Supplementary appendix

Appendix 1. Subjects and genotyping arrays of GWASs

1.1 East-GWAS and North-GWAS: We newly genotyped 6,483 cases and 5,488 controls using the Illumina Global Screening Array (GSA), which included two studies (East-GWAS: 4,872 cases and 3,397 controls from Jiangsu province and Shanghai; North-GWAS: 1,611 cases and 2,091 controls from Shandong province, Hebei province and Tianjin). We consecutively recruited histopathologically confirmed gastric cancer cases from hospitals. Cancer-free controls were selected from individuals receiving routine physical examination at hospitals or those participating in community screening for non-communicable diseases. Demographic characteristics of all participants were displayed in **Table S1**.

1.2 Onco-GWAS: Histopathologically confirmed gastric cancer cases were consecutively recruited from hospitals in Jiangsu province, China. The cancer-free control subjects were selected from individuals receiving routine physical examination at hospitals or those participating in community screening for non-communicable diseases in Jiangsu province. A total of 1,140 cases and 345 controls were genotyped using the Illumina OncoArray, and 708 controls were genotyped using the Illumina OmniZhongHua chips (**Table S1**). Detailed study design and genotype calling was provided previously.[1]

1.3 NJ-GWAS and BJ-GWAS: For the NJ-GWAS and BJ-GWAS, individuals were derived from separate case-control studies conducted in Nanjing (565 cases and 1,162 controls) and Beijing (468 cases and 1,123 controls) (**Table S1**). Individuals were genotyped using the Affymetrix Genome-Wide Human SNP Array (V.6.0), which consisting ~ 900,000 markers. The details of study design and relevant data were reported previously.[2]

1.4 SX-GWAS: A total of 1625 gastric cancer cases and 2100 controls were from the Shanxi Upper Gastrointestinal Cancer Genetics Project and the Linxian Nutrition Intervention Trial (**Table S1**). All participants were genotyped using the Illumina 660W Quad chip. The study was reported elsewhere[3] and the genotype data was downloaded from dbGap (study accession: phs00361.v1.p1).

Appendix 2. Quality control and imputation procedures

2.1 Quality control. The same protocol of quality control procedures on genotyping data was applied for all six GWAS datasets. The genotyped variants were excluded if they had a call rate of <95%, a *P* value for Hardy–Weinberg Equilibrium (HWE) in controls $\leq 1.0 \times 10^{-6}$ or a minor allele frequency (MAF) of <1% in controls; and samples were removed if they were with call rates of <95%, outliers (>6 s.d. from the mean) in population stratification analysis and heterozygosity analysis, or duplicated or related individuals (PI_HAT > 0.25).

A total of 100,641 samples in the CKB cohort were genotyped with a customized Affymetrix Axiom® CKB array. Samples with call rate < 98% or gender discrepancy, or samples with extreme heterozygosity (F statistic S.D. score <5) were excluded. Variants with call rate <95% were excluded. Variants with call rate ≥ 95% and < 98%, or deviation from the expected frequency as observed in the 1000 Genomes project (the Phase III integrated variant set release, 504 East Asians), or deviation from Hardy-Weinberg disequilibrium were selected for manual inspection. Variants that failed manual inspection were excluded.

2.2 Imputation. Detailed description of the imputation process used in this study has been shown previously[1]. Briefly, we used SHAPEIT (v. 2.12)[4] to phase qualified genotypes for each chromosome. Imputation was performed for each 5-Mb interval with IMPUTE2 (v. 2.3.1)[5] using a merged reference panel from 1000 Genomes Project (the Phase III integrated variant set release, across 2,504 samples). In particular, we re-imputed the HLA region (Chr6: 28Mb-34Mb) using SNP2HLA with a Chinese reference panel of 10,689 samples. The imputed variants were excluded from the analysis if they were poorly imputed (imputation quality score INFO < 0.5) or had low-frequency (MAF < 0.005). As a result, 7.68 to 8.30 million qualified variants were retained in each GWAS dataset, and a total of 6,957,974 shared variants were available for genetic association analysis. The imputation procedures used in the CKB cohort were similar to above process, except that IMPUTE v4 were used for the imputation analysis.

Table S1. The characteristics of subjects included in the six GWASs

Variables	East-GWAS Cases (n=4872)	East-GWAS Controls (n=3397)	North-GWAS Cases (n=1611)	North-GWAS Controls (n=2091)	BJ-GWAS Cases (n=456)	BJ-GWAS Controls (n=1118)
Age (Mean ± S.D.), y	61.34±10.14	60.29±10.44	59.74±11.48	59.36±9.02	56.79±12.42	62.44±9.17
Age (%)						
<60	1926 (39.53)	1501 (44.19)	703 (43.64)	1003 (47.97)	256 (56.14)	437 (39.09)
≥60	2946 (60.47)	1896 (55.81)	908 (56.36)	1088 (52.03)	200 (43.86)	681 (60.91)
Gender (%)						
Male	3535 (72.56)	1977 (58.20)	1191 (73.93)	1335 (63.85)	322 (70.61)	873 (78.09)
Female	1337 (27.44)	1420 (41.80)	420 (26.07)	756 (36.15)	134 (29.39)	245 (21.91)
Smoking status (%) [b]						
Smokers	--	--	--	--	129 (28.29)	590 (52.77)
Nonsmokers	--	--	--	--	327 (71.71)	528 (47.23)
Drinking status (%) [c]						
Drinkers	--	--	--	--	86 (18.86)	415 (37.12)
Nondrinkers	--	--	--	--	370 (81.14)	703 (62.88)

Table S1. The characteristics of subjects included in the six GWASs (Continue)

Variables	NJ-GWAS Cases (n=550)	NJ-GWAS Controls (n=1155)	Onco-GWAS [a] Cases (n=1140)	Onco-GWAS [a] Controls (n=1053)	SX-GWAS Cases (n=1625)	SX-GWAS Controls (n=2100)
Age (Mean ± S.D.), y	58.24±11.91	59.02±9.74	61.42±10.76	61.04±8.68	--	--
Age (%)						
<60	292 (53.09)	583 (50.48)	442 (38.77)	541 (51.37)	731 (44.98)	1000 (47.62)
≥60	258 (46.91)	572 (49.52)	698 (61.23)	512 (48.62)	894 (55.02)	1100 (52.38)
Gender (%)						
Male	392 (71.27)	823 (71.26)	844 (74.04)	751 (71.32)	1260 (77.54)	1430 (68.10)
Female	158 (28.73)	332 (28.74)	296 (25.96)	302 (28.68)	365 (22.46)	670 (31.90)
Smoking status (%) [b]						
Smokers	250 (45.45)	565 (48.92)	441 (40.35)	524 (49.76)	--	--
Nonsmokers	300 (54.55)	590 (51.08)	652 (59.65)	528 (50.14)	--	--
Drinking status (%) [c]						
Drinkers	212 (38.55)	621 (53.77)	392 (35.73)	397 (37.70)	--	--
Nondrinkers	338 (61.45)	534 (46.23)	705 (74.27)	656 (62.30)	--	--

[a] Smoking and drinking status were available for a subset of subjects in the Onco-GWAS.

[b] Smokers were defined as individuals who smoked at least one cigarette per day for more than one year during their lifetime; otherwise, they were considered nonsmokers. [6]

[c] Drinkers were defined as individuals who drank an average of twice or more in a week for at least one year in their lifetime; otherwise, they were considered nondrinkers.

Abbreviations: GWAS, genome-wide association study; SD, standard deviation.

Table S2. Polygenic risk scores (PRSs) derivation and associations with gastric cancer risk in the China Kadoorie Biobank (CKB) cohort

| Threshold of P | Number of SNPs | | | Association of PRS with gastric cancer (per SD) [a] | |
	GWAS meta-analysis	Shared SNPs for CKB	PRS deviation after LD pruning	HR (95% CI)	P value
5×10^{-8}	764	612	12	1.24 (1.15-1.33)	3.42E-08
5×10^{-7}	1033	829	18	1.26 (1.17-1.36)	9.83E-10
5×10^{-6}	1299	1007	38	1.26 (1.17-1.36)	1.25E-09
5×10^{-5}	**2220**	**1633**	**112**	**1.27 (1.17-1.37)**	**7.56E-10**
5×10^{-4}	8259	5739	539	1.18 (1.09-1.27)	1.67E-05

[a] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry. HR denotes the increase of hazard risk of gastric cancer per SD for PRS.

Abbreviations: CI, confidence interval; CKB, China Kadoorie Biobank; GWAS, genome-wide association study; HR, hazards ratio; LD, linkage disequilibrium; OR, odds ratio; PRS, polygenic risk score; SD, standard deviation; SNP, single nucleotide polymorphism.

Table S3. Summary results of polygenic risk score involving 112 SNPs in GWAS meta-analysis and CKB cohort

No.	Chr.	Position (bp)	SNP ID	Alleles (eff./ref.)	GWAS meta-analysis [a]			CKB cohort [b]		
					EAF	OR	P	EAF	HR	P
1	1	28181891	rs12039878	A/G	0.24	1.11	2.90E-05	0.25	1.00	9.80E-01
2	1	147826275	rs543163	T/G	0.60	1.10	1.43E-05	0.62	0.96	4.99E-01
3	1	155168930	rs7366775	A/G	0.80	1.41	1.39E-34	0.78	1.30	2.41E-04
4	1	155254135	rs116867045	A/G	0.05	0.79	2.29E-05	0.06	0.77	4.53E-02
5	1	155434900	rs147823566	A/G	0.99	1.60	1.87E-05	0.99	2.06	4.22E-02
6	1	155797492	rs72706172	A/C	0.82	1.29	1.47E-15	0.82	1.05	4.93E-01
7	1	156089034	rs138554234	A/G	0.05	0.64	4.26E-15	0.06	0.70	8.17E-03
8	1	208048468	rs150333442	T/C	0.03	0.71	1.08E-05	0.02	1.01	9.77E-01
9	1	230303236	rs56946451	A/G	0.29	1.11	1.38E-05	0.30	0.89	5.69E-02
10	1	240496700	rs4460625	T/C	0.24	0.89	1.53E-05	0.22	1.05	4.41E-01
11	2	23860168	rs2288697	A/G	0.49	1.10	1.80E-05	0.47	0.92	1.44E-01
12	2	41506889	rs10202299	T/C	0.53	1.10	9.44E-06	0.56	1.01	8.83E-01
13	2	43359127	rs12471190	T/C	0.75	0.90	4.24E-06	0.74	0.95	4.48E-01
14	2	52227956	rs77885082	A/C	0.97	0.76	4.28E-05	0.96	0.85	2.36E-01
15	3	10306931	rs2544001	T/G	0.98	0.74	4.51E-06	0.97	1.26	2.14E-01
16	3	38382348	rs151629	A/C	0.50	1.10	2.57E-05	0.49	1.06	2.71E-01
17	3	100384446	rs56232573	T/C	0.95	0.81	2.68E-05	0.95	0.89	3.05E-01
18	3	169039385	rs11927141	A/G	0.83	0.88	1.88E-05	0.84	0.93	3.41E-01
19	4	90709741	rs3775439	A/G	0.55	0.92	4.47E-05	0.54	1.06	3.01E-01
20	4	110580949	rs138933533	A/G	0.95	1.26	2.17E-05	0.96	1.10	5.00E-01
21	4	125452372	rs11098765	T/C	0.51	1.12	1.70E-07	0.52	1.07	2.31E-01
22	4	125457123	rs11937064	A/G	0.67	0.87	8.73E-10	0.66	0.96	4.98E-01
23	4	160181374	rs11100212	T/C	0.89	1.18	1.68E-05	0.90	1.52	7.16E-05
24	4	185244318	rs793885	T/C	0.41	0.90	2.43E-06	0.40	1.03	5.97E-01
25	4	185265401	rs793880	A/G	0.45	1.11	5.40E-06	0.45	0.98	7.74E-01
26	5	1295349	rs2853669	A/G	0.63	1.12	2.57E-06	0.65	1.10	9.31E-02
27	5	3201714	rs12374504	A/G	0.21	0.90	4.77E-05	0.21	1.02	7.48E-01
28	5	22095055	rs4701568	T/C	0.05	1.25	1.12E-05	0.05	1.29	2.09E-02
29	5	40622940	rs4957138	A/G	0.52	1.10	1.88E-05	0.50	1.00	9.88E-01
30	5	40623638	rs114080964	T/G	0.14	1.18	1.02E-07	0.12	1.14	8.77E-02
31	5	40671099	rs4432939	A/G	0.76	1.11	2.59E-05	0.75	1.15	3.51E-02
32	5	40761422	rs78955876	A/G	0.06	0.81	4.72E-05	0.07	0.87	2.28E-01
33	5	40779705	rs77651534	A/G	0.28	0.79	3.46E-21	0.28	0.93	2.46E-01
34	5	40790551	rs10074991	A/G	0.51	0.78	3.11E-30	0.52	0.85	1.89E-03
35	5	40805847	rs863241	T/G	0.70	0.83	6.19E-15	0.71	0.95	3.76E-01
36	5	40805979	rs36012714	A/G	0.21	1.27	1.48E-19	0.21	1.19	5.02E-03
37	5	40960278	rs2675982	T/C	0.09	1.18	4.75E-06	0.08	1.08	4.03E-01
38	5	149546426	rs2302274	A/G	0.51	0.91	2.69E-05	0.48	0.95	3.90E-01
39	5	149592833	rs6873984	A/G	0.38	1.10	1.71E-05	0.39	1.14	1.73E-02
40	5	152404629	rs62398780	T/G	0.32	1.10	2.68E-05	0.32	1.03	5.65E-01
41	5	157949553	rs2963438	T/C	0.22	1.11	3.48E-05	0.24	0.97	6.66E-01
42	6	4918883	rs7738459	A/G	0.89	0.87	2.34E-05	0.88	1.02	8.11E-01
43	6	16814729	rs1407534	T/C	0.36	0.91	2.98E-05	0.35	0.92	1.25E-01
44	6	27310533	rs9461366	A/G	0.15	1.15	2.46E-06	0.15	1.05	4.97E-01
45	6	28140293	rs16893741	T/C	0.85	0.87	2.35E-06	0.82	0.97	6.69E-01
46	6	28270584	rs2799081	T/C	0.69	1.11	4.22E-06	0.72	1.09	1.74E-01
47	6	28962143	rs56158159	A/C	0.14	0.88	1.66E-05	0.13	0.91	2.56E-01
48	6	29442975	rs77454196	A/C	0.05	0.78	2.78E-06	0.05	1.13	3.10E-01
49	6	29543646	rs362525	T/G	0.76	0.90	2.83E-05	0.76	1.08	2.28E-01
50	6	29690669	rs9258183	A/G	0.47	1.10	9.69E-06	0.40	1.00	9.94E-01
51	6	29905452	rs28749114	A/G	0.60	1.14	5.41E-09	0.65	1.11	5.95E-02
52	6	29913759	rs9260219	A/G	0.44	1.11	1.24E-06	0.47	1.08	1.59E-01
53	6	30571471	rs1264423	T/C	0.45	0.91	4.68E-06	0.43	0.93	1.55E-01
54	6	30578335	rs2267637	T/C	0.31	1.11	2.50E-06	0.35	1.05	4.17E-01
55	6	31357916	rs3128983	T/C	0.62	0.90	3.82E-06	0.64	1.03	5.42E-01
56	6	40984262	rs9381024	T/G	0.23	1.14	1.24E-07	0.25	1.17	1.00E-02
57	7	6567114	rs1105664	A/G	0.57	0.91	1.59E-05	0.55	1.08	1.69E-01
58	7	18022260	rs11560253	A/G	0.75	1.12	4.30E-06	0.77	1.00	9.80E-01
59	7	33953879	rs10243847	T/C	0.14	0.88	4.10E-05	0.12	1.03	7.52E-01
60	7	135804243	rs1386780	A/G	0.96	0.80	4.19E-05	0.95	1.04	7.57E-01
61	7	141230901	rs1527643	A/G	0.30	0.90	1.24E-05	0.30	1.00	9.80E-01
62	8	2958945	rs76332192	A/G	0.15	0.88	3.03E-05	0.16	0.97	6.63E-01
63	8	3486575	rs9644346	A/G	0.53	0.91	3.38E-05	0.54	1.03	5.23E-01
64	8	12706911	rs9644015	A/G	0.13	0.86	1.19E-05	0.13	0.94	4.82E-01
65	8	36663518	rs11775036	T/C	0.20	0.88	7.02E-06	0.18	0.91	2.05E-01
66	8	54434997	rs13439440	A/G	0.17	1.13	2.24E-05	0.17	1.02	7.55E-01
67	8	143671615	rs10105842	T/C	0.68	0.87	4.15E-09	0.66	0.92	1.62E-01
68	8	143680772	rs66671632	T/C	0.55	0.89	9.98E-07	0.54	0.88	2.25E-02
69	8	143776717	rs2585177	A/C	0.84	0.80	2.05E-13	0.84	0.80	1.57E-03
70	8	143814554	rs10087426	A/G	0.22	1.14	4.24E-07	0.24	1.05	4.79E-01
71	9	2177492	rs10491697	A/G	0.62	1.11	5.02E-06	0.64	1.06	2.68E-01

5

72	9	6043593	rs343471	T/C	0.83	1.15	2.96E-06	0.83	0.93	3.29E-01
73	9	37390630	rs10814529	A/G	0.17	0.88	1.85E-05	0.17	0.98	7.57E-01
74	9	74377685	rs13300174	A/G	0.66	1.10	4.60E-05	0.67	0.96	4.83E-01
75	9	136146597	rs550057	T/C	0.23	1.11	1.50E-05	0.22	1.19	5.11E-03
76	10	35948710	rs618443	A/G	0.19	0.89	1.22E-05	0.17	1.03	6.38E-01
77	10	37592538	rs7100025	A/G	0.76	1.11	1.58E-05	0.77	0.96	5.73E-01
78	10	37802532	rs2475206	A/G	0.50	0.91	7.71E-06	0.49	1.03	6.35E-01
79	10	96038686	rs11187838	A/G	0.56	1.12	3.56E-07	0.56	1.10	7.06E-02
80	10	96067947	rs10509670	A/G	0.79	0.78	1.78E-22	0.78	0.82	1.43E-03
81	10	96141697	rs11591693	A/G	0.68	1.11	1.08E-05	0.69	1.08	1.72E-01
82	10	96456685	rs56083530	A/G	0.90	0.86	3.13E-05	0.91	0.82	2.30E-02
83	11	95288142	rs11021216	A/G	0.15	1.13	2.37E-05	0.17	1.05	4.78E-01
84	11	95293681	rs59316824	T/C	0.03	1.28	3.96E-05	0.04	1.18	1.84E-01
85	11	128592032	rs644571	A/G	0.30	1.10	1.98E-05	0.32	0.91	1.25E-01
86	12	51117593	rs11169530	A/C	0.07	1.19	4.41E-05	0.07	1.08	4.75E-01
87	12	59205963	rs11172733	T/C	0.84	0.87	1.15E-06	0.84	0.93	3.49E-01
88	12	62076515	rs117949478	T/C	0.07	1.19	4.71E-05	0.08	0.99	9.39E-01
89	13	24887434	rs9551092	T/C	0.27	1.12	1.04E-05	0.25	1.02	7.58E-01
90	13	59565064	rs1856810	T/C	0.24	1.11	4.79E-05	0.25	0.99	8.42E-01
91	13	66733074	rs9540669	A/G	0.64	1.10	2.64E-05	0.63	0.97	5.23E-01
92	13	80244121	rs61291112	T/C	0.09	1.19	8.94E-06	0.11	0.91	3.35E-01
93	13	99855313	rs9517648	A/G	0.49	0.91	1.39E-05	0.47	0.94	2.49E-01
94	13	99968151	rs9517683	A/G	0.46	1.10	1.22E-05	0.44	1.08	1.39E-01
95	14	36477954	rs11851309	T/C	0.38	0.89	2.65E-07	0.35	1.03	5.76E-01
96	14	39171481	rs56758829	T/C	0.87	0.88	4.26E-05	0.85	0.94	4.20E-01
97	15	40795198	rs999197	T/C	0.11	1.18	1.08E-05	0.11	1.04	6.04E-01
98	15	69361169	rs311915	A/G	0.96	1.28	2.35E-05	0.96	1.33	7.74E-02
99	16	11023868	rs4072864	A/C	0.31	0.91	4.18E-05	0.30	0.97	6.34E-01
100	16	19580621	rs117430025	A/G	0.09	0.82	9.38E-07	0.08	1.15	1.29E-01
101	16	48536809	rs7206781	T/C	0.09	1.16	3.47E-05	0.11	1.00	9.88E-01
102	17	32005754	rs9906617	A/G	0.86	0.88	3.54E-05	0.84	1.03	7.11E-01
103	17	36077863	rs12951345	A/C	0.65	0.91	2.05E-05	0.64	0.90	4.58E-02
104	18	21561790	rs1258143	A/G	0.28	1.10	4.74E-05	0.29	1.06	3.01E-01
105	19	42142877	rs141851899	T/C	0.10	1.16	1.38E-05	0.09	0.95	6.09E-01
106	20	4339724	rs77070411	A/G	0.13	0.87	2.88E-05	0.11	1.08	3.74E-01
107	20	10079396	rs6039695	A/G	0.34	0.89	1.41E-06	0.33	0.95	3.59E-01
108	21	43867892	rs35735840	A/G	0.83	1.13	3.07E-05	0.84	1.00	9.97E-01
109	21	47928792	rs11089073	T/C	0.26	1.11	1.61E-05	0.25	1.12	5.51E-02
110	22	26382534	rs6004874	A/G	0.08	1.18	4.62E-05	0.08	0.99	9.06E-01
111	22	31355354	rs2285918	A/G	0.72	0.91	2.83E-05	0.73	1.05	4.14E-01
112	22	33751863	rs2283895	T/C	0.40	0.90	2.53E-05	0.38	1.05	4.17E-01

[a] Derived from the meta-analysis of six GWAS datasets.

[b] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

Abbreviations: CKB, China Kadoorie Biobank; EAF, effect allele frequency; GWAS, genome-wide association study; HR, hazards ratio; OR, odds ratio; SNP, single nucleotide polymorphism.

Table S4. Association results of polygenic risk score with characteristics of participants and lifestyle factors

Variable	Polygenic risk score (per SD) [a]	
	OR (95% CI)	*P* value
Sex		
Male	Ref.	
Female	1.00 (1.00-1.00)	0.991
Age at baseline, years		
-40	Ref.	
40-50	1.00 (1.00-1.00)	0.914
50-60	1.00 (0.99-1.01)	0.719
60-70	1.00 (1.00-1.00)	0.914
70-	1.01 (0.99-1.04)	0.356
Residential area		
Rural	Ref.	
Urban	0.99 (0.96-1.01)	0.190
Current smoking		
No	Ref.	
Yes	0.99 (0.96-1.01)	0.151
Alcohol drinking		
Never	Ref.	
Ever	1.01 (0.99-1.02)	0.385
Consumption of preserved foods		
Low	Ref.	
High	1.00 (0.99-1.00)	0.695
Intake of fresh fruits and vegetables		
Frequent	Ref.	
Less frequent	0.99 (0.97-1.01)	0.150

[a] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry where appropriate.

Abbreviations: CI, confidence interval; OR, odds ratio; SD, standard deviation.

Table S5. Risk of incident gastric cancer according to quintiles of polygenic risk score

Quintiles of PRS	No. of cases	Person years	Model 1 [a]		Model 2 [b]	
			HR (95% CI)	*P* value	HR (95% CI)	*P* value
1	88	198004	Ref.		Ref.	
2	120	197559	1.35 (1.03-1.78)	3.25E-02	1.36 (1.03-1.79)	3.04E-02
3	141	197399	1.58 (1.21-2.06)	7.96E-04	1.58 (1.21-2.07)	7.68E-04
4	152	197250	1.68 (1.29-2.19)	1.11E-04	1.69 (1.30-2.20)	9.38E-05
5	191	197384	2.09 (1.62-2.69)	1.66E-08	2.10 (1.63-2.71)	1.26E-08
P value for trend				3.25E-09		2.38E-09

[a] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

[b] Adjusted for age, sex, residential area, lifestyle factors and the first 10 principal components of ancestry.

Abbreviations: CI, confidence interval; HR, hazards ratio; PRS, polygenic risk score.

Table S6. Risk of incident gastric cancer according to risk levels of genetic risk

Genetic risk [a]	No. of cases	Person years	Model 1 [b] HR (95% CI)	P value	Model 2 [c] HR (95% CI)	P value
Low	88	198,004	Ref.		Ref.	
Intermediate	413	592,200	1.54 (1.22-1.94)	2.67E-04	1.54 (1.22-1.94)	2.38E-04
High	191	197,384	2.08 (1.61-2.69)	1.83E-08	2.09 (1.62-2.70)	1.38E-08
P value for trend			<0.0001		<0.0001	

[a] Defined by polygenic risk score: low (the bottom quintile), intermediate (quintiles 2-4) and high (the top quintile).

[b] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

[c] Adjusted for age, sex, residential area, lifestyle factors and the first 10 principal components of ancestry.

Abbreviations: CI, confidence interval; HR, hazards ratio.

Table S7. Risk of incident gastric cancer according to different risk levels of polygenic risk score (PRS)

Genetic risk	No. of cases	Person years	Model 1 [a]		Model 2 [b]	
			HR(95% CI)	*P* value	HR(95% CI)	*P* value
Based on quartiles of PRS [c]						
Low	123	247,358	Ref.		Ref.	
Intermediate	339	493,176	1.36 (1.11-1.67)	3.58E-03	1.36 (1.11-1.68)	3.58E-03
High	230	247,054	1.81 (1.45-2.26)	1.43E-07	1.82 (1.46-2.28)	1.04E-07
P value for trend			<0.0001		<0.0001	
Based on tertiles of PRS [d]						
Low	171	329,725	Ref.		Ref.	
Intermediate	225	328,439	1.30 (1.07-1.59)	0.009	1.30 (1.07-1.59)	0.009
High	296	329,424	1.68 (1.39-2.03)	1.03E-07	1.69 (1.39-2.04)	7.87E-08
P value for trend			<0.0001		<0.0001	

[a] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

[b] Adjusted for age, sex, residential area, the first 10 principal components of ancestry and lifestyle factors.

[c] Defined by quartiles of PRS: low (the bottom quartile), intermediate (quartiles 2-3) and high (the top quartile).

[d] Defined by tertiles of PRS: low (the lowest tertiles), intermediate (the mid tertiles) and high (the highest tertiles).

Abbreviations: CI, confidence interval; HR, hazards ratio; PRS, polygenic risk score.

10

Table S8. Risk of incident gastric cancer according to risk levels of polygenic risk score (PRS) derived from different significance thresholds of GWAS meta-analysis

Genetic risk [a]	No. of cases	Person years	Model 1 [b]		Model 2 [c]	
			HR(95%CI)	P value	HR(95%CI)	P value
Based on quintiles of PRS derived from 12 SNPs with $P<5\times10^{-8}$						
Low	105	198,274	Ref.		Ref.	
Intermediate	406	591,706	1.26 (1.02-1.56)	0.035	1.26 (1.02-1.56)	0.035
High	181	197,608	1.65 (1.30-2.11)	4.82E-05	1.66 (1.30-2.12)	4.06E-05
P value for trend			<0.0001		<0.0001	
Based on quintiles of PRS derived from 18 SNPs with $P<5\times10^{-7}$						
Low	103	198,182	Ref.		Ref.	
Intermediate	397	591,799	1.27 (1.02-1.57)	0.034	1.27 (1.02-1.58)	0.032
High	192	197,607	1.81 (1.42-2.30)	1.53E-06	1.81 (1.43-2.31)	1.31E-06
P value for trend			<0.0001		<0.0001	
Based on quintiles of PRS derived from 38 SNPs with $P<5\times10^{-6}$						
Low	87	198,297	Ref.		Ref.	
Intermediate	419	591,476	1.61 (1.27-2.02)	6.36E-05	1.61 (1.27-2.03)	6.10E-05
High	186	197,814	2.09 (1.62-2.71)	1.71E-08	2.12 (1.64-2.74)	1.11E-08
P value for trend			<0.0001		<0.0001	
Based on quintiles of PRS derived from 539 SNPs with $P<5\times10^{-4}$						
Low	110	197,965	Ref.		Ref.	
Intermediate	410	592,434	1.24 (1.01-1.54)	0.043	1.24 (1.01-1.54)	0.044
High	172	197,189	1.55 (1.22-1.97)	3.70E-04	1.55 (1.22-1.97)	3.90E-04
P value for trend			0.00030		0.00031	

[a] Defined by polygenic risk score: low (the bottom quintile), intermediate (quintiles 2-4) and high (the top quintile).

[b] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

[c] Adjusted for age, sex, residential area, the first 10 principal components of ancestry and lifestyle factors.

Abbreviations: CI, confidence interval; GWAS, genome-wide association study; HR, hazards ratio; PRS, polygenic risk score; SNP, single nucleotide polymorphism.

Table S9. Risk of incident gastric cancer according to risk levels of genetic risk after excluding incident cases during the first year of follow-up

Genetic risk [a]	No. of cases	Person years	Model 1 [b]		Model 2 [c]	
			HR (95% CI)	*P* value	HR (95% CI)	*P* value
Low	82	198,001	Ref.		Ref.	
Intermediate	383	592,185	1.53 (1.21-1.95)	4.86E-04	1.54 (1.21-1.95)	4.30E-04
High	181	197,381	2.12 (1.63-2.76)	2.42E-08	2.13 (1.64-2.78)	1.79E-08
P value for trend			<0.0001		<0.0001	

[a] Defined by polygenic risk score: low (the bottom quintile), intermediate (quintiles 2-4) and high (the top quintile).

[b] Adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

[c] Adjusted for age, sex, residential area, lifestyle factors and the first 10 principal components of ancestry.

Abbreviations: CI, confidence interval; HR, hazards ratio.

Table S10. Incident gastric cancer by baseline variables of participants in the CKB cohort

Variable	Incident gastric cancer (n=692)		No gastric cancer (n=99,528)		Incidence rate (per 100,000 person years)
	No.	%	No.	%	
Smoking					
Never	338	48·84	66,326	66·64	50.5
Former	66	22·61	4,759	4·78	148.34
Current	288	41·62	28,443	28·58	105.17
Pack-Year	30.20±20.94		27.48±20.65		
-10	54	7·80	6,309	6·34	87.83
10-20	67	9·68	7,708	7·74	89.44
20-30	77	11·13	6,967	7·00	112.97
30-	156	22·54	12,218	12·28	137.11
Alcohol drinking					
Never	448	64·74	76,577	76·94	58.7
Former	55	7·95	4,548	4·57	131.39
Current	189	27·31	18,403	18·49	103.47
Daily Amount (g)	55.92±40.87		52.17±37.49		
<15	13	1·88	1,100	1·10	119.23
15-30	38	5·49	3,565	3·58	107.52
30-60	49	7·08	4,871	4·89	101.54
60-	58	8·38	5,764	5·79	101.7
Consumption of preserved foods					
Daily	169	24·42	15,995	16·07	104.02
4-6 days per week	55	7·95	6,857	6·89	82.08
1-3 days per week	164	23·7	28,187	28·32	59.09
Monthly	170	24·57	30,323	30·47	56.41
Never/rarely	134	19·36	18,166	18·25	74.76
Intake of fresh fruits					
Daily	102	14·74	19,012	19·1	52.67
4-6 days per week	44	6·36	8,917	8·96	48.41
1-3 days per week	208	30·06	31,089	31·24	66.82
Monthly	264	38·15	33,472	33·63	81.12
Never/rarely	74	10·69	7,038	7·07	111.48
Intake of fresh vegetables					
Daily	653	94·36	94,127	94·57	69.71
4-6 days per week	29	4·19	3,752	3·77	81.06
1-3 days per week	8	1·16	1,353	1·36	63.75
Monthly	2	0·29	272	0·27	85.03
Never/rarely	0	0	24	0·02	–
Peptic ulcers					
Yes	47	6·79	3,942	3·96	118.55
No	645	93·21	95,586	96·04	68.04
COPD					
Yes	110	15·9	9,531	9·58	162.22
No	582	84·1	89,997	90·42	64.63
Diabetes					
Yes	37	5·34	3,661	3·68	110.62
No	655	64·66	95,867	96·32	68.64

Table S11. Risk of incident gastric cancer according to lifestyle categories in the CKB cohort

Lifestyle category [a]	No. of cases	Person years	Model 1 [b]		Model 2 [c]	
			Hazards ratio (95%CI)	*P* value	Hazards ratio (95%CI)	*P* value
Favorable (n=13,576)	52	139,479	Ref.		Ref.	
Intermediate (n=68,270)	408	672,521	1.34 (1.00-1.81)	0.052	1.34 (0.99-1.80)	0.054
Unfavorable (n=18,374)	232	175,587	2.03 (1.46-2.83)	<0.0001	2.02 (1.45-2.82)	<0.0001
P value for trend			<0.0001		<0.0001	

[a] Participants were divided into favorable (4 healthy lifestyle factors), intermediate (2 or 3 healthy lifestyle factors), or unfavorable (0 or 1 healthy lifestyle factor) according to the number of healthy lifestyle factors.

[b] Model 1: Cox proportional hazards regression adjusted for age, gender, residential area, and the first 10 principal components of ancestry; *P* value for trend calculated treating the lifestyle as a ordinal variable.

[c] Model 2: Cox proportional hazards regression adjusted for Model 1, as well as polygenic risk score; *P* value for trend calculated treating the lifestyle as a ordinal variable.

Table S12. Risk of incident gastric cancer according to redefined lifestyle categories in the CKB cohort [a]

Lifestyle category [b]	No. of cases	Person years	Model 1 [c]		Model 2 [d]	
			Hazards ratio (95% CI)	P value	Hazards ratio (95% CI)	P value
Favorable (n=9,157)	37	94,088	Ref.		Ref.	
Intermediate (n=71,466)	414	705,813	1.32 (0.93-1.86)	0.121	1.33 (0.94-1.88)	0.109
Unfavorable (n=19,597)	241	187,686	1.97 (1.35-2.88)	0.00042	1.99 (1.37-2.91)	0.00034
P value for trend			<0.0001		<0.0001	

[a] The frequent intake of fresh fruits and vegetables was redefined as eating vegetables and fruits very day.

[b] Participants were divided into favorable (4 healthy lifestyle factors), intermediate (2 or 3 healthy lifestyle factors), or unfavorable (0 or 1 healthy lifestyle factor) according to the number of healthy lifestyle factors.

[c] Model 1: Cox proportional hazards regression adjusted for age, gender, residential area, and the first 10 principal components of ancestry; P value for trend calculated treating the lifestyle as a ordinal variable.

[d] Model 2: Cox proportional hazards regression adjusted for Model 1, as well as polygenic risk score; P value for trend calculated treating the lifestyle as a ordinal variable.

Table S13. Risk of incident gastric cancer according to lifestyle categories within each genetic risk category based on quartiles or tertiles of polygenic risk score [a]

Lifestyle category	Low genetic risk			Intermediate genetic risk			High genetic risk		
	Unfavorable	Intermediate	Favorable	Unfavorable	Intermediate	Favorable	Unfavorable	Intermediate	Favorable
Based on quartiles of PRS [b]									
No. of cases/Person years	41/42,861	75/169,192	7/35,305	120/89,014	193/335,452	26/68,710	71/43,712	140/167,877	19/35,465
HR (95%CI)	Ref.	0.64 (0.41-0.99)	0.37 (0.15-0.88)	Ref.	0.58 (0.45-0.76)	0.47 (0.29-0.75)	Ref.	0.76 (0.55-1.04)	0.59 (0.33-1.03)
P value		0.047	0.025		<0.001	<0.001		0.089	0.062
P value for trend		0.011			<0.001			0.037	
Absolute risk (‰) -10 years (95% CI)	9.92 (6.84-13.00)	4.56 (3.50-5.61)	2.02 (0.52-3.53)	13.84 (11.33-16.35)	5.88 (5.03-6.73)	3.85 (2.37-5.34)	15.33 (11.68-18.98)	7.83 (6.47-9.19)	5.00 (2.73-7.26)
Absolute risk reduction (‰)-10 years (95% CI)	Ref.	5.36 (2.13-8.52)	7.90 (4.37-11.44)	Ref.	7.96 (5.29-10.49)	9.99 (7.01-12.64)	Ref.	7.50 (3.54-11.14)	10.33 (6.20-14.48)
Numbers needed -10 years [d]		186	127		126	100		133	97
Based on tertiles of PRS [c]									
No. of cases/Person years	56/58,039	103/224,839	12/46,847	79/58,929	130/223,467	16/46,043	97/58,619	175/224,216	24/46,589

16

HR (95%CI)	Ref.	0.64 (0.44-0.94)	0.46 (0.23-0.92)	Ref.	0.64 (0.46-0.88)	0.47 (0.26-0.85)	Ref.	0.66 (0.50-0.87)	0.52 (0.32-0.85)
P value		0.021	0.029		0.007	0.013		0.004	0.009
P value for trend		0.009			0.003			0.002	
Absolute risk (‰) -10 years (95% CI)	9.85 (7.23-12.47)	4.64 (3.72-5.56)	2.58 (1.11-4.04)	14.04 (10.91-17.17)	6.07 (5.01-7.13)	3.62 (1.84-5.39)	15.78 (12.56-19.00)	7.41 (6.26-8.55)	4.86 (2.90-6.81)
Absolute risk reduction (‰)-10 years (95% CI)	Ref.	5.21 (2.35-7.93)	7.27 (4.36-10.35)	Ref.	7.97 (4.47-11.17)	10.42 (6.56-14.15)	Ref.	8.37 (4.89-11.54)	10.92 (7.20-14.81)
Numbers needed -10 years [d]		192	137		125	96		119	92

[a] Cox proportional hazards regression adjusted for age, sex, residential area, and the first 10 principal components of ancestry.

[b] Defined by quartiles of PRS: low (the bottom quartile), intermediate (quartiles 2-4) and high (the top quartile).

[c] Defined by tertiles of PRS: low (the lowest tertiles), intermediate (the mid tertiles) and high (the highest tertiles).

[d] The number needed to adhere to a healthy lifestyle to prevent one gastric cancer case in 10 years.

Abbreviations: CI, confidence interval; HR, hazards ratio; PRS, polygenic risk score.

Figure S1. Workflow chart for the procedures of GWAS meta-analysis

Figure S2. Population structure assessed by top two principal components of ancestry between cases and controls for each study. (A) East-GWAS; (B) North-GWAS; (C) BJ-GWAS; (D) NJ-GWAS; (E) Onco-GWAS; (F) SX-GWAS. Consecutive plots show the first two PCs of samples from each study and each point indicates an individual. Patients with gastric cancer are labeled as cases in brown, whereas the controls are labeled in green.
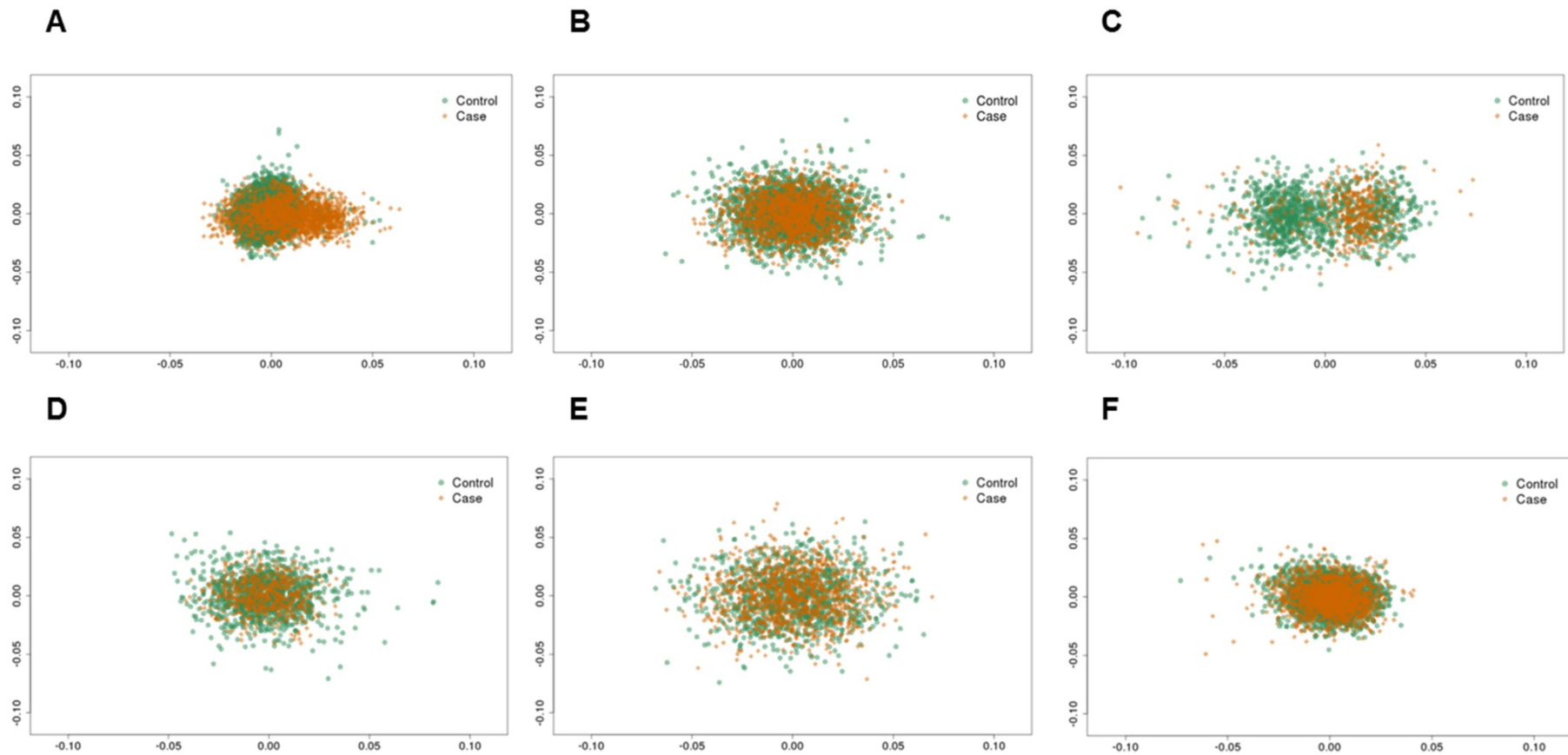
Figure S3. Quantile-quantile plot and genomic inflation factor lambda for associations with gastric cancer risk in each study and the meta-analysis of six studies. Observed *P* values are plotted as a function of theoretical *P* values. This figure shows the Q-Q plot for the two-sided *P*-values obtained from each GWAS study and the meta-analysis. The X axis shows the expected distribution of the observed -$\log_{10}$(*P*-values) under the null hypothesis of no association. The Y axis shows the distribution of the observed -$\log_{10}$(*P*-values) of each study. The black indicator line shows where X=Y. Gray areas indicate 95% confidence intervals from a null distribution of *P* values. Genomic inflation values (lambda) for East-GWAS and meta-analysis (with sample size larger than 2,000 cases and 2,000 controls) were adjusted as equal to a study of 1,000 cases and 1,000 controls. The genomic inflation values (lambda) we calculated indicate that the results are not affected by model mis-specification. A high inflation factor might indicate presence of population stratification, unknown familial relationships, undetected sample duplications, technical problems with the data, or application of incorrect statistical methods.
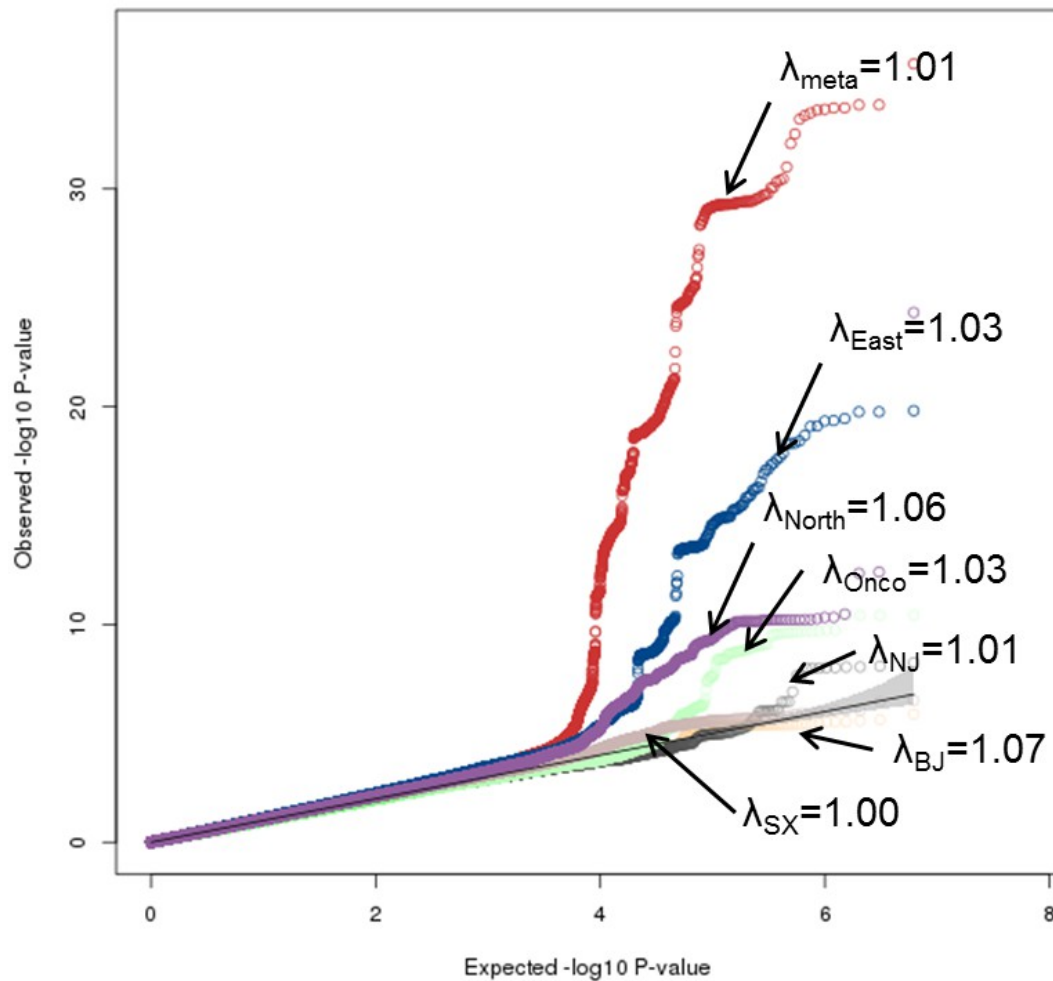
Figure S4. Manhattan plots from the individual GWAS dataset of gastric cancer. The associations (-log10 (*P*) values, Y-axis) are plotted against genomic position (X-axis by chromosome and the chromosomal position). The red horizontal line corresponds to a *P* value of 5.0×10^{-5}. (A) East-GWAS; (B) North-GWAS; (C) BJ-GWAS; (D) NJ-GWAS; (E) Onco-GWAS; (F) SX-GWAS.
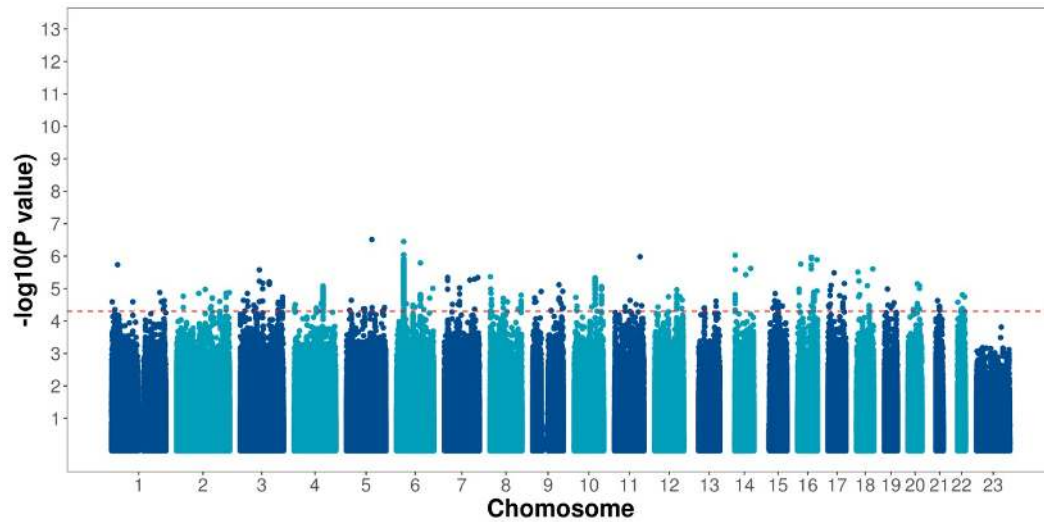
A: East-GWAS
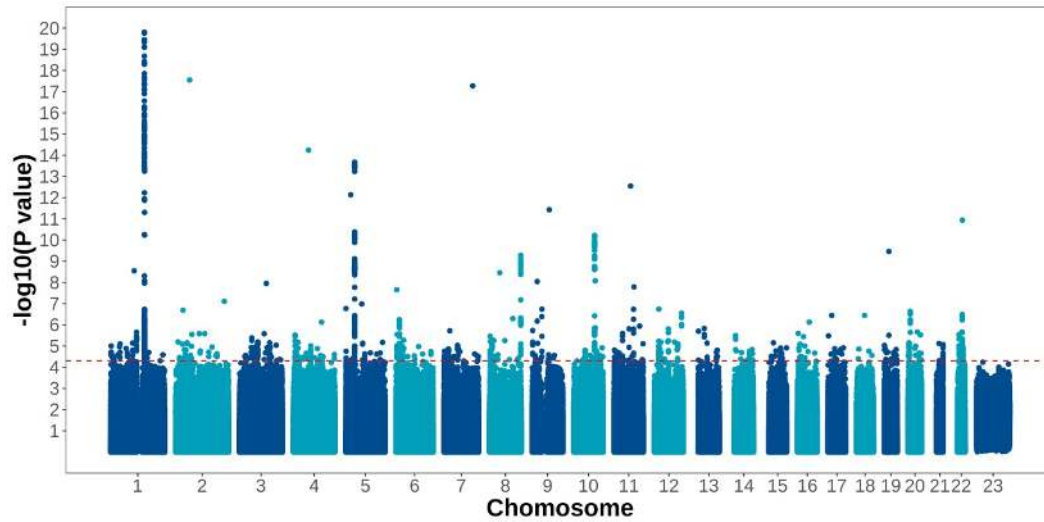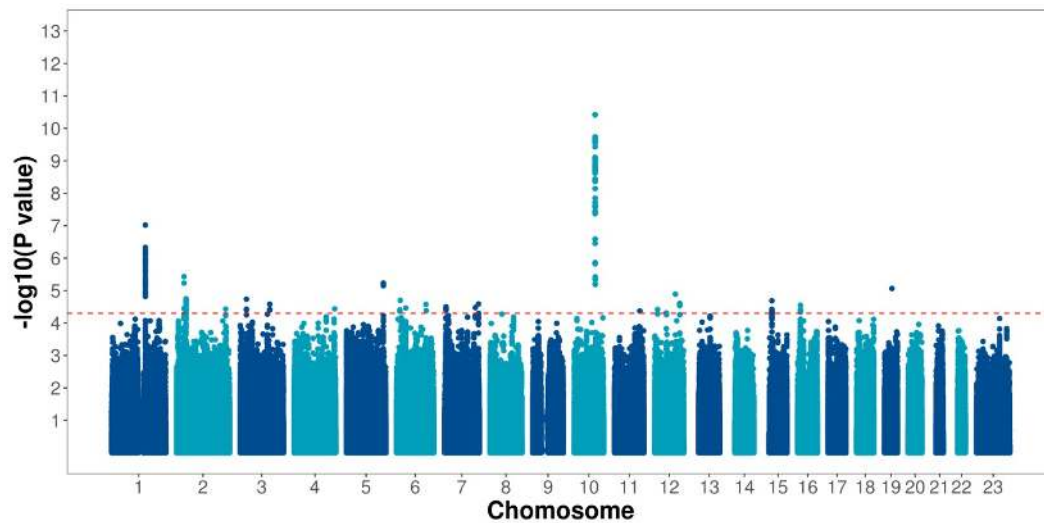

B: North-GWAS
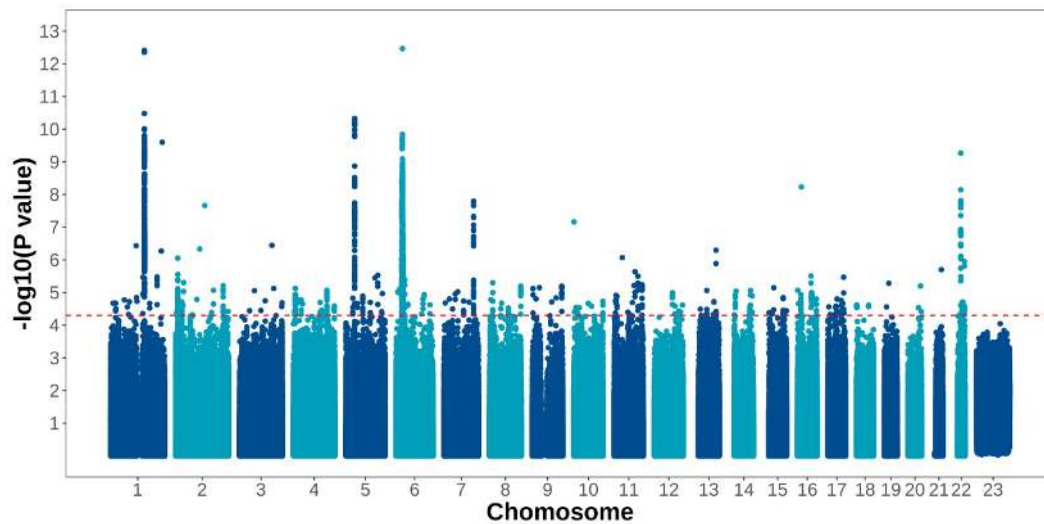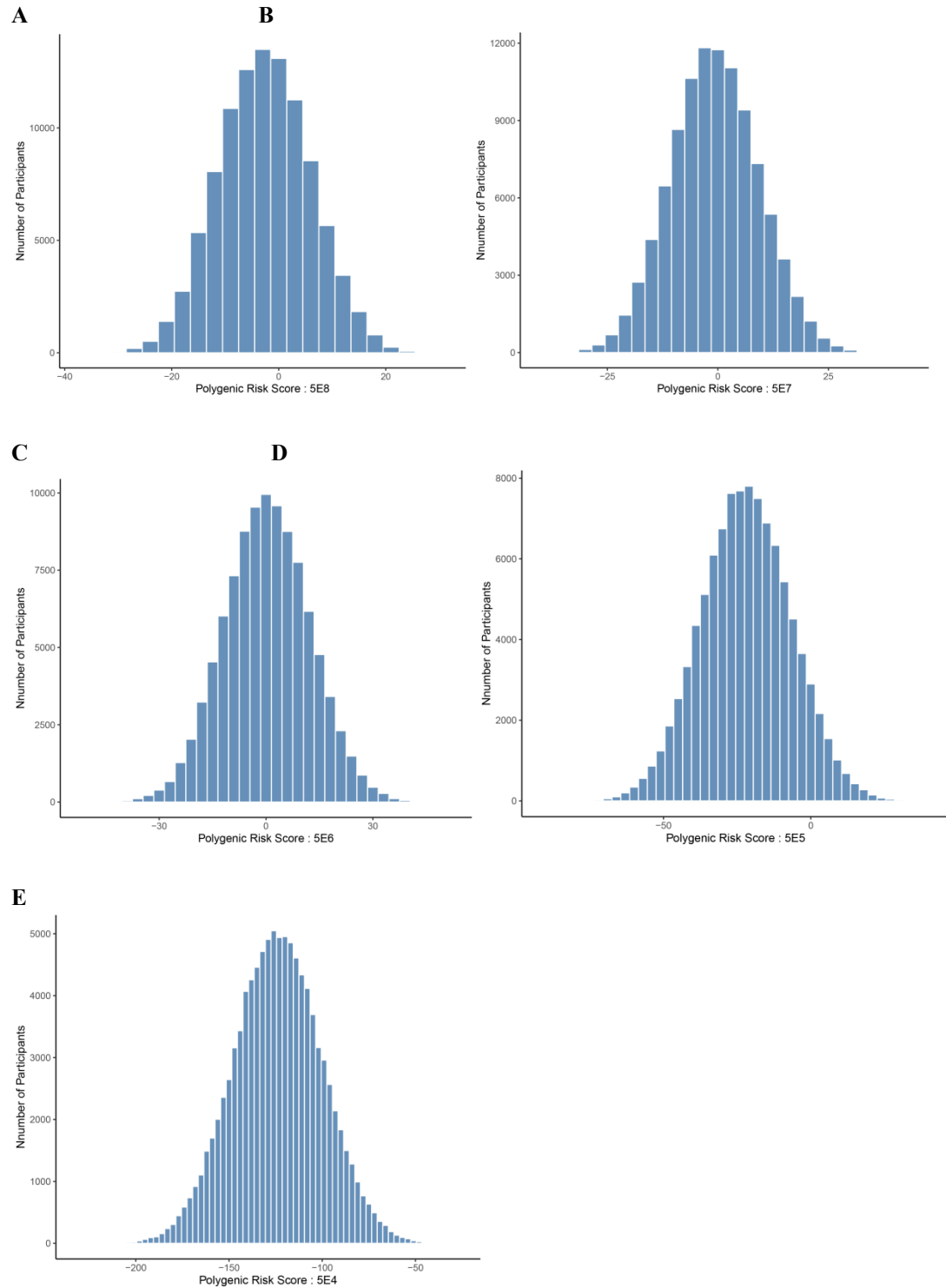

C: BJ-GWAS
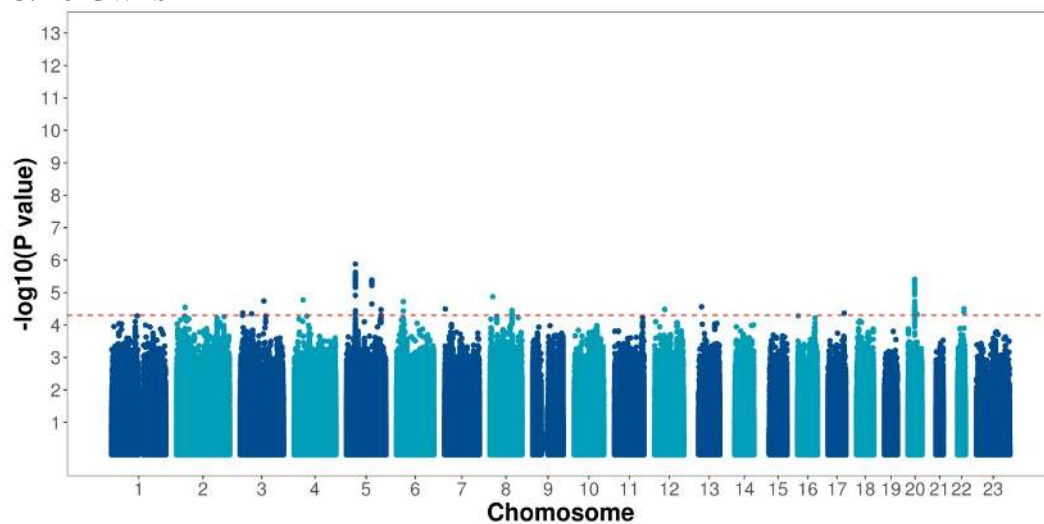
D: NJ-GWAS
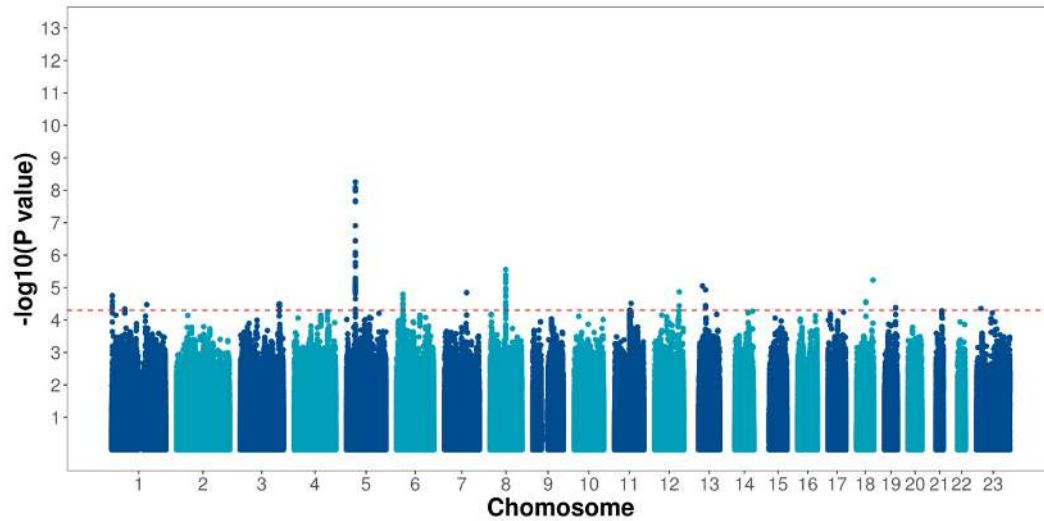
E: Onco-GWAS

F: SX-GWAS

Figure S5. Manhattan plot for meta-analysis of gastric cancer GWASs. The associations (-log10 (*P*) values, Y-axis) are plotted against genomic position (X-axis by chromosome and the chromosomal position). The red horizontal line corresponds to a *P*-value of 5.0×10^{-8} and the black horizontal line corresponds to a *P*-value of 5.0×10^{-5}.
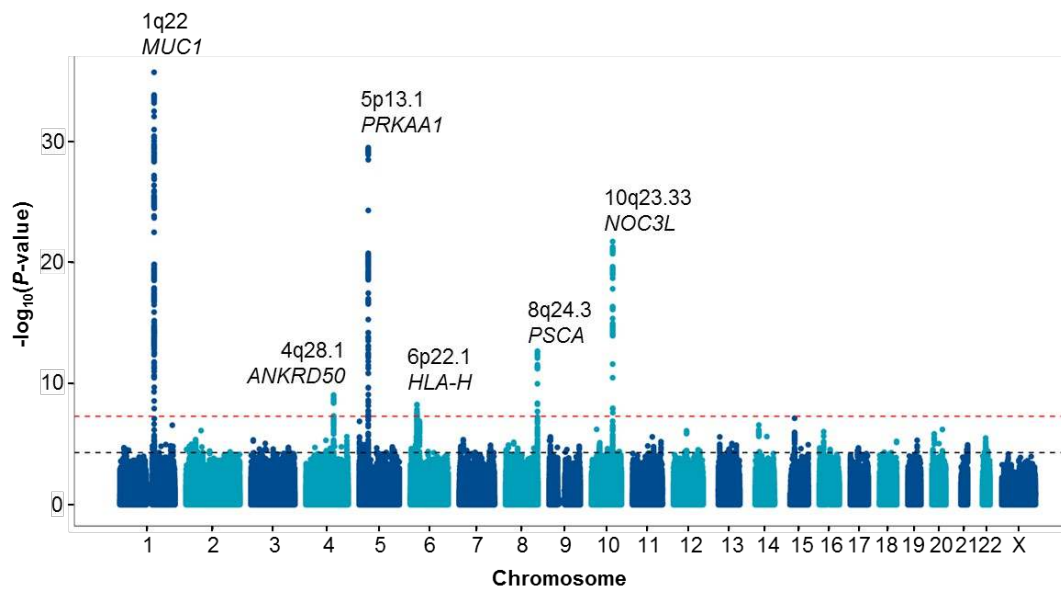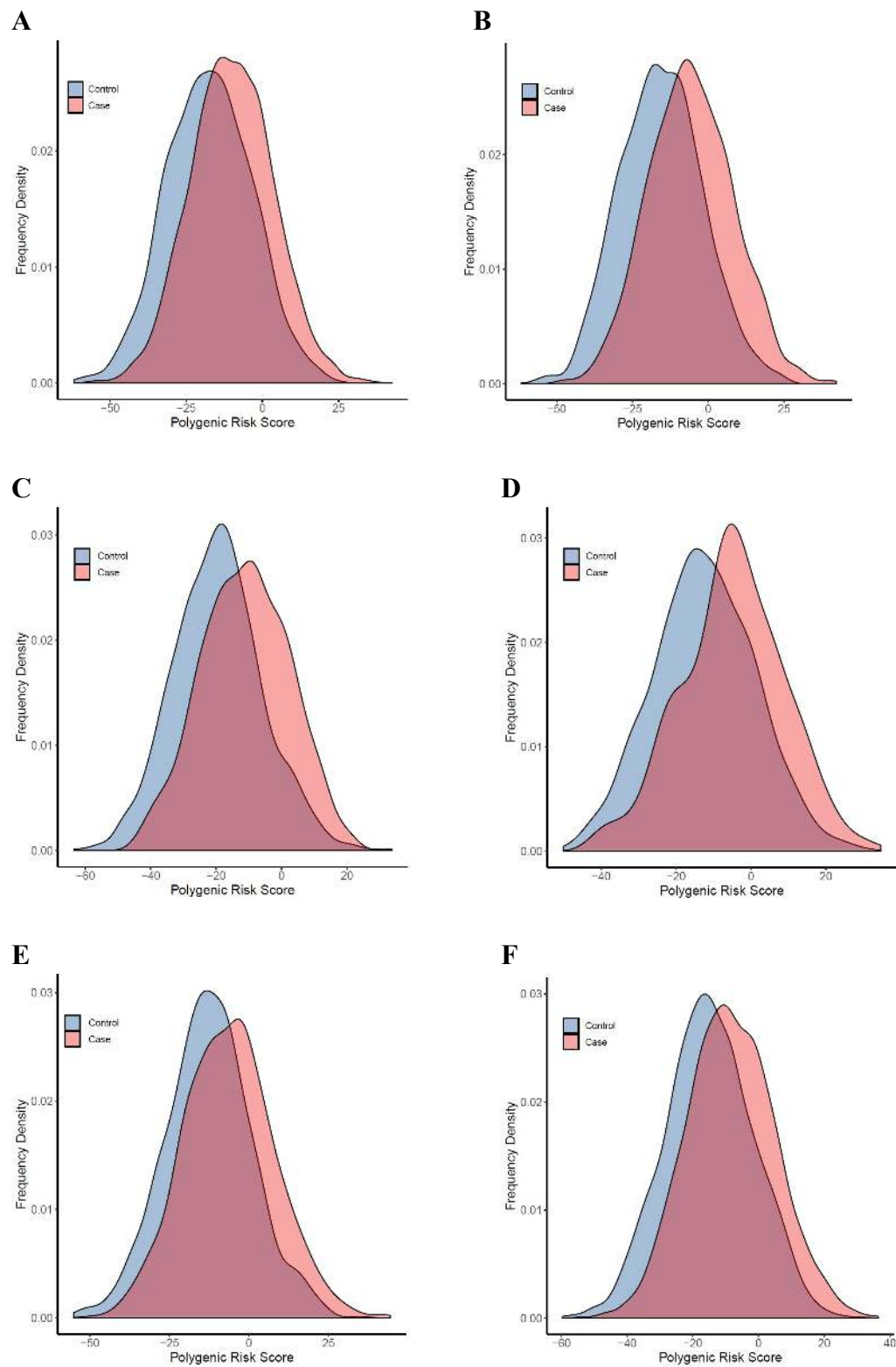
Figure S6. Distribution of polygenic risk score in the CKB cohort. The frequency (Y-axis) is plotted against polygenic risk score (X-axis). (A) $P \leq 5 \times 10^{-8}$; (B) $P \leq 5 \times 10^{-7}$; (C) $P \leq 5 \times 10^{-6}$; (D) $P \leq 5 \times 10^{-5}$; (E) $P \leq 5 \times 10^{-4}$. The blue bar corresponds to number of participants for each category of polygenic risk score in the CKB cohort.

Figure S7. The distribution of polygenic risk scores between cases and controls of GWASs. (A) East-GWAS; (B) North-GWAS; (C) BJ-GWAS; (D) NJ-GWAS; (E) Onco-GWAS; (F) SX-GWAS.

Figure S8. The distribution and relationship of polygenic risk score (PRS) with incident gastric cancer in the China Kadoorie Biobank (CKB) cohort. (A) Distribution of PRS in participants affected with or without gastric cancer of CKB cohort; (B) Linear relationship between PRS and gastric cancer risk was assessed using a restricted cubic spline analysis, and hazards ratios (HRs) were estimated with adjustment for age at risk, sex, residential area, and the first 10 principal components of ancestry.
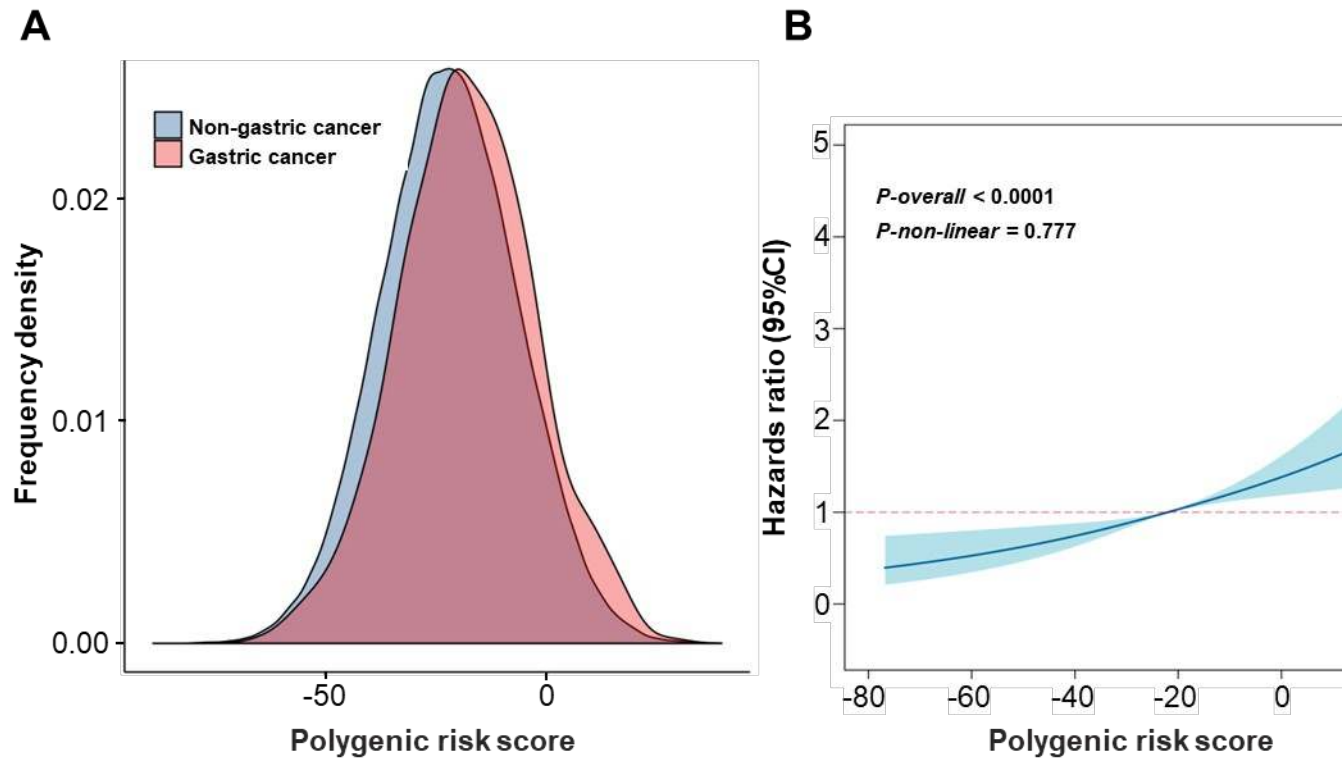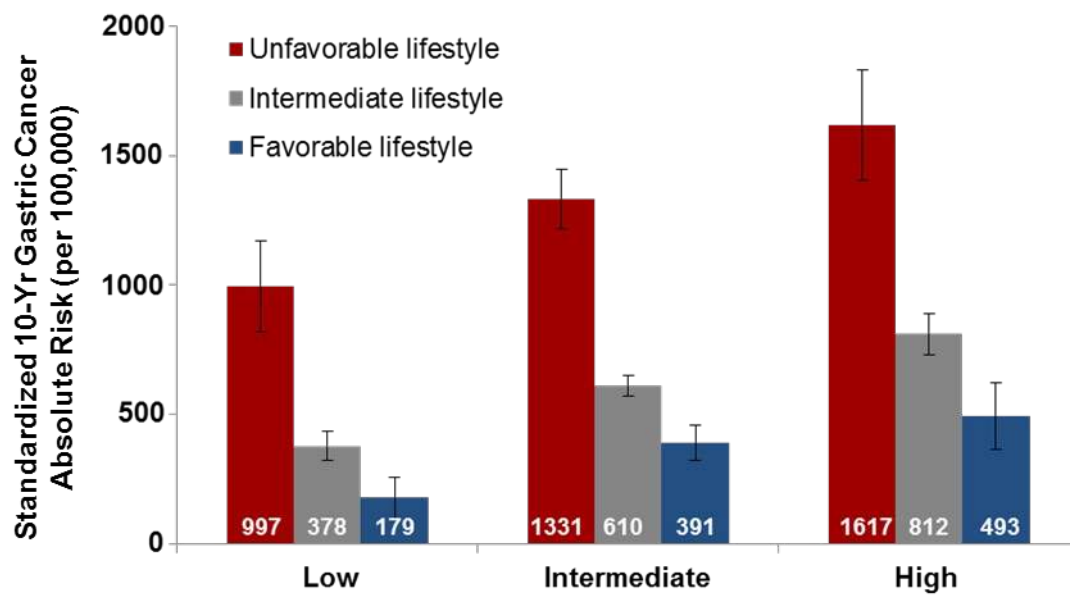
Figure S9. Ten-year gastric cancer absolute risk, according to lifestyle and genetic risk in the China Kadoorie Biobank (CKB) cohort. The bars represent 95% confidence intervals.

References

1. Yan C, Zhu M, Ding Y, et al. Meta-analysis of genome-wide association studies and functional assays decipher susceptibility genes for gastric cancer in Chinese populations. *Gut*. 2020;**69(4)**:641-651.

2. 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**:1215-8.

3. 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**:764-7.

4. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2011;**9**:179-81.

5. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;**5**:e1000529.

6. Gilpin E, Cavin SW, Pierce JP. Adult smokers who do not smoke daily. *Addiction*. 1997;**92(4)**:473-480.