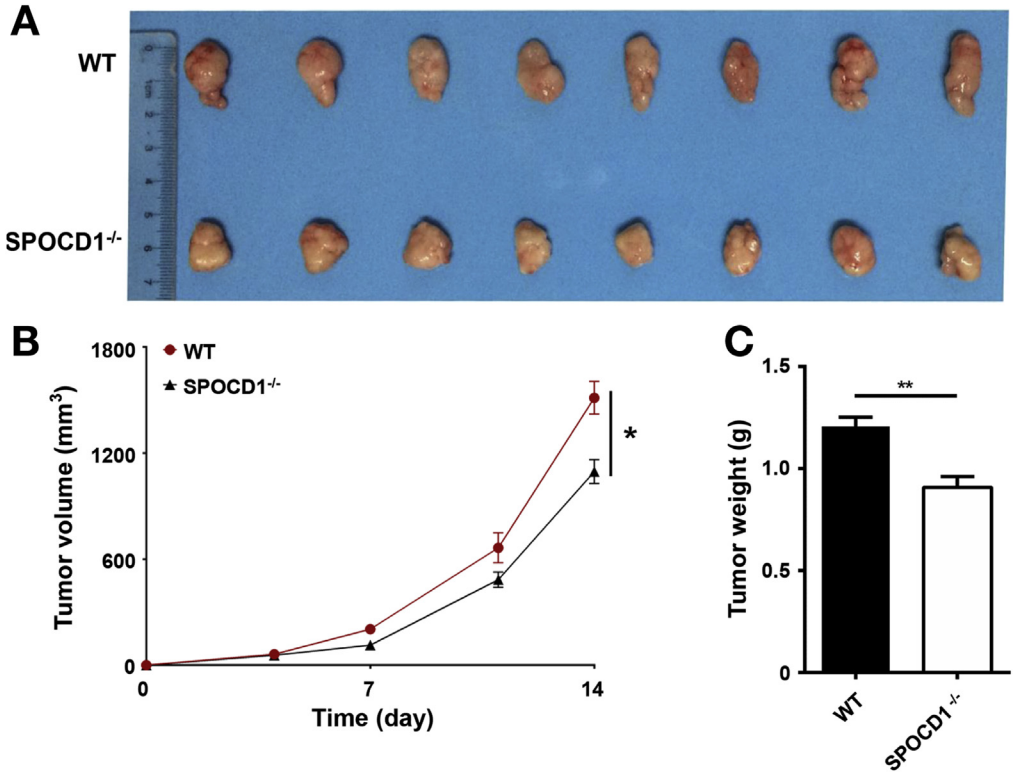

Figure 2. *SPOCD1* knockout suppresses tumorigenicity of gastric cancer cells in nude mice. In vivo xenograft tumor formation assays were performed using *SPOCD1* knockout (*SPOCD1*$^{-/-}$) and negative control (WT) cells (5.0×10^6) subcutaneously injected into the bilateral armpit (*left*: WT; *right*: *SPOCD1*$^{-/-}$) of 5-week-old male BALB/C nude mice. Tumor growth was measured twice a week. At day 14, mice were sacrificed and tumors were photographed. (*A*) Photograph of excised tumor tissues from mice (*top*, tumors from WT; *bottom*, tumors from *SPOCD1*$^{-/-}$). (*B*) Mean volumes of xenograft tumors in WT or *SPOCD1*$^{-/-}$ groups. (*C*) Average tumor weight of WT or *SPOCD1*$^{-/-}$ groups when the tumors were harvested. Data analysis was conducted by 2-sided *t*-test. *$P < .05$; **$P < .01$.

(Supplementary Table 8). To evaluate the role of *BTN3A2* in gastric cancer development, we knocked down *BTN3A2* by RNA interference, and found that knockdown of *BTN3A2* attenuated malignant cellular phenotypes in both HGC27 and BGC823 cell lines (Supplementary Figure 13). Furthermore, we knocked out the *BTN3A2* in BGC823 cell line using CRISPR/Cas9 system (Supplementary Figure 6), and then rescued the expression of *BTN3A2* in the established *BTN3A2*$^{-/-}$ BGC823 cell line. In consistent with the results from RNA interference, deletion of *BTN3A2* significantly inhibited cellular malignant phenotypes (Figure 4*A–C* and Supplementary Figure 12*C*), while these phenotypes were recovered in the rescued BGC823 cell line (Figure 4*D–F* and Supplementary Figure 12*D*). These findings suggest that *BTN3A2* at 6p22.1 probably contributes to gastric cancer development as a susceptibility gene, even though we did not observe significant change in the xenograft model (Supplementary Figure 14).

Discussion

In this study, we performed exome-wide association analysis of gastric cancer using 5800 cases and 7628 controls from Chinese populations, and identified a low-frequency missense variant, rs112754928 at 1p35.2 and a common variant rs1679709 at 6p22.1, that conferred gastric cancer risk. Further analysis showed that the missense variants in strong LD with rs112754928 were all restricted to *SPOCD1*, and were predicted to reduce the protein stability of SPOCD1. *SPOCD1* was significantly up-regulated in gastric tumors, and knockout of *SPOCD1* reduced the gastric cancer cell proliferation, clonogenicity, and migration or invasion abilities, which could be recovered after rescuing the expression of *SPOCD1*. Moreover, the oncogenic role of *SPOCD1* was also supported by xenograft model. The protective allele rs1679709-A at 6p22.1 was associated with a decreased expression of *BTN3A2*, which was up-regulated in gastric tumors. The haplotype rs2799077-T– rs2799079-C was strongly correlated with the rs1679709-A allele; in addition, it was mapped to an enhancer and was observed to reduce its activity using a luciferase assay. Deletion of the enhancer significantly reduced the expression of *BTN3A2*. These results suggested that rs2799077 and rs2799079 are functional variants associated with gastric cancer risk, by regulating *BTN3A2* expression through modulating its enhancer activity. Deletion of *BTN3A2* could inhibit the cellular malignant phenotypes, and these phenotypes could be recovered by rescuing its expression. These findings indicate that genetic variants at 1p35.2 and 6p22.1 contribute to gastric cancer susceptibility, and that *SPOCD1* and *BTN3A2* are involved in gastric carcinogenesis.

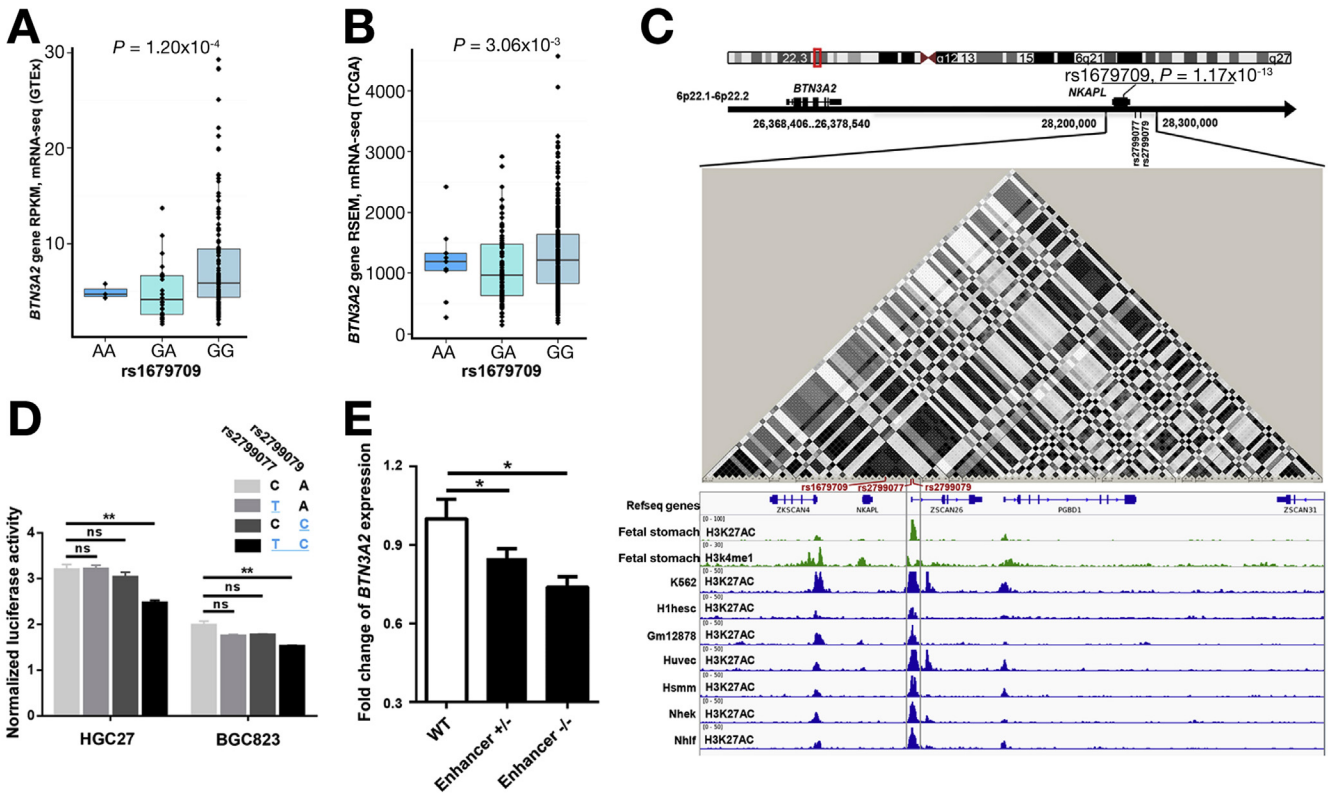

Figure 3. The variant haplotype of rs2799077-T– rs2799079-C reduces *BTN3A2* expression through reducing an enhancer activity. (*A–B*) The expression quantitative trait loci analysis between rs1679709 and *BTN3A2* in GTEx and TCGA. (*C*) Active epigenetic signature of chromatin at the 6p22.1 locus. *Top*, LD analysis of 100-kb region around rs1679709 based on CHB of 1000 Genomes Project data. *Bottom*, 2 single nucleotide polymorphisms (rs2799077 and rs2799079) were annotated into an enhancer region based on H3K27AC ChIP-seq data from ENCODE and Gene Expression Omnibus (GSM1102783 and GSM1102794). (*D*) Luciferase reporter assay for *BTN3A2* enhancer region. The minor alleles of rs2799077 and rs2799079 are underlined. All constructs were cotransfected with pRL-SV40 to standardize the transfection efficiency. The results of luciferase activity were normalized by PGL3 promoter. (*E*) Expression level of *BTN3A2* was measured by qRT-polymerase chain reaction in BGC823 cells with and without deletion of enhancer region. Error bars in *D* and *E* represent SEM. *P* values were derived from *t*-tests: $^{*}P < .05$, $^{**}P < .01$.

GWAS have successfully identified thousands of disease-/trait-related variants. However, these studies have mainly focused on common variants and failed to evaluate the roles of low-frequency variants. The exome array used in this study was specifically developed to capture low-frequency variants in coding regions on the basis of genetic variants identified from the whole-exome sequencing of >12,000 individuals. This platform has also been used to identify low-frequency variants in other complex diseases and traits.[13,35–38] To the best of our knowledge, this is the first study to evaluate the effects of low-frequency variants on gastric cancer susceptibility, and we identified a low-frequency missense variant in *SPOCD1* that was consistently associated with gastric cancer risk in different Chinese populations. The frequency of this variant, however, is very low or monomorphic in other populations, including the European ancestry (0.005), according to the 1000 Genomes project data. In addition, we also analyzed the low-frequency coding variants in *ATM* that have been reported to be associated with gastric cancer risk in European ancestry,[11] but we did not identify any *ATM* variants associated with gastric cancer risk in the current study. The disparities of genetic associations between ethnicities are more prominent for low-frequency variants than for common variants because low-frequency variants arose recently in an extended pedigree and are thus likely to be of recent origin.[39] These variants can have causative roles in pathogenesis but are likely to be restricted to a specific population. Nevertheless, the identification of *ATM* and *SPOCD1* suggests that low-frequency variants are also important in gastric cancer susceptibility, and that they may contribute to the missing heritability.

SPOCD1 was predicted to encode a protein belonging to the TFIIS family of transcription factors.[40] To date, few studies have examined the function of this gene, although our results clearly showed a promoting role of *SPOCD1* in gastric carcinogenesis. KEGG enrichment analysis of *SPOCD1* co-expressed genes indicated that *SPOCD1* might acts as a gastric cancer susceptibility gene by potentially regulating the expression of genes in the key pathways of carcinogenesis, such as the extracellular matrix receptor interaction pathway and the PI3K-Akt signaling pathway. Of interest, the variant alleles of the identified low-frequency variants in *SPOCD1* were protective against gastric cancer risk. Similar

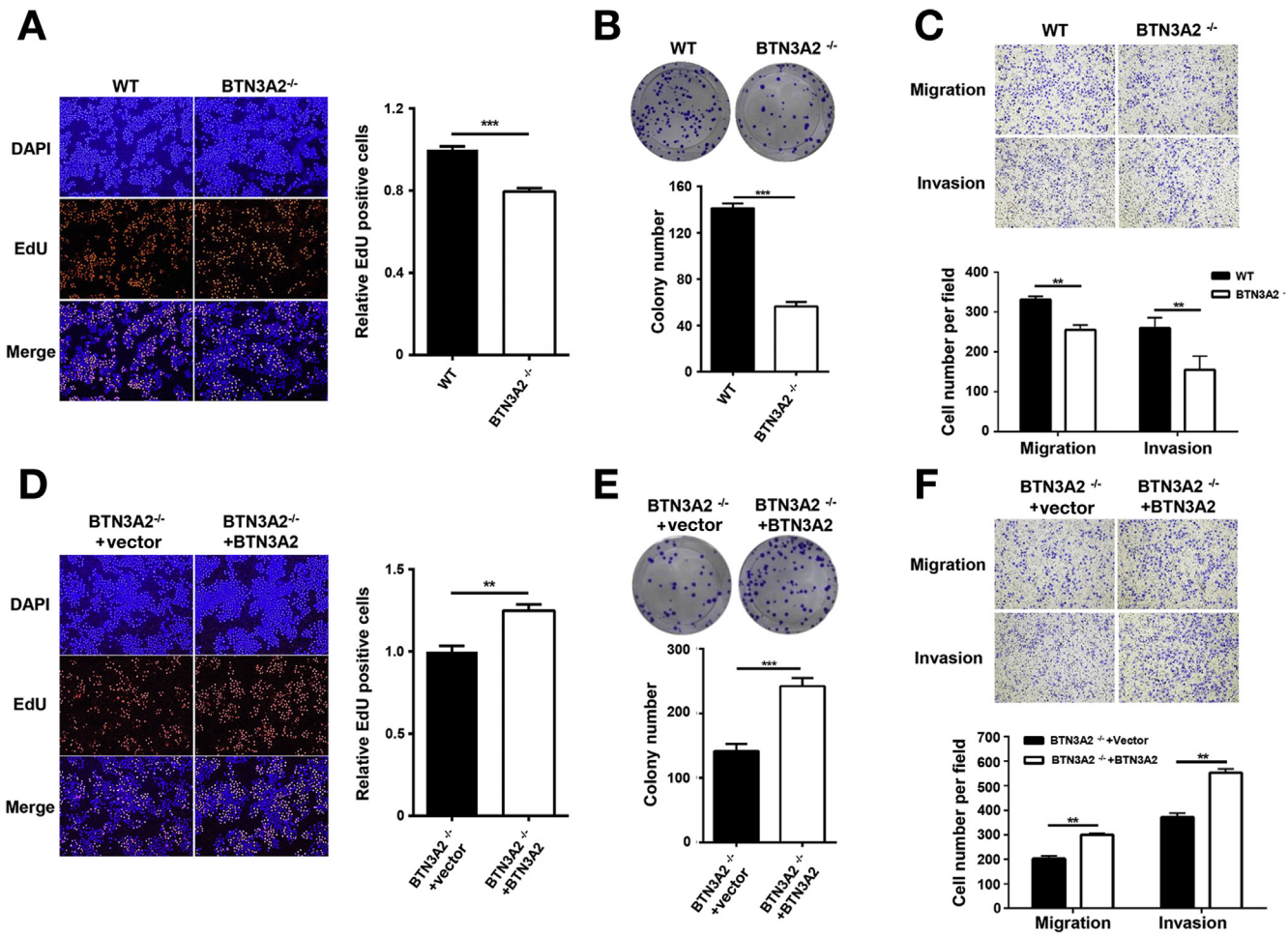

Figure 4. BTN3A2 promotes gastric cancer cellular malignant phenotypes. (*A–C*) Knockout of *BTN3A2* (*BTN3A2*$^{-/-}$) reduced cell proliferation, colony formation, migration, and invasion of BGC823 cells. (*D–F*) Overexpression of *BTN3A2* in the *BTN3A2*$^{-/-}$ BGC823 cells rescued the vitalities of cell proliferation, colony formation, migration, and invasion. Representative images (*top*) and quantification (*bottom*) are shown in colony formation and transwell (migration or invasion) assays. Error bars represent SEM, n = 3. All experiments were performed at least 3 times and 2-sided *t*-test was used for analysis. $^{**}P < .01$, $^{***}P < .001$.

results were also found in other studies, including rs146753414 in *KRT83* for gastric cancer (MAF = 0.006, OR = 0.26),[11] rs17879961 in *CHEK2* for lung cancer (MAF = 0.01, OR = 0.38),[12] and a del12 mutation in *ASGR1* for coronary artery disease (MAF ranging from 0.27% to 0.83%, OR = 0.66).[41] These evidences indicate that the low-frequency variants may not always be deleterious. These variants might be in the absence of selective pressure because they are implicated with late-onset diseases and have no influence on reproduction.[42] The amino acid changes resulted from low-frequency variants in *SPOCD1* were consistently predicted to decrease protein stability. Therefore, it is biologically plausible that the variant alleles tagged by rs112754928 at 1p35.2 may reduce the oncogenic role of SPOCD1 and lead to a decreased risk of gastric cancer.

In addition to the low-frequency variants in *SPOCD1*, we also identified a common variant at 6p22.1. This locus has not been reported in any previous GWAS, and the lead variant rs1679709 was also nominally associated with gastric cancer risk in our previous GWAS (OR = 0.84, $P = .054$) and GWAS-NCI (OR = 0.83, $P = .006$). In contrast with other cancers, such as prostate[43] and breast cancers,[44] which have been widely investigated and associated with approximately 100 loci by GWAS with a large sample size, gastric cancer has been less frequently studied and more attention should be taken. Identification of the variant rs1679709 may improve our understanding of gastric cancer susceptibility. Importantly, we found that the variant rs1679709 might be a proxy of the functional haplotype of rs2799077-rs2799079, which was located in an enhancer region and affected the enhancer activity in luciferase assay. The genotypes of rs1679709 were also related to the *BTN3A2* expression and the deletion of the enhancer region resulted in decreased expression of *BTN3A2*. *BTN3A2* encodes a protein (also known as CD277) that belongs to the human BTN 3 molecules and usually acts as a co-regulator of the immune signal in T and natural killer cells.[45] *BTN3A2* has been reported to play roles in T-cell responses in the adaptive immune response,[46] and to inhibit the release of interferon-gamma from an activated natural killer cell line.[46] The important role of *BTN3A2* in immune

regulation was also supported by KEGG enrichment analysis based on co-expressed genes in gastric cancer tumors of TCGA. Our functional experiments indicate that *BTN3A2* can promote gastric cancer cell proliferation, migration, and invasion in vitro, but no effect was observed in xenograft model. Future studies in genetically engineered mouse models with *BTN3A2* would provide more direct evidence for the oncogenic role of *BTN3A2*. Collectively, the common variants at 6p22.1 may regulate the expression of *BTN3A2* and involve in gastric cancer development, possibly via the regulation of immune related pathways.

In summary, we have identified 1 missense low-frequency variant in *SPOCD1* and 1 common locus of *BTN3A2* that are associated with gastric cancer risk. The phenotypic variance explained by these 2 newly identified variants was about 0.47%, and this value could increase to 1.74% in combination with 4 previously reported loci (1q22, 5p13.1, 8q24.3, and 10q23). We also provided further evidence for the biological role of the *SPOCD1* and *BTN3A2* genes in gastric cancer development. These findings indicate the contribution of low-frequency variants to gastric cancer heritability, and may improve our understanding of the pathogenesis of gastric cancer. However, this study was also limited by the content of the exome array, which included low-frequency variants mainly in coding regions, and further studies would benefit from whole genome sequencing-based analysis, as the costs of sequencing have steadily decreased, to assess the roles of noncoding and rarer variants in gastric cancer susceptibility.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2017.02.017.

References

1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115–132.
2. Lochhead P, El-Omar EM. Gastric cancer. Br Med Bull 2008;85:87–100.
3. Sampson JN, Wheeler WA, Yeager M, et al. Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. J Natl Cancer Inst 2015;107:djv279.
4. Tan P, Yeoh KG. Genetics and molecular pathogenesis of gastric adenocarcinoma. Gastroenterology 2015;149:1153–1162.
5. Saeki N, Saito A, Choi IJ, et al. A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. Gastroenterology 2011;140:892–902.
6. **Shi Y, Hu Z, Wu C**, et al. 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.
7. **Wang Z, Dai J, Hu N**, et al. Identification of new susceptibility loci for gastric non-cardia adenocarcinoma: pooled results from two Chinese genome-wide association studies. Gut 2017;66:581–587.
8. **Hu N, Wang Z, Song X**, et al. Genome-wide association study of gastric adenocarcinoma in Asia: a comparison of associations between cardia and non-cardia tumours. Gut 2016;65:1611–1618.
9. **Sakamoto H, Yoshimura K, Saeki N**, et al. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. Nat Genet 2008;40:730–740.
10. **Abnet CC, Freedman ND, Hu N**, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet 2010;42:764–767.
11. **Helgason H, Rafnar T**, Olafsdottir HS, et al. Loss-of-function variants in ATM confer risk of gastric cancer. Nat Genet 2015;47:906–910.
12. **Wang Y, McKay JD**, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46:736–741.
13. **Jin G, Zhu M**, Yin R, et al. Low-frequency coding variants at 6p21.33 and 20q11.21 are associated with lung cancer risk in Chinese populations. Am J Hum Genet 2015;96:832–840.
14. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet 2008;40:17–22.
15. **Loveday C, Turnbull C**, Ramsay E, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. Nat Genet 2011;43:879–882.
16. **Rafnar T, Gudbjartsson DF, Sulem P**, et al. Mutations in BRIP1 confer high risk of ovarian cancer. Nat Genet 2011;43:1104–1107.
17. **Ewing CM, Ray AM, Lange EM**, et al. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med 2012;366:141–149.
18. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. Nat Protoc 2014;9:2643–2662.
19. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–287.
20. **Wang C, Gu Y, Zhang K**, et al. Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. Nat Commun 2016;7:10499.
21. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
22. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
23. Lin DY, Tang ZZ. A general framework for detecting disease associations with rare variants in sequencing studies. Am J Hum Genet 2011;89:354–367.
24. Wang X. Firth logistic regression for rare variant association tests. Front Genet 2014;5:187.
25. Tukiainen T, Pirinen M, Sarin A-P, et al. Chromosome x-wide association study identifies loci for fasting insulin

and height and evidence for incomplete dosage compensation. PLoS Genet 2014;10:e1004127.
26. **Wu MC, Lee S**, Cai T, et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89:82–93.
27. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008;83:311–321.
28. **Price AL, Kryukov GV**, de Bakker PI, et al. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 2010;86:832–838.
29. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 2006;62:1125–1132.
30. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33:W306–W310.
31. Chen CW, Lin J, Chu YW. iStable: off-the-shelf predictor integration for predicting protein stability changes. BMC Bioinformatics 2013;14:S5.
32. **Pruim RJ, Welch RP**, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336–2337.
33. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 2012;22:1760–1774.
34. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat 2013;34:E2393–E2402.
35. Huyghe JR, Jackson AU, Fogarty MP, et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet 2013;45:197–201.
36. Igartua C, Myers RA, Mathias RA, et al. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat Commun 2015;6:5965.
37. Tachmazidou I, Dedoussis G, Southam L, et al. A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. Nat Commun 2013;4:2872.
38. **Tang CS, Zhang H, Cheung CY**, et al. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. Nat Commun 2015;6:10206.
39. Lupski JR, Belmont JW, Boerwinkle E, et al. Clan genomics and the complex architecture of human disease. Cell 2011;147:32–43.
40. Kimura K, Wakamatsu A, Suzuki Y, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res 2006;16:55–65.
41. Nioi P, Sigurdsson A, Thorleifsson G, et al. Variant ASGR1 associated with a reduced risk of coronary artery disease. N Engl J Med 2016;374:2131–2141.
42. McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010;141:210–217.
43. **Al Olama AA, Kote-Jarai Z**, Berndt SI, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet 2014;46:1103–1109.
44. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet 2015;47:373–380.
45. Messal N, Mamessier E, Sylvain A, et al. Differential role for CD277 as a co-regulator of the immune signal in T and NK cells. Eur J Immunol 2011;41:3443–3454.
46. **Palakodeti A, Sandstrom A**, Sundaresan L, et al. The molecular basis for modulation of human Vgamma9V-delta2 T cell responses by CD277/butyrophilin-3 (BTN3A)-specific antibodies. J Biol Chem 2012;287:32780–32790.

Author names in bold designate shared co-first authorship.

Received March 18, 2016. Accepted February 21, 2017.

Reprint requests
Address requests for reprints to: Guangfu Jin, MD, PhD, Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China. e-mail: guangfujin@njmu.edu.cn; fax: +86-25-8686-8499; or Yongyong Shi, PhD, Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200030, China. e-mail: shiyongyong@gmail.com; or Hongbing Shen, MD, PhD, Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China. e-mail: hbshen@njmu.edu.cn; fax: +86-25-8686-8499.

Conflicts of interest
The authors disclose no conflicts.

Funding
This work was supported by grants from the National Basic Research Program (973) (grant no. 2013CB910304); the National Key Research and Development Program of China (grant no. 2016YFC1302703); the National Natural Science Foundation of China (grant nos. 81422042, 81373090, 8152100014, 81230067, 81301823, 31550003); the Science Foundation for Distinguished Young Scholars in Jiangsu (grant no. BK20130042); the Key Grant of Natural Science Foundation of Jiangsu Higher Education Institutions (grant no. 15KJA330002); the Top-notch Academic Programs Project of Jiangsu Higher Education Institution (grant no. PPZY2015A067); and the Priority Academic Program for the Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).