Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett’s esophagus

Barrett’s esophagus is an increasingly common disease that is strongly associated with reflux of stomach acid and usually a hiatus hernia, and it strongly predisposes to esophageal adenocarcinoma (EAC), a tumor with a very poor prognosis. We report the first genome-wide association study on Barrett’s esophagus, comprising 1,852 UK cases and 5,172 UK controls in the discovery stage and 5,986 cases and 12,825 controls in the replication stage. Variants at two loci were associated with disease risk: chromosome 6p21, rs9257809 (Pcombined = 4.09 × 10−8; odds ratio (OR) = 1.21, 95% confidence interval (CI) = 1.13–1.28), within the major histocompatibility complex locus, and chromosome 16q24, rs9936833 (Pcombined = 2.74 × 10−10; OR = 1.14, 95% CI = 1.10–1.19), for which the closest protein-coding gene is FOXL1, which is implicated in esophageal development and structure. We found evidence that many common variants of small effect contribute to genetic susceptibility to Barrett’s esophagus and that SNP alleles predisposing to obesity also increase risk for Barrett’s esophagus.

Barrett’s esophagus is one of the most common premalignant lesions in the western world. It affects over 2% of the adult population and, unlike bowel polyps, lacks any proven effective therapy. In the majority of cases, Barrett’s esophagus is associated with chronic gastroesophageal reflux disease (GERD), including esophagitis. Over 80% of affected individuals have a hiatus hernia in the lower esophagus that facilitates the reflux of acid and bile into the esophagus. The measured annual risk of EAC in individuals with Barrett’s esophagus varies widely but is approximately 0.4–1% (refs. 5–7). Notably, the incidence of EAC has been rising by 3% each year for the last 30 years; it is now the fifth most common cancer in the UK. Despite modern multimodality therapy, the prognosis for EAC remains poor, with a 9–15% 5-year survival rate.

The etiology of Barrett’s esophagus is not well characterized. Environmental factors, such as diet, are weakly associated with GERD, Barrett’s esophagus and EAC, and obesity is a known risk factor for all three conditions. There is also evidence implicating genetic factors: relative risks are increased by 2- to 4-fold for GERD, Barrett’s esophagus and EAC when one first-degree relative is affected. A segregation analysis of 881 pedigrees of familial Barrett’s esophagus supports an incompletely dominant inheritance model with a polygenic component. Extensive candidate gene and linkage searches have to date been unsuccessful in identifying genetic variants that are associated with risk of Barrett’s esophagus.

As part of the Wellcome Trust Case Control Consortium 2 (WTCCC2) study of 15 common disorders and traits, we present the results of the first genome-wide association study of Barrett’s esophagus susceptibility. Using a discovery cohort from the UK (with case samples from the Aspirin and Esomeprazole Chemoprevention Trial of Cancer in Barrett’s esophagus (AspECT)) and five replication cohorts (including case samples from CHemoprevention Of Premalignant Intestinal Neoplasia (ChOPIN) and Esophageal Adenocarcinoma GenEthics Consortium (EAGLE) studies), we identified two variants associated with Barrett’s esophagus, each with combined evidence at P < 5 × 10−8. The analysis workflow is outlined in Supplementary Figure 1, and characteristics of the case and control samples that were included can be found in the Online Methods and Supplementary Table 1.

For the discovery analysis, cases with histologically confirmed Barrett’s esophagus (Online Methods) were recruited from sites across the UK (Supplementary Table 2). Population controls were taken from the WTCCC2 common set of 1958 Birth Cohort (58C) and National Blood Service (UKBS) samples as previously described.

Figure 1 Plot of the genome-wide association results after fitting the multiplicative model in SNPTEST. Results are shown for the 521,744 SNPs that passed quality control filters. Chromosomes are labeled on the x axis. The y axis shows the −log10 Passociation values. Regions in red show the loci newly identified to be associated with Barrett’s esophagus (described in Table 1).

1A full list of authors and affiliations appears at the end of the paper. 2A full list of members and affiliations is provided in the Supplementary Note.

Received 1 June; accepted 15 August; published online 9 September 2012; doi:10.1038/ng.2408
Cases were genotyped on the Illumina 660W-Quad array, and controls were genotyped on the Illumina custom Human 1.2M-Duo array, with analysis performed on the overlapping set of SNPs. After quality control (Online Methods, Supplementary Fig. 2, Supplementary Table 3 and Supplementary Note), a total of 521,744 SNPs typed in 1,852 cases and 5,172 controls (2,499 UKBS and 2,673 58C) were included in the discovery analysis.

Association analysis was carried out under a logistic regression model implemented in SNPTEST. The genomic overdispersion factor23, $\lambda$, was 1.10, and this was reduced to 1.05 when the first principal component was incorporated as a covariate, suggesting that population structure was not a major confounder in the discovery analyses (Supplementary Fig. 3). For all of the following results presented, unless otherwise stated, the first principal component was used as a covariate.

After analysis of the genome-wide association results (Fig. 1), we adopted a staged approach to replication (outlined below and in Supplementary Fig. 1).

In stage 1, 100 SNPs that showed evidence of association in the discovery data (at $P < 5 \times 10^{-8}$) were analyzed in another UK sample set. This comprised 1,105 cases from ChOPIP and EAGLE and 4,421 controls from the 58C control data set, all of which were genotyped on the Illumina Immunochip23. WTCCC2 contributed SNPs to the Immunochip design to allow for its replication studies, and the set of 100 SNPs followed up in our stage 1 replication were all on the Immunochip, and an additional set of 2,578 UK controls (the People of the British Isles (PoBI) collection)24 genotyped on the Illumina custom Human 1.2M-Duo array. Results of this first stage of replication are shown in Supplementary Table 4.

In stage 2, the 16 top SNPs ($P_{\text{combined}} < 1 \times 10^{-5}$) from meta-analysis of the discovery stage and stage 1 replication were replicated in silico in a Dutch collection of 473 cases and 1,780 controls genotyped on the Immunochip23. Results from stage 2 replication are shown in Supplementary Table 5.

In stage 3, two SNPs with $P_{\text{combined}} < 5 \times 10^{-8}$ after stage 2 replication (rs9257809 on chromosome 6p21 and rs9936833 on chromosome 16q24) were studied in three additional replication sample sets. They were directly genotyped in an Irish cohort of 245 cases and 473 controls and a UK cohort of 1,765 cases and 1,586 controls, and data for these SNPs were retrieved from the Barrett’s Esophagus and Esophageal Adenocarcinoma Consortium (BEACON) for 2,398 cases and 2,167 controls from European, Australian and US individuals with European ancestry.

After these three stages of replication, the two SNPs on chromosomes 6p21 and 16q24 showed compelling evidence for association, with combined $P$ values of $4.09 \times 10^{-9}$ for rs9257809, OR (95% CI) = 1.10–1.21, and $2.74 \times 10^{-10}$ for rs9936833, OR (95% CI) = 1.14–1.19 (Figs. 2 and 3 and Table 1).

We performed tests for pairwise interaction (Supplementary Note) in the discovery data between all pairs of the 16 SNPs taken forward to stage 2 replication (Supplementary Table 5), but no significant interactions ($P < 0.01$) were found.

Imputation was carried out in the discovery data for the chromosome 6p21 and 16q24 regions, using the 1000 Genomes Project June 2010 Utah residents of Northern and Western European ancestry (CEU) reference panel. In each case, rs9257809 and rs9936833 remained the strongest signal of association in the corresponding genomic region (Supplementary Fig. 4).

The lead SNP at 6p21, rs9936833, maps 24 kb away from the spliced, non-coding LOC732275 transcript. The closest coding gene, 141 kb toward the telomere, is FOXF1, a forkhead family transcription factor that acts in the Hedgehog signaling pathway. The FOXF1 protein is known to have a role in development of the gastrointestinal tract and has been reported to cause esophageal structural alterations, especially atresia, when inactivated25. The region around rs9936833 contains multiple binding sites for specific transcription factors, such as FOXP2, that are known to control FOXF1 expression (assessed using Encyclopedia of DNA Elements (ENCODE) data; see URLs).

The lead SNP at 6p21, rs9257809, lies on the telomeric edge of the major histocompatibility complex (MHC) region between olfactory receptor genes OR2D12 and OR2D13.
It is in strong long-range linkage disequilibrium (LD; $r^2 > 0.6$, calculated in the control data) with SNPs over 1 Mb away, including two at which stage 2 replication was attempted: rs13211507 ($P_{\text{combined}} = 8.77 \times 10^{-6}$) and rs9262143 ($P_{\text{combined}} = 2.18 \times 10^{-6}$). When conditioning on rs9257809, no other SNP in the MHC region was associated with significance of $P < 1 \times 10^{-5}$.

To further investigate the SNP signal in the MHC region, we took two approaches: GENECLUSTER, which is a Bayesian tree building method$^{26,27}$, and HLA*IMP, which is a method for imputing classical human leukocyte antigen (HLA) alleles from SNP data$^{28}$. Both methods provided evidence of association in the discovery data for reduced risk conferred by three classical HLA alleles that are in strong LD with each other (HLA-C*07:01, HLA-A*01:01 and HLA-B*08:01) (Supplementary Table 6). However, conditional analysis suggested that rs9257809 better captured the association in our discovery data, and none of these three classical HLA alleles showed an association signal in the replication data ($P > 0.1$; Supplementary Table 6).

We used standard UK criteria, in accordance with the British Society of Gastroenterology, for diagnosis of Barrett’s esophagus. However, some countries use the American College of Gastroenterology criteria that require the presence of intestinal metaplasia for diagnosis with Barrett’s esophagus. In order to investigate the effect of having included non intestinal metaplasia cases, we analyzed the two replicated loci, using only the subset of discovery and replication cases with histological evidence of intestinal metaplasia (86%). Both signals remained significant, with combined evidence across discovery and all stages of replication of $P < 5 \times 10^{-8}$ (Supplementary Table 7a, b).

We also investigated associations with the related quantitative traits of circumferential extent ($C$) and maximal extent ($M$) of the length of the Barrett’s segment. In the discovery cohort, the $C$ measurement was available for 1,744 cases, and the $M$ measurement was available for 1,618 cases. In a linear regression analysis of cases, neither SNP showed evidence of association with $C$ or $M$ status (for rs9936833, $P = 0.63$ and 0.87, respectively; for rs925809, $P = 0.10$ and 0.09, respectively). We then extended the $C$ and $M$ analysis across the genome. No SNP reached $P < 1 \times 10^{-6}$ in the analysis of $C$. One SNP (rs1023313) reached $P < 1 \times 10^{-6}$ in the analysis of $M$, but this association was not confirmed in stage 1 or stage 2 replication (Supplementary Table 8).

There is an established sex bias in susceptibility to Barrett’s esophagus, with men being at greater risk than women$^{3,29}$. The ratio of males to females was 4:1 in our case discovery data. To determine whether there might be sex-specific effects of any predisposition SNPs, we performed a sex-stratified analysis for the 16 SNPs analyzed in stage 2 (Supplementary Table 9). The SNP showing the most evidence for a sex-specific effect from the combined discovery set and stage 1 and 2 replication was rs9257809. The association signal was stronger in males than females (uncorrected $P = 0.01$ for the difference of effects between sexes), corresponding to a male odds ratio of 1.38 (95% CI = 1.25–1.53; $P_{\text{combined}} = 1.71 \times 10^{-10}$) and a female odds ratio of 1.11 (95% CI = 0.95–1.30; $P_{\text{combined}} = 0.19$) (see Supplementary Note for further details). This finding warrants further investigation.

Previous genome-wide association studies of common diseases or phenotypes have found evidence for a model where many common variants of small effect influence risk$^{30,31}$. We looked for these combined effects in Barrett’s esophagus using two methods (Online Methods). First, taking the top $k$ SNPs (for different values of $k$) in independent regions in the discovery data, we performed a sign test to determine whether there was an excess (over the proportion expected under the null of 50%) of SNPs for which the effect was in the same direction in the stage 1 replication data. Second, a disease score test analysis was undertaken, as described by the International

**Table 1. Loci associated with risk of Barrett’s esophagus**

| Chromosome | Locus | Risk Allele | Minor Allele | Discovery UK replication 1 | Replication (5,172 controls) | $\text{Combined OR} \quad 95\% \text{CI}$ |
|------------|-------|-------------|--------------|----------------------------|-----------------------------|------------------------------------------|
| 6p21       | rs9257809 | A            | G            | 0.89                       | 0.91                        | 0.88 ($0.99–0.79$)                  |
| 5p15       | rs9936833 | C            | G            | 0.87                       | 0.87                        | 0.87 ($0.99–0.79$)                  |
| 8p22       | rs1023313 | A            | G            | 0.87                       | 0.87                        | 0.87 ($0.99–0.79$)                  |

Genomic positions are based on NCBI Build 36.
Schizophrenia Consortium. Both methods found evidence of an excess of SNPs that have the same risk allele in both cohorts. The strongest evidence in the sign test was for the top 1,100 SNPs, for which the sign test gave $P_{uncorrected} = 2.30 \times 10^{-3}$ (Supplementary Fig. 5). From the disease score analysis, the strongest evidence was for the top 1,710 SNPs, with $P_{uncorrected} = 7.07 \times 10^{-11}$ (Supplementary Fig. 6). Both analyses thus implicate a large number of common SNPs of small effect in susceptibility to Barrett’s esophagus.

There is a well-established link between Barrett’s esophagus and obesity. To investigate whether this link may in part reflect genetic effects, we repeated the sign test at 40 of the SNPs that have been found to be associated with either body mass index (BMI) or waisthip ratio (WHR) where genotype data or tag SNPs were available in our discovery samples. In our discovery data, a total of 29 out of 40 BMI- and/or WHR-associated SNPs (14 genotyped, 15 tagging; Supplementary Table 10a,b) shared the same risk alleles in Barrett’s esophagus as they did for BMI and/or WHR ($P = 6.42 \times 10^{-3}$).

Our results provide direct evidence that Barrett’s esophagus etiology has a genetic component. Inference of the underlying genes must be undertaken cautiously, especially for the variant (tagged by rs9257809) in the gene-rich MHC region in which LD is long-range and complex. However, the location of the other associated SNP, rs9368333, near FOXL1 suggests a role for structural factors in the esophagus and stomach as disease-predisposing factors, consistent with the fact that changes such as hiatus hernia are known to be strongly associated with Barrett’s esophagus. We also found evidence to show that SNPs associated with body weight measures are more likely than expected by chance to show effects in the same direction in association with Barrett’s esophagus, suggesting that genetic effects may in part underpin the epidemiological observation that BMI is a risk factor for Barrett’s esophagus. Given that Barrett’s esophagus has an accepted status as a precursor lesion, the SNPs that we have identified could also essentially be risk factors for EAC and may give clues as to the biology of both of these important phenotypes.

**Note:** Supplementary information is available in the online version of the paper.

**ACKNOWLEDGMENTS**

We acknowledge the individuals with Barrett’s esophagus who participated in the study and the physicians who helped in recruitment. We thank the many research nurses who also helped recruit, including S. Prew. We thank S. Bertrand, J. Bryant, S.L. Clark, J.S. Conquer, T. Dibling, J.C. Eldred, S. Gamble, C. Hind, A. Wilk, C.R. Stirling and S. Taylor of the Wellcome Trust Sanger Institute’s Sample and Genotyping Facilities for technical assistance. We also thank J. Barrett for design of the Illuminum array. Funding for this study was provided by the WTCCC2 project (085473/Z/08/Z and 085475/Z/08/Z), the UK National Blood Service controls, funded by the Wellcome Trust, and the People of the British Isles collection, funded by the Wellcome Trust. This work was also supported by the Esophageal Adenocarcinoma GenE Consortium incorporating the ChOOPIN project (grant C548/A5675) and Inherited Predisposition of neoplasia analysis of genomic DNA (IPOND) from AspECT grant RS/A13672/1. We acknowledge the Dublin Centre for Clinical Research (funded by the Health Research Board and the Wellcome Trust) for support in sample collection and the Irish Blood Transfusion Service and TCD DNA Biobank for provision of control samples. R.M. is supported by the Science Foundation Ireland grant 09/IN.1/B2640.

**AUTHOR CONTRIBUTIONS**

J.A.Z.J., J.T., L.J.G. and M.N. oversaw cohort collection for the discovery and replication data sets. N.T., N.B., P.B., S. Paterson, C.E., I.P., K.V., Y.A., L.M., P.P., P. Mullins, H.D., K. Koss, D.C., M. Griffin, D.A., H.W., S. Panter, I.S., H.S., A. Dhar, H.M., H.A., M.R., A.T., D. Morris, C.N., R.L., P.I., C.R., K.R., C.M., C.H., D. Monk, G.D., S.W., D.J., M. Gibbons, S.C., N.C., D.G., S.A., P.W., J.D.C., H.B. and J.A.Z.J. recruited over 50 cases to the AspECT and/or ChOOPIN studies. D.T.R., K.P., K.H., P.A., A.M.M. and N.E.G. processed AspECT and/or ChOOPIN samples. The AspECT and ChOOPIN management groups (P.S., A.T.T., P.B., D.J., M.A., C. Brooks, I.H., D.F., R.B., I.B., S.L., S.A., P.W., R.F.H., P. Moayedy, J.D.C., H.B. and J.A.Z.J.) monitored the appropriate use of samples and data from these studies. A.v.d.W., N.T., M.P.P., L.J.V.v.d.L., E.I.K., J.P.H.D., W.H.P., J.V.R., D.P.K., R.M., H.G., H.P., R.B., K. Krishnadat, P.D.S., J.W.P.m.v.B., M.M., R. Petty, R.G. and S.C.C. provided samples as part of the EAGLE Consortium. T.L. was involved in EAGLE project design. The BEACON Consortium (D.C.W., D.L., W.Y., A.H.W., N.C.B., N.J.S., I.J.M., L.B., Y.R., I.J.H., R.Z., D.A.C., H.A.R., B.J.R., M.D.G., G.L., A.G.C., W.-H.C., S.M., W.E. and T.L.V.) provided data on the top 2 lead SNPs for the second replication phase. G.T. and C.W. provided Dutch control samples for the first replication phase. The WTCCC2 DNA, genotyping, data quality control and informatics group (S.D., S.E.H., S.E., E.G., C.L., S.C.P., A.T.G. and L.P.) executed genome-wide association study sample handling, genotyping and quality control. The WTCCC2 data and analysis group (Z.S., A.S., C.C.A.S., G.B., C. Bellenguez, C.F., M.P. and P. Donnelly) led the statistical analyses. C.P., E.S., F.L., P.S. and I.B.C. also undertook statistical analyses. A.S., C.P., I.T., J.A.Z.J., C.C.A.S. and P. Donnelly contributed to writing the manuscript. The WTCCC2 management committee (J.M.B., E.B., M.A.R., J.B.C., A.C., P. Deloukas, P. Donnelly (chairperson), A. Duncanson, J.A.Z.J., H.S.M., C.G.M., C.N.A.P., L.P., R. Plomin, A.R., S.S., R.C.T., A.C.V. and N.W.) monitored the execution of the WTCCC2 project (grant RGAG1G7R); Cancer Research UK (Cancer Research UK (A.S., C.W., C.C.A.S., G.B., C. Bellenguez, C.F., M.P. and P. Donnelly) monitored the execution of the EAGLE project. The BEACON Consortium (D.C.W., C.P., E.S., F.L., P.S. and I.B.C. also undertook statistical analyses. A.S., C.P., I.T., J.A.Z.J., C.C.A.S. and P. Donnelly contributed to writing the manuscript. The WTCCC2 management committee (J.M.B., E.B., M.A.R., J.B.C., A.C., P. Deloukas, P. Donnelly (chairperson), A. Duncanson, J.A.Z.J., H.S.M., C.G.M., C.N.A.P., L.P., R. Plomin, A.R., S.S., R.C.T., A.C.V. and N.W.) monitored the execution of the genome-wide association study. All authors reviewed the final manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

**URLs**

SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; 1000 Genomes Project, http://www.1000genomes.org/; ENCODE, http://genome.ucsc.edu/encode/.

**METHODS**

Methods and any associated references are available in the online version of the paper.

Published online at http://www.nature.com/dofind10.1038/ng.2408. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Ronkainen, J. et al. Prevalence of Barrett’s esophagus in the general population: an endoscopic study. Gastroenterology 129, 1825–1831 (2005).
2. Jankowski, J.A., Harrison, R.F., Perry, L., Balkwill, F. & Tselipis, C. Barrett’s metaplasia. Lancet 356, 2079–2085 (2000).
3. Kulig, M. et al. Risk factors of gastroesophageal reflux disease: methodology and first epidemiological results of the ProGERD study. J. Clin. Epidemiol. 57, 580–589 (2004).
4. Koek, G.H., Sifrim, D., Lerut, T., Janssens, J. & Tack, J. Multivariate analysis of the association of acid and duodeno-gastro-oesophageal reflux exposure with the presence of oesophagitis, the severity of oesophagitis and Barrett’s oesophagus. Gut 57, 1056–1064 (2008).
5. Sikkema, M., de Jonge, P.J., Steyenberg, E.W. & Kuipers, E.J. Risk of esophageal adenocarcinoma and mortality in patients with Barrett’s esophagus: a systematic review and meta-analysis. Clin. Gastroenterol. Hepatol. 8, 235–244; quiz e32 (2010).
6. Jankowski, J.A., Provenzale, D. & Moayyedi, P. Esophageal adenocarcinoma arising from Barrett’s metaplasia has regional variations in the west. Gastroenterology 122, 588–590 (2002).
7. Thomas, T., Abrams, K.R., De Caestecker, J.S. & Robinson, R.J. Meta-analysis: cancer risk in Barrett’s oesophagus. Aliment. Pharmacol. Ther. 26, 1465–1477 (2007).
8. Corley, D.A. & Buffler, P.A. Oesophageal and gastric cardia adenocarcinomas: analysis of regional variation using the Cancer Incidence in Five Continents database. Int. J. Epidemiol. 30, 1415–1425 (2001).
9. Jankowski, J., Barr, H., Wang, K. & Delaney, B. Diagnosis and management of Barrett’s oesophagus. Br. Med. J. 341, c4551 (2010).
The authors of this paper are:

Zhan Su1, Laura J Gay2, Amy Strange1, Claire Palles1, Gavin Band1, David C Whiteman4, Francesco Lesci5, Cordelia Langford6, Manoj Nanj2, Sarah Edkins6, Anouk van der Winkel7, David Levine8, Peter Sasieni9, Céline Bellenguez10, Kimberley Howarth9, Colin Freeman8, Nigel Trudgill10, Art T Tucker11, Matti Pirinen1, Maikel P Peppelenbosch2, Luc J W van der Laan12, Ernst J Kuipers7, Joost P H Drenth13, Wilbert H Peters13, John V Reynolds14, Dermot P Kelleher15, Ross McManus15, Heike Grabsch16, Hans Prener17, Raf Bisschops18, Kausila Krishnadath18, Peter D Siersema19, Jantine W P M van Baal19, Mark Middleton20, Russell Petty21, Richard Gillies20, Nicola Burch22, Pradeep Bhandari23, Stuart Paterson24, Cathryn Edwards25, Ian Penman26,27, Kishor Vaidya28, Yung Ang29, Iain Murray30, Praful Patel31, Weimin Ye32, Paul Mullins33, Anna H Wu34, Nigel C Bird35, Helen Dallal36, Nicholas J Shaheen37, Liam J Murray38, Konrad Koss39, Leslie Bernstein40, Yvonne Romero41, Laura J Hardie42, Rui Zhang8, Helen Winter43, Douglas A Corley44, Simon Panter45, Harvey A Risch46, Brian J Reid47, Ian Sargeant48, Marilie D Gammon49, Howard Smart50, Anjan Dhar51, Hugh McMurtry52, Haythem Ali53, Geoffrey Liu54–56, Alan G Casson57, Hong–Chow58, Matt Rutter59, Ashref Tawil60, Danielle Morris61, Chuka Nwokolo62, Peter Isaacs63, Colin Rodgers64, Kris Ragunath65, Chris MacDonald66, Chris Haigh67, David Monks68, Gareth Davies69, Saj Wajed70, David Johnston71, Michael Gibbons72, Sue Cullen73, Nicholas Church74, Ruth Langley75, Michael Griffin76, Derek Alderson77, Panos Deloukas78, Sarah E Hunt79, Emma Gray80, Serge Dronov81, Simon C Potter82, Avazeh Tashakkori–Ghanbaria83, Mark Anderson84, Claire Brooks85, Jenefer M Blackwell86,87, Elvira Bramon82,83, Matthew A Brown84, Juan P Casas85,86, Aiden Corvin87, Audrey Duncanson88, Hugh S Markus89, Christopher G Mathew90, Colin N A Palmer91, Robert Plomnin92, Anna Rautanen93, Stephen J Sawcer94, Richard C Trembath95, Ananth C Viswanathan96, Nicholas Wood97,98, Gosia Trynka99, Cisca Wijmenga100, Jean–Baptiste Cazier101, Paul Atherfold102–104, Anna M Nicholson2, Nichola L Gellaty2, Deborah Glancy2, Sheldon C Cooper10, David Cunningham105, Tore Lind106, Julie Hapeshi107, David Ferry108, Barrie Rathbone109, Julia Brown110, Sharon Love111, Stephen Attwood112, Stuart MacGregor4, Peter Watson106, Scott Sanders107, Weronica Ek4, Rebecca F Harrison108, Paul Moayyedi109, John de Caestecker110, Hugh Barr111, Elia Stupka112,113, Thomas L Vaughan114, Leena Peltonen116, Chris C A Spencer1, Ian Tomlinson1,3, Peter Donnelly1,114,118 & Janusz A Z Jankowski2,20,22,115,118

24. Winney, B. et al. People of the British Isles: preliminary analysis of genotypes and surnames in a UK control population. Eur. J. Hum. Genet. 20, 203–210 (2012).
25. Martin, V. & Shaw-Smith, C. Review of genetic factors in intestinal malrotation. Pediatr. Surg. Int. 26, 769–781 (2010).
26. Su, Z. & Cardin, N. Welcome Trust Case Control Consortium, Donnelly, P. & Marchini, J. A Bayesian method for detecting and characterizing allelic heterogeneity and boosting signals in genome-wide association studies. Stat. Sci. 24, 430–450 (2009).
27. Sawyer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214–219 (2011).
28. Dittrich, A.T., Moutsianas, L., Leslie, S. & McKeown, G. HLA-1MP—an integrated framework for imputing classical HLA alleles from SNP genotypes. Bioinformatics 27, 968–972 (2011).
29. Shalaula, M.D. & Saad, R. Barrett’s esophagus. Am. Fam. Physician 69, 2113–2118 (2004).
30. Purcell, S.M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752 (2009).
31. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467, 832–838 (2010).
32. Akiyama, T. et al. Visceral obesity and the risk of Barrett’s esophagus. Digestion 83, 142–145 (2011).
33. Lagergren, J. Influence of obesity on the risk of esophageal disorders. Nat. Rev. Gastroenterol. Hepatol. 8, 340–347 (2011).
34. Spielotes, E.K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. 42, 937–948 (2010).
35. Loos, R.J. et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat. Genet. 40, 768–775 (2008).
36. Lindgren, C.M. et al. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. PLoS Genet. 5, e1000508 (2009).
37. Heid, I.M. et al. Meta-analysis of the INSIG2 association with obesity including 74,345 individuals: does heterogeneity of estimates relate to study design? PLoS Genet. 5, e1000694 (2009).
38. Heard-Costa, N.L. et al. NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. PLoS Genet. 5, e1000539 (2009).
39. Moayyedi, P. et al. Mortality rates in patients with Barrett’s esophagus. Aliment. Pharmacol. Ther. 27, 315–320 (2008).
ONLINE METHODS

Samples. Cases from discovery, stage 1 and 2 replication and stage 3 UK and Irish sets. For the discovery stage, we ascertained cases of histologically confirmed Barrett’s esophagus through the UK-based AspECT clinical trial of proton pump–inhibitor (esomeprazole) and aspirin as preventive agents for progression of Barrett’s esophagus to EAC.49 UK, Irish and Dutch replication cases were from the ChOPIN genetic study and the EAGLE Consortium. Replication cases were diagnosed with Barrett’s esophagus with lengths of at least 1 cm (CIMI) circumferential Barrett’s esophagus or at least a 2-cm tongue (COM2), according to the Prague criteria40. Case collection was in accordance with British Society of Gastroenterology criteria41, the standard practice for collaborating histopathologists in the UK and much of Europe. We found that 90% of our discovery samples (for which full clinicopathological data were available) had evidence of intestinal metaplasia and therefore also met the American College of Gastroenterology criteria that are widely used in the United States42. Full details of the ancestry, age and sex distributions and Prague criteria measurements of the cases are provided in Supplementary Table 1.

Discovery. The full data set comprised 1,991 cases and 5,667 controls. After quality control, 1,852 cases and 5,172 controls were analyzed. Controls were taken from the WTCCC2 set, made up of samples from 58C and UKBS.

Samples were genotyped at the Wellcome Trust Sanger Institute (WTSI), cases on the Illumina Human660W-Quad array and controls on the Illumina custom Human 1.2M-Duo array. Primary analysis was performed on the overlapping set of SNPs.

Stage 1. After quality control, the UK replication set totaled 1,105 cases and 6,819 controls. The controls were from the PoBI cohort13 (2,578) and 58C samples (4,241) that were not genotyped in the discovery phase. The case and 58C control samples were genotyped on the Illumina Immunochip, and the PoBI samples were genotyped on the Illumina custom Human 1.2M-Duo array. The Immunochip is a custom–designed chip containing 196,524 SNPs in total, of which ~2,400 were selected on the basis of our discovery GWAS study.

Stage 2. The Dutch replication cohort consisted of 473 cases and 1,780 controls. These samples were all genotyped on the Illumina Immunochip but in two separate locations; the case samples were genotyped at the WTSI, and the control samples were genotyped as described in a previous report43.

Information on DNA sample preparation is provided in the Supplementary Note.

Circumferential and maximal extent phenotypes. Length of the Barrett’s segment was available for a subset of discovery and replication phase samples. Where baseline measurements were not available, the earliest measurement taken after baseline was used. A small number of cases were excluded on the basis of reporting errors (if C > M or if either value exceeded 25 cm). Of the discovery phase individuals after quality control, 1,744 had C measurements, and 1,618 had M measurements: C mean = 4.05 (range 0–22); M mean = 4.60 (range 1–24). M measurements were available for 1,015 samples from the stage 1 replication (M mean = 4.66 (range 1–23)) and for 240 samples from the stage 2 replication (M mean = 4.44 (range 1–15)). Both C and M phenotypes were square-root transformed before analysis to improve the fit of the linear regression model.

Stage 3. Irish replication. The Irish replication cohort consisted of 245 cases and 473 controls. Cases were provided by St. James’ Hospital and Mater Misericordiae University Hospital, Dublin, as part of EAGLE. Controls were provided by Trinity Biobank. Cases (168) were genotyped on the Illumina Immunochip at WTSI. rs9257809 and rs9936833 were genotyped in 77 cases and all controls using competitive allele-specific PCR KASPar chemistry (Kbiosciences). Details on primers, probes and conditions used are available on request. Genotyping quality control was tested using duplicate DNA samples within studies and SNP assays, together with direct sequencing of subsets of samples to confirm genotyping accuracy. For all SNPs, >99% concordant results were obtained.

UK replication 2. We ascertained 1,765 cases using the diagnostic criteria and sampling from AspECT as described for the discovery stage. Controls (1,586) were collected as part of the Colorectal Tumour Gene Identification (CoRGI) Consortium44. Controls were spouses or partners unaffected by cancer and without a personal family history (in up to second-degree relatives) of colorectal neoplasia. All were of European ancestry, 45% were male, and the mean age ± standard deviation was 45.1 ± 15.9 years. All samples were genotyped using KASPar competitive allele-specific PCR.

Written informed consent was obtained from all subjects. The ethics of the project were reviewed by the East London and the City Research Ethics Committee (8 June 2004, 04/Q0603/1). All UK studies were performed with national ethical committee approval (MREC numbers: AspECT 04/Q0603/1; ChOPIN/IP06/Q1603/07; HANDDEL 09/H0050/23; and CORGI MREC/06/Q1702/99). The Irish replication cohort (part of EAGLE) was collected with approval from the Research Ethics Committee Board of St. James’ Hospital. The Dutch replication cohort was collected with approval from the ethics committee or institutional review board of all participating institutions.

BEACON replication. We analyzed 2,398 cases and 2,167 controls. Samples were collected as part of a genome-wide association study (BEAGESS) undertaken by the BEACON collaboration. Samples were collected from sites in Australia (cases n = 325, controls n = 361), Europe (England, Ireland and Sweden; cases n = 363, controls n = 333) and North America (Canada and United States, cases n = 1,710, controls n = 1,273). Samples were genotyped at the Fred Hutchinson Cancer Research Center on the Illumina Omni1M Quad array. The Institutional Review Board or Research Ethics Committee for each study within BEACON approved data collection.

Quality control. Samples. As previously described27,45, we identified and removed samples whose genome-wide patterns of diversity differed from those of the collection at large, interpreting them as likely to be due to biases or artifacts. Further details are given in the Supplementary Note. After sample quality control, our final discovery data set consisted of 1,852 cases and 5,172 controls (Supplementary Table 3).

SNPs. For all arrays, normalized probe intensities were exported using the BeadStudio program, and genotypes were called at the WTSI using Illuminu46. SNPs were excluded from analysis if in any of the data sets (58C, UKBS or cases) they had a very low minor allele frequency (defined as <0.01%); showed extreme departure from Hardy–Weinberg equilibrium (P < 1 × 10−4); or showed a strong plate effect (P < 1 × 10−4). SNPs were also excluded if the observed statistical (Fisher) information about the allele frequency was less than 98% of the information contained in a hypothetical sample of the same size and expected minor allele frequency but with no missing data. We removed 45 SNPs after visual inspection of cluster plots. In total, 521,744 autosomal SNPs were available for analysis after quality control.

To confirm the genotyping accuracy of the different platforms used in the study, 5% of the UK, Irish and Dutch samples typed on each platform were regenotyped at rs9257809 and rs9936833 using competitive allele-specific PCR KASPar. Concordance was >99% (Supplementary Table 11), suggesting genotyping robustness across platforms.

HLA allele imputation. Classical HLA alleles were imputed using HLA*IMP28. Further details can be found in the Supplementary Note.

Statistical analysis. Genome-wide case-control analysis was performed using frequentist tests under a missing data logistic regression model, as implemented in SNPTYPE. Unless otherwise stated, we assumed a multiplicative model for allelic risk by encoding the genotypes at each SNP as a discrete explanatory variable, with an indicator of case status as the binary response and the first principal component as a covariate (Supplementary Note). Quantitative C and M measurements were analyzed using frequentist tests under a missing data linear regression model, as implemented in SNPTYPE. To combine the evidence of association across the discovery and replication data sets, we conducted an inverse variance–weighted fixed-effect meta-analysis in the statistical package R (Supplementary Note). To test for interactions (Supplementary Note) between SNPs or between a SNP and sex and to compare models that included additional SNPs or classical HLA alleles as predictors, we used logistic regression models implemented in R. These analyses used thresholded (posterior probability > 0.9) genotype calls.
SNP imputation was performed using IMPUTE2 (ref. 47), which adopts a two-stage approach using both a haploid reference panel and a diploid reference panel.

BEACON data were analyzed under an additive logistic regression model, including the first four principal components as covariates (Supplementary Note). Genomic inflation \( \hat{\lambda} \) was 1.037.

Combined analysis was carried out on the discovery and stage 1 data. To reduce possible population structure (such analyses are sensitive to this), we restricted the stage 1 control set to the 58C individuals. SNPs with minor allele frequency of >0.01 that were genotyped in both the discovery (Illumina 670K and Illumina custom Human 1.2M-Duo arrays) and the replication (Illumina Immunochip) stages were pruned to remove strong LD. This was done by ranking the SNPs by Bayes factor calculated under an additive model in SNPTEST and successively selecting SNPs from the top, requiring that they be at least 0.125 cM and 25 kb away from any SNPs that had already been selected. We obtained 7,673 SNPs from a total of 28,972 (after quality control) that were typed in discovery and UK Immunochip data. For the \( k \) SNPs showing the strongest signal of association, the sign test compared the direction of effect of each SNP in the discovery and replication samples. Using a likelihood-ratio test, we compared the null model where the probability of the same direction of effect is assumed to be 0.5 to a model where the probability is not 0.5 (two sided).

The disease score test aims to measure indirectly the collective effect of many weakly associated alleles. We determined the risk allele and odds ratio for each pruned SNP from the discovery data. Then, we used the top \( k \) SNPs to calculate the 'score' for each individual in the replication data as the number of risk alleles carried by each individual weighted by the log of the odds ratio estimated from the discovery data. Under the null hypothesis, the risk alleles and odds ratios in the discovery and replication samples are independent. We tested a logistic regression model of disease status on the score in the replication data, conditioning on the first principal component, to control for population structure, and the number of missing genotypes (called with maximum probability of <0.9), to control for potential differences in genotyping rate, as covariates.

40. Sharma, P. et al. The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. Gastroenterology 131, 1392–1399 (2006).
41. Watson, A., Heading, R.C. & Shepherd, N.A. Guidelines for the diagnosis and management of Barrett's columnar-lined oesophagus. in A Report of the Working Party of the British Society of Gastroenterology 1–3 (Q3 Print Project Management, Loughborough, UK, 2005).
42. Sampliner, R.E. Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus. Am. J. Gastroenterol. 97, 1888–1895 (2002).
43. Trynka, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. 43, 1193–1201 (2011).
44. Houlston, R.S. et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat. Genet. 42, 973–977 (2010).
45. Spencer, C.C. et al. Dissection of the genetics of Parkinson’s disease identifies an additional association 5′ of SNCA and multiple associated haplotypes at 1q21. Hum. Mol. Genet. 20, 349–353 (2011).
46. Teo, Y.Y. et al. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics 23, 2741–2746 (2007).
47. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).