

DR. SHUAI YUAN (Orcid ID : 0000-0001-5055-5627)

DR. SUSANNA LARSSON (Orcid ID : 0000-0003-0118-0341)

Article type : Letter to the Editor

Assessing the protective role of allergic disease in gastrointestinal tract cancers using Mendelian randomization analysis

Running head: Allergic disease and gastrointestinal tract cancer

Shuai Yuan¹, Mathew Vithayathil², Siddhartha Kar³, Paul Carter⁴, Amy M. Mason^{5,6}, Shao-Hua Xie⁷, Stephen Burgess^{4,8}, Susanna C. Larsson^{1,9*}

¹ Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

² MRC Cancer Unit, University of Cambridge, Cambridge, United Kingdom

³ MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK

⁴ Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁵ British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

⁶ National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals, Cambridge, UK

⁷ Upper Gastrointestinal Surgery, Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

⁸ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

⁹ Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

***Corresponding author:**

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ALL.14616](https://doi.org/10.1111/ALL.14616)

This article is protected by copyright. All rights reserved

32 Susanna C. Larsson, PhD, Associate professor
33 Institute of Environmental Medicine, Karolinska Institutet, Nobelsväg 13, Stockholm, 17177,
34 Sweden; Department of Surgical Sciences, Uppsala University, Dag Hammarskjölds Väg 14B,
35 Uppsala, 75185, Sweden.
36 Tel: 46-852486059; E-mail: susanna.larsson@ki.se

37 To the Editor,

38

39 Immune hypersensitivity featured among allergic individuals has been proposed to enhance
40 immune surveillance, thereby inhibiting abnormal cell growth and thus reduce cancer risk.^{1,2}

41 Similarly, the prophylaxis theory suggests that allergy symptoms may prevent cancer development
42 by removing potential carcinogens.² This hypothesis envisages a preventive role specifically for
43 cancers of tissues that interface with the external environment, such as gastrointestinal tract
44 cancers.² A reduced risk of gastrointestinal tract cancer among individuals with self-reported
45 allergic conditions has been observed in several but not all observational studies.^{3,4} However,
46 whether the associations are causal remains unclear as observational studies are prone to residual
47 confounding. Therefore, we conducted a Mendelian randomization study⁵ to determine the causal
48 role of genetic liability to allergic disease in esophageal, gastric and colorectal cancers.

49

50 One hundred and thirty-six single-nucleotide polymorphisms (SNPs) associated with at least one
51 allergic disease (asthma, allergic rhinitis or eczema) at $p < 3 \times 10^{-8}$ in 180,129 cases and 180,709
52 non-cases of European ancestry, were considered as instrumental variables.⁶ The 136 SNPs were
53 independent and not in linkage disequilibrium defined by distance $> 1\text{Mb}$ and $r^2 < 0.02$. The SNPs
54 explained around 2.6% of phenotypic variance. Summary-level data for the associations of those
55 SNPs with esophageal, gastric and colorectal cancers were obtained from the UK Biobank,
56 BioBank Japan and FinnGen consortium. The main analysis was based on the multiplicative
57 random-effects inverse-variance weighted method and estimates from the different studies were
58 combined using fixed-effects meta-analysis. The inverse-variance weighted method under a fixed-
59 effects model and the weighted median, MR-Egger regression and MR-PRESSO methods were
60 employed as sensitivity analyses. The I^2 (%) statistic and Cochran's Q value was calculated to
61 assess heterogeneity among estimates of individual SNPs. We considered associations
62 with p values below 0.017 (where $p = 0.05/3$) to represent strong evidence of causal associations,
63 and associations with p values below 0.05 but above 0.017 as suggestive evidence of associations.

64

65 Genetic liability to allergic disease was associated with lower risk of the gastrointestinal tract
66 cancers in the meta-analysis of UK Biobank, BioBank Japan and FinnGen, though the association
67 with gastric cancer was not statistically significant after correcting for multiple testing (Figure 1).
68 The combined odds ratio was 0.85 (95% confidence interval, 0.75, 0.97; $p = 0.016$) for esophageal

69 cancer, 0.90 (95% confidence interval, 0.82, 0.98; $p=0.020$) for gastric cancer, and 0.91 (95%
70 confidence interval, 0.86, 0.96; $p=0.001$) for colorectal cancer for one unit increase in log-
71 transformed odds of allergic disease. The associations were consistent across data sources with the
72 exception of the estimate for gastric cancer in FinnGen which was above one but with a broad
73 confidence interval that overlapped the estimates in UK Biobank and BioBank Japan. Results
74 were consistent in sensitivity analyses for all three cancer outcomes (Table 1).

75
76 A potential limitation arises from the fact that we used an instrument for liability to allergic
77 disease that was developed using genetic data from individuals of European ancestry and applied
78 to a Japanese population. However, a consistent effect of risk alleles of SNPs for allergic disease
79 on asthma in BioBank Japan indicates a high validity of used instrumental variables and a
80 negligible population bias. In addition, a robust pattern of established associations in both Asian
81 and European populations suggested that the associations were portable across populations of
82 distinct ancestries. A further limitation of this work is that we cannot make specific inferences
83 about the impact of food allergies on cancer risk.

84
85 In summary, this MR study supports a protective effect of allergic disease against esophageal and
86 colorectal cancer and possibly gastric cancer. This finding may be useful for further laboratory
87 studies to understand the immunological pathways involved in gastrointestinal tract
88 carcinogenesis.

89 References

- 90 1. Rigoni A, Colombo MP, Pucillo C. Mast cells, basophils and eosinophils: From allergy to
91 cancer. *Semin Immunol.* 2018;35:29-34.
- 92 2. Sherman PW, Holland E, Sherman JS. Allergies: their role in cancer prevention. *Q Rev*
93 *Biol.* 2008;83(4):339-362.
- 94 3. D'Arcy M, Rivera DR, Grothen A, Engels EA. Allergies and the Subsequent Risk of
95 Cancer among Elderly Adults in the United States. *Cancer Epidemiol Biomarkers Prev.*
96 2019;28(4):741-750.
- 97 4. Ma W, Yang J, Li P, Lu X, Cai J. Association between allergic conditions and colorectal
98 cancer risk/mortality: a meta-analysis of prospective studies. *Sci Rep.* 2017;7(1):5589.
- 99 5. Burgess S, Thompson SG. *Mendelian randomization: methods for using genetic variants*
100 *in causal estimation.* CRC Press; 2015.
- 101 6. Ferreira MA, Vonk JM, Baurecht H, et al. Shared genetic origin of asthma, hay fever and
102 eczema elucidates allergic disease biology. *Nat Genet.* 2017;49(12):1752-1757.

103 **Acknowledgement**

104 Genetic instruments for allergic disease were obtained from a published meta-analysis of genome-
105 wide association studies.⁶ Genetic association estimates for esophageal, gastric and colorectal
106 cancers were obtained from the UK Biobank study, BioBank Japan and FinnGen consortium. The
107 authors thank all investigators for sharing these data. Analyses of UK Biobank data were
108 performed under application 29202.

109 **Conflicts of interest**

110 The authors report no conflicts of interest.

111 **Funding information**

112 SK is supported by a Cancer Research UK programme grant, the Integrative Cancer Epidemiology
113 Programme (C18281/A19169). AMM is supported by EC-Innovative Medicines Initiative
114 (BigData@Heart). SB is supported by Sir Henry Dale Fellowship jointly funded by the Wellcome
115 Trust and the Royal Society (204623/Z/16/Z). SCL is supported by research grants from the
116 Swedish Research Council for Health, Working Life and Welfare (Forte; grant no. 2018-00123),
117 the Swedish Research Council (Vetenskapsrådet; grant no. 2019-00977), and the Swedish Heart-
118 Lung Foundation (Hjärt-Lungfonden; grant no. 20190247). This work is supported by funding
119 from the National Institute for Health Research (to SB) [Cambridge Biomedical Research Centre
120 at the Cambridge University Hospitals NHS Foundation Trust] [*]. *The views expressed are
121 those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health
122 and Social Care.

123 **Ethical approval**

124 All studies included in cited genome-wide association studies had approved by a relevant review
125 board. The present MR analyses were approved by the Swedish Ethical Review Authority (2019-
126 02793).

Table 1. Associations of genetic predisposition to allergic disease with esophageal, gastric and colorectal cancer in sensitivity analyses

	Esophageal cancer			Gastric cancer			Colorectal cancer		
	OR	95% CI	<i>p</i>	OR	95% CI	<i>p</i>	OR	95% CI	<i>p</i>
UK Biobank									
IVW-fixed effects method	0.89	0.73, 1.08	0.242	0.88	0.72, 1.09	0.253	0.91	0.84, 0.98	0.015
Weighted median method	0.92	0.69, 1.23	0.571	0.84	0.61, 1.15	0.270	0.96	0.84, 1.08	0.489
MR-Egger regression	1.01	0.6, 1.68	0.976	1.23	0.71, 2.13	0.468	0.91	0.73, 1.15	0.442
MR-PRESSO method ^a	n/a	n/a	n/a	n/a	n/a	n/a	0.90	0.83, 0.98	0.012
Heterogeneity ^b	<i>I</i> ² =0%; Cochrane's Q=114; <i>p</i> _{het} =0.877			<i>I</i> ² =0%; Cochrane's Q=137; <i>p</i> _{het} =0.393			<i>I</i> ² =21%; Cochrane's Q=168; <i>p</i> _{het} =0.021		
Pleiotropy ^c	Intercept=-0.007; <i>p</i> _{pleiotropy} =0.601			Intercept=-0.018; <i>p</i> _{pleiotropy} =0.208			Intercept=0.000; <i>p</i> _{pleiotropy} =0.944		
BioBank Japan									
IVW-fixed effects method	0.79	0.66, 0.96	0.017	0.88	0.81, 0.96	0.004	0.91	0.83, 0.99	0.022
Weighted median method	0.65	0.49, 0.86	0.003	0.94	0.82, 1.07	0.332	0.87	0.76, 0.99	0.034
MR-Egger regression	0.65	0.37, 1.14	0.130	0.83	0.61, 1.13	0.237	0.84	0.66, 1.08	0.181
MR-PRESSO method ^a	n/a	n/a	n/a	0.93	0.84, 1.02	0.138	n/a	n/a	n/a
Heterogeneity ^b	<i>I</i> ² =0%; Cochrane's Q=92; <i>p</i> _{het} =0.833			<i>I</i> ² =31%; Cochrane's Q=154; <i>p</i> _{het} =0.002			<i>I</i> ² =3%; Cochrane's Q=109; <i>p</i> _{het} =0.407		
Pleiotropy ^c	Intercept 0.011; <i>p</i> _{pleiotropy} 0.453			Intercept=0.003; <i>p</i> _{pleiotropy} =0.682			Intercept=0.004; <i>p</i> _{pleiotropy} =0.541		
FinnGen consortium									
IVW-fixed effects method	0.93	0.52, 1.67	0.803	1.11	0.79, 1.56	0.537	0.93	0.79, 1.09	0.352
Weighted median method	0.79	0.34, 1.85	0.586	1.21	0.72, 2.02	0.477	1.11	0.88, 1.4	0.38
MR-Egger regression	0.30	0.06, 1.43	0.129	1.00	0.37, 2.67	0.999	0.88	0.55, 1.39	0.578
MR-PRESSO method ^a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Heterogeneity ^b	$I^2=0\%$; Cochrane's Q=81; $p_{het}=0.998$	$I^2 14\%$; Cochrane's Q 140; $p_{het} 0.103$	$I^2 16\%$; Cochrane's Q 143; $p_{het} 0.078$
Pleiotropy ^c	Intercept=0.061; $p_{pleiotropy}=0.127$	Intercept 0.006; $p_{pleiotropy} 0.818$	Intercept 0.003; $p_{pleiotropy} 0.797$

^a There was no corrected estimate (n/a) if no outlier was identified. One outlier was identified and removed in the analysis of colorectal cancer based on UK Biobank (p value for the distortion test was 0.815) and two outliers were identified and removed in the analysis of gastric cancer based on Biobank Japan (p value for the distortion test was 0.137). ^b We observed significant heterogeneity among used SNPs in the analysis of colorectal cancer based on UK Biobank and gastric cancer based on Biobank Japan ($p_{het}<0.05$). ^c Pleiotropy was measured by the intercept from MR-Egger regression. Pleiotropy was not detected in any analysis (all $p>0.05$).

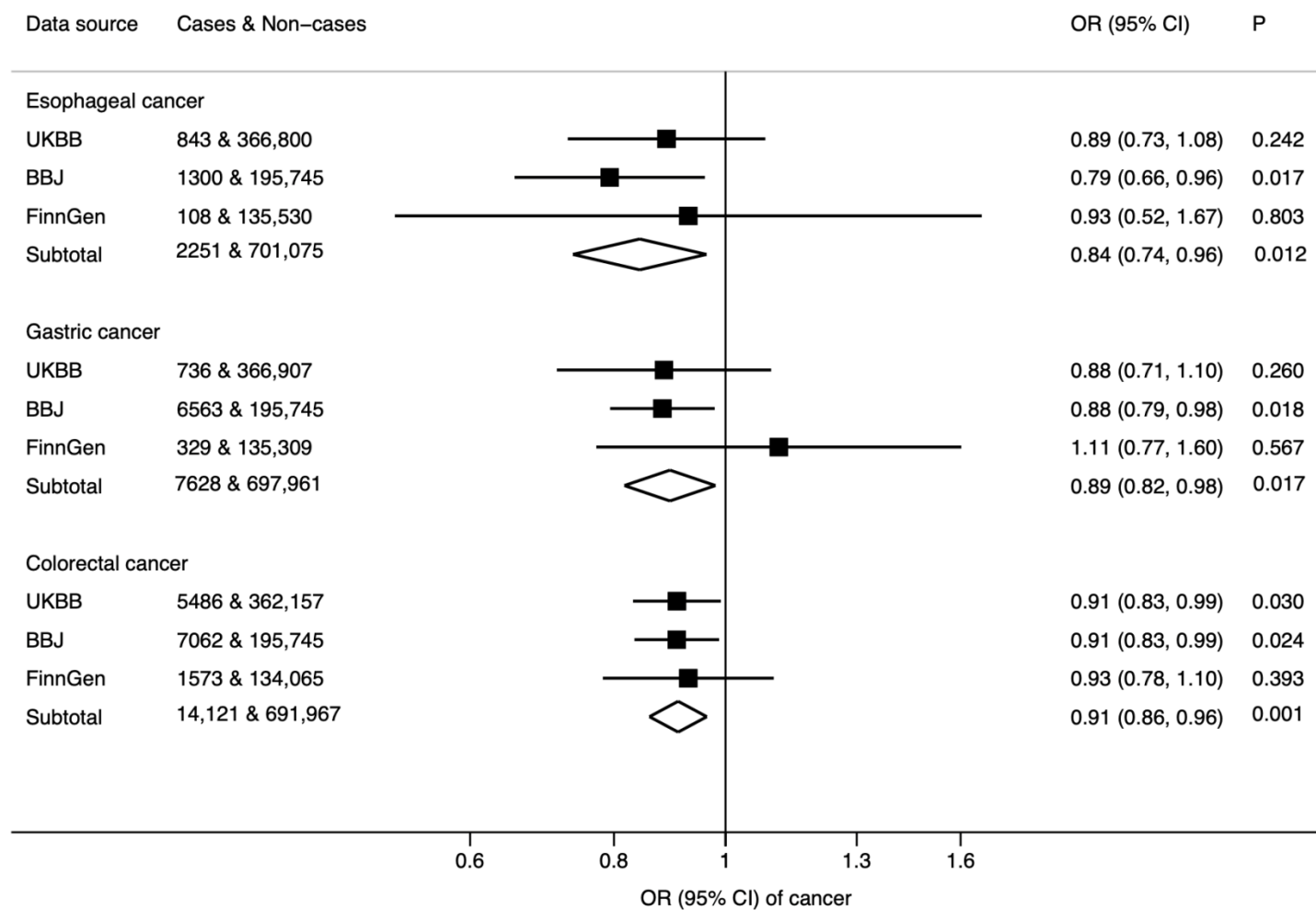


Figure 1. Associations of genetic predisposition to allergic disease with esophageal, gastric and colorectal cancer. BBJ indicates BioBank Japan; CI, confidence interval; OR, odds ratio. UKBB, UK Biobank.

Study hypothesis

A reduced risk of gastrointestinal tract cancer among individuals with self-reported allergic conditions, such as asthma, allergic rhinitis, and eczema, has been found in several but not all observational studies.^{1,2} Nevertheless, due to potential methodological limitations, such as residual confounding and reverse causality, embedded in observational studies, it remains inconclusive whether allergic disease exerts a causal effect on gastrointestinal tract cancer. For example, infection of *helicobacter pylori*, alcohol consumption, obesity and socioeconomic status showed associations with both allergy³ and certain types of gastrointestinal tract cancers.^{4,5} Given that few studies took all possible confounders into consideration, the observed inverse association between allergic disease and gastrointestinal tract cancer could be skewed by the unobserved confounding. A clear appraisal of the causal link from allergic disease to major gastrointestinal cancers is of great importance not only for building accurate risk prediction models to potentially discriminate high-risk individuals from the general population, but also shedding light on the etiology of these cancers, thereby facilitating prevention and treatment strategies.

A brief introduction to Mendelian randomization

Through the use of genetic variants as instruments for an exposure (e.g. genetic liability to allergic disease), the Mendelian randomization (MR) analysis can strengthen the causal inference for an exposure-outcome association by minimizing residual confounding and bias due to reverse causality.^{6,7} The rationale for minimization of residual confounding is that genetic variants are randomly allocated at conception, and therefore, genetically proxied exposures are generally not correlated with other confounding traits. The MR design can also minimize potential bias due to reverse causality because alleles are fixed at birth and cannot be modified by the onset or progression of the disease.^{6,7} Therefore, if a genetic variant robustly associated with an exposure is also associated with an outcome and the effect on the outcome is not stronger than that on the exposure, this offers evidence that the exposure exerts causal effects on the outcome but not *vice versa*. While allergic disease is in itself a disorder, genetic liability to allergic disease can be considered an exposure in the MR setting.

Detailed information on genome-wide association studies on allergic disease

One hundred and thirty-six single-nucleotide polymorphisms associated with at least one allergic disease (asthma, allergic rhinitis or eczema) at $p < 3 \times 10^{-8}$ in a meta-analysis of 13 genome-wide association studies including 180,129 cases (83,827 cases with one allergic disease, 30,418 cases with two allergic diseases, 6,418 cases with three allergic diseases and 59,466 unclassifiable cases) and 180,709 non-cases of European ancestry were considered as instrumental variables.⁸ The analyses had been adjusted for age, sex and population principal components. The mean age of participants in the meta-analysis was around 52 years and there were 52% women. The 136 SNPs were independent and not in linkage disequilibrium defined by distance $> 1\text{Mb}$ and $r^2 < 0.02$. All SNPs together explained around 2.6% of phenotypic variance. To evaluate the applicability of the instrumental variables for allergic disease identified in individuals of European descent for the Japanese population, we examined the effect of the genetic instrument on risk of asthma using data from BioBank Japan. A significant elevated risk of asthma (OR 2.38; 95% CI, 2.06, 2.74; $p = 1.1 \times 10^{-32}$) among Japanese individuals with risk alleles of allergic disease indicated the validity of the genetic instrument.

Detailed information on data sources of cancers

UK Biobank

The UK Biobank study recruited around 500,000 adults aged 40 to 69 years in 2006-2010.⁹ In the present analysis, we included 367,643 individuals of European ancestry after excluding related individuals (third-degree relatives or closer), low call rate, and excess heterozygosity (3 or more standard deviations from the mean). Participants were followed up until 31 March 2017 or the date of death (recorded until 14 February 2018) in the present study with a median follow-up of around 8.0 years. Cancer outcomes were defined based on registry data and self-reported information validated by nurse interview. We calculated genetic association estimates for cancer (beta coefficients and corresponding standard errors) using logistic regression with adjustment for age, sex and the first ten genetic principal components.

BioBank Japan

We obtained summary-level estimates from BioBank Japan and the FinnGen consortium. BioBank Japan is a genetic study with a total of 212,453 Japanese individuals.¹⁰ Individuals with a sample call rate of <0.98 and outliers from East Asian clusters and genetic variants with call rate <99%, p value for Hardy–Weinberg equilibrium $<1.0 \times 10^{-6}$, and fewer than five heterozygotes were excluded in the quality control stage of genome-wide association analyses. All samples in 1KG Phase3 (version 5) were used as the reference population in imputation stage and analyses were adjusted for age, sex and the first five genetic principal components.

The FinnGen consortium

The FinnGen consortium encompasses 135,638 Finnish individuals in the last publicly available (R3) data release. The consortium excluded participants with ambiguous gender, high genotype missingness (>5%), excess heterozygosity (± 4 standard deviation) and non-Finnish ancestry and genetic variants with high missingness (>2%), low Hardy-Weinberg equilibrium p -value ($p < 5 \times 10^{-6}$) and minor allele count <3. The population specific SISu v3 reference panel was used in the imputation stage and genome-wide association tests adjusted for age, sex, the first ten genetic principal components, and genotyping batch. Detailed information about FinnGen is available on its website (<https://www.finnngen.fi/>).

Detailed description of statistical analysis

The main analysis was based on the multiplicative random-effects inverse-variance weighted method,¹¹ and estimates from the different studies were combined using fixed-effects meta-analysis. The inverse-variance weighted method under a fixed-effects model and the weighted median,¹² MR-Egger regression¹³ and MR-PRESSO¹⁴ methods were employed as sensitivity analyses to examine the robustness of results and correct for potential pleiotropy. The F-statistic was estimated to measure the strength of the genetic instrument. The I^2 (%) statistic¹⁵ and Cochrane's Q value was calculated to assess heterogeneity among estimates of individual SNPs. Odds ratios (ORs) and corresponding confidence intervals (CIs) of cancer were scaled to one-unit increase in log-transformed odds of allergic disease. To account for multiple testing, we considered associations with p values below 0.017 (where $p = 0.05/3$ (3 cancer outcomes)) to represent strong evidence of causal associations, and associations with p values below 0.05 but above 0.017 as suggestive evidence of associations. All analyses were two-sided and performed using the `mrrobust` package¹⁶ in Stata/SE 15.0 (Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.), and the MRPRESSO¹⁴ and

TwoSampleMR¹⁷ packages in R Software 3.6.0 (R Core Team. R Foundation for Statistical Computing. Vienna, Austria. 2019. <https://www.R-project.org>).

Supplementary results

The number of SNPs available as genetic instruments for allergic condition in the UK Biobank, BioBank Japan and FinnGen datasets was 134, 110 and 121, respectively. The F-statistic for individual SNPs was over 10 (ranging from 32 to 296) and the average F-statistic was around 71. Results of sensitivity analyses are shown in Table 1. Overall, results were consistent in sensitivity analyses for all three cancer outcomes. We observed significant heterogeneity among estimates across SNPs in the analysis of colorectal cancer based on UK Biobank and gastric cancer based on BioBank Japan ($p_{\text{het}} < 0.05$). However, no pleiotropy was detected by MR-Egger regression (all p for intercept > 0.05). One outlier SNP was identified in the analysis of colorectal cancer based on UK Biobank and two outlier SNPs were identified in the analysis of gastric cancer based on BioBank Japan. Outlier-corrected associations persisted and the p values for two distortion tests were over 0.05, indicating no significant difference between estimates derived from analysis before and after removal of outlier SNPs.

Discussion on findings

The present MR study found support for causal inverse associations between genetic liability to allergic disease and risk of esophageal and colorectal cancers using data from three independent populations. Genetic liability to allergic disease also showed a suggestive inverse association with risk of gastric cancer. Our findings are overall consistent with results of most traditional observational studies. A systematic review study that comprehensively summarized the association between allergies and site-specific cancers found that individuals with any type of allergy had a lower risk of esophageal, gastric and colorectal cancer.¹⁸ The associations were subsequently replicated in studies with a large sample size,¹ especially for colorectal cancer.^{2,19,20} In a meta-analysis of 12 studies including 515,379 individuals and 10,345 colorectal cancer cases, the relative risk for the association between allergic disease and colorectal cancer was 0.88 (95% CI 0.83, 0.92).² However, with regard to esophageal and gastric cancer, a hospital-based cohort study with 92,986 asthma patients found a modest increased risk of these cancers among asthma patients compared with the general population, but the excessive risks were caused by a higher prevalence of gastro-esophageal reflux among asthma patients.²¹ The discrepancy in results for esophageal cancer might be caused by different effects of allergic disease on esophageal squamous cell carcinoma and esophageal adenocarcinoma. The protective effect of allergy may be stronger for or restricted to esophageal squamous cell carcinoma.¹ The stronger association observed between allergic disease and esophageal cancer in BioBank Japan than in the UK Biobank and FinnGen consortium in the present study may also support this hypothesis considering the higher prevalence of esophageal squamous cell carcinoma in the Japanese population than in European populations.²²

Two major hypotheses have been proposed to explain the link between allergic disease and reduced risk of gastrointestinal tract cancers. The mainstay one is the immunosurveillance hypothesis assuming that individuals with allergic disease own a generally enhanced immune responsiveness, thereby having the propensity to detect and eradicate dysregulated cells. In detail, atopic allergens can enter human body via several systems at numerous points, including the respiratory system and the digestive system,

where mast cells, basophils and eosinophils are omnipresent²³. Type I immunoglobulin E-mediated hypersensitivity reactions will be activated by this process and so does an excessive T-helper cell type II response due to the overstimulation of eosinophils and mast cells. Enhanced permeability and inflammation after mast cells degranulating, immunoglobulin E antibody-dependent cellular cytotoxicity in precancerous and cancerous cells and direct anticancer effects may help clear dysregulated cells and limit the growth of abnormal cells^{24,25}. In addition, an increased level of tumor associated tissue eosinophils has been showed associations with lower rates of venous invasion, lymph node metastases and better clinic outcome in esophageal squamous cell carcinoma²⁶. The other hypothesis is the prophylaxis hypothesis, which suggests that the physical effect of allergic responses and inflammation in mucosal surfaces removes mutagenic triggers, such as sneezing or coughing out the microbial carcinogen Aflatoxin B1 produced by the fungus *Aspergillus flavus* before the start of malignant transformation²⁴.

There are several strengths and limitations of the present study. The major merit is the MR design which strengthens the causal inference in the associations of allergic disease with three major gastrointestinal tract cancers. The associations for esophageal and colorectal cancer were coherent in three independent population, thereby providing strong support for that the findings for these site-specific cancers are causal. Moreover, a high consistency across MR analyses indicated a negligible distortion by potential pleiotropy. However, the finding for gastric cancer needs verification considering the somewhat inconclusive results across studies. A part of asthma cases was not allergic but caused by other factors, which might have affected our findings. However, only around 5.6% of cases were diagnosed with only asthma.⁸ This means that the vast majority of recruited cases had allergic conditions and the estimates of used SNPs could appropriately represent the genetic effects of used instruments on allergic conditions. The influence of allergic conditions on cancer risk may differ for allergic rhinitis, asthma, and eczema¹. Whether the observed associations in this study were driven by a certain allergic disease needs to be investigated. Nevertheless, given a high comorbidity of different allergic diseases⁸, an overall causal effect revealed by our finding is of great significance in the public health and clinical setting. A further limitation of this work is that we cannot make specific inferences about the impact of food allergies on cancer risk. Whether food allergy has protective effect as allergic diseases on these cancers is of great interest and deserves examination in future studies.

References

1. D'Arcy M, Rivera DR, Grothen A, Engels EA. Allergies and the Subsequent Risk of Cancer among Elderly Adults in the United States. *Cancer Epidemiol Biomarkers Prev.* 2019;28(4):741-750.
2. Ma W, Yang J, Li P, Lu X, Cai J. Association between allergic conditions and colorectal cancer risk/mortality: a meta-analysis of prospective studies. *Sci Rep.* 2017;7(1):5589.
3. Chen Y, Blaser MJ. Inverse associations of *Helicobacter pylori* with asthma and allergy. *Arch Intern Med.* 2007;167(8):821-827.
4. Peek RM, Jr., Blaser MJ. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer.* 2002;2(1):28-37.
5. Cui Y, Hill AW. Atopy and Specific Cancer Sites: a Review of Epidemiological Studies. *Clin Rev Allergy Immunol.* 2016;51(3):338-352.

6. Burgess S, Thompson SG. *Mendelian randomization: methods for using genetic variants in causal estimation*. CRC Press; 2015.
7. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.
8. Ferreira MA, Vonk JM, Baurecht H, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet*. 2017;49(12):1752-1757.
9. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
10. Ishigaki K, Akiyama M, Kanai M, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet*. 2020;52(7):669-679.
11. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology*. 2017;28(1):30-42.
12. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304-314.
13. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525.
14. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693-698.
15. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-1558.
16. Spiller W, Davies NM, Palmer TM. Software application profile: mrrobust—a tool for performing two-sample summary Mendelian randomization analyses. *International Journal of Epidemiology*. 2019;48(3):6.
17. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.
18. Merrill RM, Isakson RT, Beck RE. The association between allergies and cancer: what is currently known? *Ann Allergy Asthma Immunol*. 2007;99(2):102-116; quiz 117-109, 150.
19. Prizment AE, Folsom AR, Cerhan JR, Flood A, Ross JA, Anderson KE. History of allergy and reduced incidence of colorectal cancer, Iowa Women's Health Study. *Cancer Epidemiol Biomarkers Prev*. 2007;16(11):2357-2362.
20. Tambe NA, Wilkens LR, Wan P, et al. Atopic allergic conditions and colorectal cancer risk in the Multiethnic Cohort Study. *Am J Epidemiol*. 2015;181(11):889-897.
21. Ye W, Chow WH, Lagergren J, et al. Risk of adenocarcinomas of the oesophagus and gastric cardia in patients hospitalized for asthma. *Br J Cancer*. 2001;85(9):1317-1321.
22. Wong MCS, Hamilton W, Whiteman DC, et al. Global Incidence and mortality of oesophageal cancer and their correlation with socioeconomic indicators temporal patterns and trends in 41 countries. *Sci Rep*. 2018;8(1):4522.

Supporting information for

Yuan S., et al. Assessing the protective role of allergic disease in gastrointestinal tract cancers using Mendelian randomization analysis

23. Rigoni A, Colombo MP, Pucillo C. Mast cells, basophils and eosinophils: From allergy to cancer. *Semin Immunol.* 2018;35:29-34.
24. Josephs DH, Spicer JF, Corrigan CJ, Gould HJ, Karagiannis SN. Epidemiological associations of allergy, IgE and cancer. *Clin Exp Allergy.* 2013;43(10):1110-1123.
25. Jensen-Jarolim E, Achatz G, Turner MC, et al. AllergoOncology: the role of IgE-mediated allergy in cancer. *Allergy.* 2008;63(10):1255-1266.
26. Ishibashi S, Ohashi Y, Suzuki T, et al. Tumor-associated tissue eosinophilia in human esophageal squamous cell carcinoma. *Anticancer Res.* 2006;26(2b):1419-1424.

Supplemental Table 1. Diagnosis of esophageal, gastric and colorectal cancers*

Data source	Definition of cancer		
UK Biobank	ICD-9 codes	ICD-10 codes	Self-reported cancer
Esophageal cancer	150, V10.03	C15, Z85.01	1017
Gastric cancer	151, V10.04	C16, Z85.028	1018
Colorectal cancer	153, 154.0, 154.1, V10.05, V10.06	C18, C19, C20, Z85.038, Z85.048	1020, 1022, 1023
FinnGen consortium	ICD-8 codes	ICD-9 codes	ICD-10 codes
Esophageal cancer	150	150	C15
Gastric cancer	151	151	C16
Colorectal cancer	153,1540, 1541	153,154	C18, C19, C20

*This information was not available for BioBank Japan.