Supplementary Methods

SUBJECTS AND METHODS

Subject populations

Esophageal Adenocarcinoma and Barrett's esophagus

Specifically; two population based studies from the Queensland Institute of Medical Research (1,2); one population based study from the University of Southern California (3); a Kaiser Permanente, Northern California community-based case-control study (4); a Nova Scotia, Canada case-control study based on cases seen at a tertiary referral center (5,6); The US Multicenter Study, a population-based case-control study involving three areas, New Jersey, Connecticut and western Washington state, in a cooperative agreement with NCI (7-9); a simultaneous population-based study of EA and BE from Ireland (Republic and Northern), using hospital, clinic and pathology records, with a shared control group representing the population from a province-wide database of practitioner records included (10,11); a population-based Swedish study of EA cases and representative population controls selected from the population registry (12); a tertiary clinic-based English case-control study of EA and BE, with non-esophagitis clinic controls (13); a Toronto, Canada hospital-based case-control study of EA using controls selected from case friends or spouse of other cancer cases (14); a prospective cohort of consecutive patients with long segment BE and EA recruited by a multi-campus, multidisciplinary, tertiary care academic consortium since September 10, 2001, located in Rochester, MN, Scottsdale, AZ, and Jacksonville, FL (15-17); a University of North Carolina study of BE and GERD controls (18); a western Washington community-clinic based study of newly diagnosed BE cases and population controls (19); and a western Washington cohort of persons with BE under active surveillance for the development of EA from the Seattle Barrett's Esophagus Research Program (20,21) .

EA cases had clinically diagnosed and histologically confirmed EA. BE patients had histologically confirmed intestinal metaplasia and endoscopically-evident columnar epithelium in the tubular esophagus (a more detailed definition of BE for each study can be found in Supplementary Table 1). Samples were genotyped on the Illumina Omni1-Quad array. Quality control procedures removed low-quality and incorrectly identified samples. Further, related individuals were identified and only one individual from each family was used in any particular analysis.

GERD

Swedish Twin Registry; Trained professional interviewers conducted computer-assisted telephone interviews. A structured questionnaire was used to elicit a history of reflux symptoms. Individuals were asked if they ever had "heartburn", "pain behind the breastbone", or "regurgitation of bitter or sour fluid into the mouth". If a positive response was given to any one of these questions, seven more questions were asked to determine whether these reflux symptoms were specific to GERD. These questions assessed the frequency and duration of symptoms, radiation or discomfort toward the neck, night waking, antacid relief, and use of histamine-receptor antagonistic or proton pump inhibitor medications. GERD was defined as a symptom frequency of at least once per week, of either retrosternal pain with antacid relief, retrosternal burning with antacid relief, or radiation toward the neck; or regurgitation of bitter fluid. Twins that answered "no" to the question of whether they had had heartburn or reflux during the last year were included as controls. One twin from each pair was included in the study. To get a more representative set of controls, we also excluded all monozygotic (MZ) twin pairs with discordant answers (if one MZ twin had never had reflux while the sibling had any of the symptoms, both were discarded).

The UK St Thomas Adult Twin Registry; Each individual was sent a 25-item questionnaire, covering demographic details, as well as symptoms of heartburn and acid regurgitation during the past year. GERD was defined as having at least weekly symptoms of heartburn or acid regurgitation. One individual from each family was included in the analysis. In the case of symptom discordance between MZ twins, both twins were excluded from analyses. Controls comprised individuals that answered "no" to whether they had had heartburn or reflux during past year.

Statistical Methods

Estimates of variance explained can be biased by genotyping errors and we therefore applied a stricter quality control than for typical GWAS analyses. SNPs with minor allele frequency < 0.01, P values for Hardy-Weinberg equilibrium < 10^{-4} and missing call rates > 0.05, were excluded from the analysis. Individuals with a missing call rate greater than 1% were excluded. Individuals were excluded to ensure that no pairs had an estimated genetic relationship > 0.025 (approximately a second cousin relationship). Related pairs were excluded to avoid the possibility that the phenotypic resemblance between close relatives could be because of non-genetic effects (for example, shared environment). Quality control for the polygenic overlap analysis was as for GCTA analysis above.

A genetic correlation above zero suggests that the two traits are influenced by common genes. A genetic correlation close to 1 (as in our estimations) means that BE and EA share a large proportion of common genes. The genetic correlation was estimated to 0.96 between males and females, which suggests that the same genes influence the trait in both males and females, and that the higher prevalence of BE and EA in males is not explained by specific sex effects. To ensure that bivariate analysis uses two independent data sets for EA and BE, we allocated some controls to EA, with the remainder allocated to BE. To determine

if our particular allocation influenced results we randomly reassigned controls to either BE or EA. Since all the analysis was done within continent (i.e. a factor for continent was included in analysis), random reassignment was done within continent (America, Europe, Australia). With the reallocated control set, we estimated the genetic correlation (rG). Based on 200 randomly generated data sets, the median rG was 0.87, suggesting that even if we had chosen a different allocation of controls to case sets, our conclusion that the genetic correlation was very high (close to 1) would be unchanged.

To demonstrate the robustness of the methods, we tested our approach by performing a control-control analysis. We took the controls used for BE and compared them to the controls used for EA, where we coded one control group as cases. Since 'lifetime' risk could not be defined for this situation, we did the calculation on the observed case-control scale. Note that the significance of the genetic component is not affected by scaling for lifetime risk, only the proportion of variance explained. This calculation did not show a significant h^2_g or r_g.

Since we only used a finite number of SNPs to predict the genetic relationship at trait associated loci (i.e. the genomic region that is associated with the trait), we needed to adjust prediction errors due to imperfect linkage disequilibrium (LD). Here we assumed that the trait associated loci had a similar distribution of allele frequencies as the genotyped SNPs.

We examined PC plots for PCs1-4 (Supplementary Figures 1A-B).

Limitations

We could not measure directly the possible effects of shared environmental effects between distantly related persons. We examined the effect of changing the degree of relatedness on our estimates of variance explained. The estimates for BE were 36%, 36% and 35% after excluding participants at relatedness thresholds of 10%, 5% and 2.5% respectively. Estimates

for EA were similarly stable, arguing against a strong effect of shared environment as an explanation for our findings.

Acknowledgement

Members of the Mayo Clinic Romero Registry Consortium ([1]Mayo Clinic Florida, [2]Mayo Clinic Rochester, [3]Mayo Clinic Arizona); Sami R.Achem, M.D.[1], David A. Ahlquist, M.D.[2], Steven R. Alberts, M.D.[2] , Jeffrey A. Alexander, M.D.[2] , Mark S. Allen, M.D.[2], Amindra S. Arora, MBBChir[2], Jonathan B. Ashman, M.D., Ph.D.[3], Pamela J. Atherton[2], Lisa A. Boardman, M.D.[2], Ernest P. Bouras, M.D.[1], Vicki A. Bryhn[2], Patrick A. Burch, M.D.[2], George E. Burdick, M.D.[2], Navtej S. Buttar, M.D.[2], John K. Camoriano, M.D.[3], John R. Cangemi, M.D.[1], Stephen D. Cassivi, M.D.[2], Frances K. Cayer[1], Amy C. Clayton, M.D.[2], Michael D. Crowell, M.D., Ph.D.[3], Julie M. Cunningham, Ph.D.[2], Mariza de Andrade, Ph.D.[2], Piet de Groen, M.D.[2], Giovani De Petris, M.D.[3], Claude Deschamps, M.D.[2], Kenneth R. DeVault, M.D.[1], Robert B. Diasio, M.D.[2], John K. DiBaise, M.D.[3], Eric S. Edell, M.D.[2], Sharon Elcombe[2], Charles Erlichman, M.D.[2], Douglas O. Faigel, M.D.[3], Tom R. Fitch, M.D.[3], David E. Fleischer, M.D.[3], Jean C. Fox, M.D.[2], Amy E., Foxx-Orenstein, D.O.[2], Dawn Francis, M.D.[2], Mary B. Fredericksen[2], Evanthia Galanis, M.D.[2], Debra M. Geno[2], Axel Grothey, M.D.[2], Michael G. Haddock, M.D.[2], Kevin C. Halling, M.D., Ph.D.[2], Denise M. Harnois, D.O.[1], Tracy W. Hilton[1], Timothy Hobday, M.D.[2], Lesley A. Houghton, Ph.D.[1], Prasad A. Iyer, M.D.[2], Dawn E. Jaroszewski, M.D.[3], Aminah Jatoi, M.D.[2], Robert B. Jenkins, M.D., Ph.D.[2], Elizabeth A. Johnson, M.D.[1], Rajni Katipamula, M.D.[2], David A. Katzka, M.D.[2], Sharon F. Kaufman[2], Andrew P. Keaveny, M.D.[1], Daniel A. Keller[2], George P. Kim, M.D.[1], Benjamin R. Kipp, Ph.D.[2], Dora M. Lam-Himlin, M.D.[2], Stephen M. Lange, M.D.[1], Louis Lanza, M.D.[3], Shauna N. Legrand[1], Paul J. Limburg, M.D.[2], Wilma L. Lingle, Ph.D.[2], Wanguo Liu, Ph.D[2], G. Richard Locke, III, M.D.[2], David S. Loeb, M.D.[1], Lori S. Lutzke[2], James A. Martenson, Jr., M.D.[2] Robert R. McWilliams, M.D.[2], Robert C. Miller, M.D.[2], Joseph A. Murray, M.D.[2], Cuong C. Nguyen, M.D.[3], Francis C. Nichols, III, M.D.[2], John A. Odell, M.D.[1], Rahul Pannala, M.D.[3], Harshita R. Paripati, M.D.[2], Alexander S. Parker, Ph.D.[1], Shabana F. Pasha, M.D.[2], Bret T. Petersen, M.D.[2], Gloria M. Petersen, Ph.D.[2], Michael F. Picco, M.D.[1], Henry C. Pitot, M.D.[2], J. Fernando Quevedo, M.D.[2], Massimo Raimondo, M.D.[1], Elizabeth Rajan, M.D.[2], Francisco (Pancho) C. Ramirez, M.D.[2], Yvonne Romero, M.D.[2], Renee M. Root[2], Helen J. Ross, M.D.[3], Joseph Rubin, M.D.[2], Schuyler O. Sanderson, M.D.[2], Daniel J. Sargent, Ph.D.[2], Daniel J. Schaid, Ph.D.[2], K. Robert Shen, M.D.[2], Qian Shi, Ph.D.[2], Frank A. Sinicrope, M.D.[2], Jeff A. Sloan, Ph.D.[2], Dan C. Smith, M.D.[1],

David I. Smith, Ph.D.[2], Thomas C. Smyrk, M.D.[2], Mark E. Stark, M.D.[1], Nicholas Talley, M.D., Ph.D.[2], Stephen N. Thibodeau, Ph.D.[2], Michael D. Van Norstrand, M.D., Ph.D.[2], Michael B. Wallace, M.D.[1], Kenneth K. Wang, M.D.[2], Richard M. Weinshilboum, M.D.[2], Dennis Wigle, M.D., Ph.D.[2], Herbert C. Wolfsen, M.D.[1], Louis M. Wong Kee Song, M.D.[2], Timothy A. Woodward, M.D.[1], Tsung-The Wu, M.D., Ph.D.[2], Harry H. Yoon, M.D.[2], Alan R. Zinsmeister, Ph.D[2].

Supplementary Table 1. Individuals included from each study site for the GWAS data used.

Location	Study/ Reference	n genotyped (n excluded)			Total Genotyped	Total Analysed
		EA[a] cases	BE[b] cases	Controls		
Australia						
Australia-wide		236	0	245	481	481
Queensland-Australia		0	326[1] (1)	323 (5)	655	649
Subtotal		236	326 (1)	568 (5)	1136	1130
Europe						
Sheffield, England		102	167[2] (7)	0	276	269
Sweden-wide		64	0	116 (1)	181	180
Ireland, Republic of Ireland		194	199[3]	218 (2)	613	611
Subtotal		360	366 (7)	334 (3)	1070	1060
North America						
Kaiser Permanente, Northern California, US		0	242[4] (30)	215 (30)	517	457
Washington & New Jersey, US	EGA study	56	0	114 (2)	172	170
Rochester, Minnesota, US	Mayo registry	503 (2)	814[5] (5)	0	1324	1317
Toronto, Ontario, Canada		248 (23)	0	259 (13)	543	507
Raleigh, North Carolina, US		0	100[6] (1)	0	101	100
Washington, US	Study Reflux Disease	0	157[7] (3)	167	327	324
Washington, US	Seattle Barrett's Esophagus Program	0	296[8] (6)	0	302	296
Nova Scotia, Canada		54	115[9] (6)	92 (1)	268	261
Los Angeles, California, US		60 (1)	0	438 (6)	505	498
Subtotal		921 (26)	1724 (51)	1285 (52)	4059	3930
Total BEACON consortium		1517 (26)	2416 (59)	2187 (60)	6265	6120

[a]Esophageal adenocarcinoma, [b]Barrett's esophagus.

[1]BE was defined as the presence of specialized intestinal metaplasia (columnar epithelium with goblet cells) in a biopsy taken from the esophagus by upper gastrointestinal endoscopy, regardless of the

length of involvement. Patients with specialized intestinal metaplasia detected only in biopsies taken from the gastric cardia were not eligible for inclusion.

[2]BE was defined as any length of histologically confirmed specialized intestinal metaplasia containing goblet cells.

[3]BE patients were eligible for inclusion if ≥3 cm of typical Barrett's mucosa were seen at endoscopy, and the presence of specialized intestinal metaplasia was confirmed by histologic examination of biopsy specimens. Patients with dysplasia on histologic examination were not included.

[4]BE was defined if the endoscopist clearly described a visible length of columnar-type epithelium proximal to the gastroesophageal junction/gastric fold and if a biopsy showed specialized intestinal epithelium. Pathology slides underwent a separate manual review by a gastrointestinal pathologist. The following patients were excluded: patients with only gastric-type metaplasia of the esophagus on all pathologic evaluations, patients with columnar metaplasia without features of intestinal metaplasia on all pathology readings, patients without a biopsy specimen of esophageal origin, biopsy specimens of only a mildly irregular squamocolumnar junction.

[5]BE was defined as long segment (≥3 cm), histologically confirmed specialized intestinal metaplasia containing goblet cells.

[6]BE was defined as any detectable upward displacement of the squamocolumnar junction into the tubular esophagus, with at least one biopsy specimen showing columnar epithelium with goblet cells. Patients with goblet cells on biopsy examination but no endoscopic appearance of BE were not eligible for inclusion.

[7]During the endoscopy procedure, the physicians recorded the presence or absence of visible columnar epithelium and, if present, its length. Based on these findings, cases were subsequently classified into 1, 2, or 3 of the following progressively exclusive groups: (1) BE cases (ie, all cases), (2) BE cases with visible columnar epithelium (visible Barrett’s esophagus) and (3) BE cases with visible column epithelium greater than 2 cm (long-segment BE).

[8]BE was diagnosis as metaplastic columnar epithelium with intestinal metaplasia in esophageal biopsies, the absence of esophageal malignancy at or prior to baseline endoscopy, and having had at least one follow-up endoscopy.

[9]The diagnosis of BE was established by the histological finding of intestinal metaplasia, which was confirmed independently by two consultant gastrointestinal histopathologists.

References

1. Whiteman D, Sadeghi S, Pandeya N, et al. Combined effects of obesity, acid reflux and smoking on the risk of adenocarcinomas of the oesophagus. *GUT.* 2008;57(2):173-180.

2. Smith K, O'Brien S, Smithers B, et al. Interactions among smoking, obesity, and symptoms of acid reflux in Barrett's esophagus. *Cancer Epidemiol Biomarkers Prev.* 2005;11(14):2481-2486.

3. Wu A, Wan P, Bernstein L. A multiethnic population-based study of smoking, alcohol and body size and risk of adenocarcinomas of the stomach and esophagus. *Cancer Causes Control.* 2001;12(8):721-732.

4. Corley D, Kubo A, Levin T, et al. Abdominal obesity and body mass index as risk factors for Barrett's esophagus. *Gastroenterol.* 2007;133(1):34-41.

5. Casson A, Zheng Z, Evans S, Veugelers P, Porter G, Guernsey D. Polymorphisms in DNA repair genes in the molecular pathogenesis of esophageal (Barrett) adenocarcinoma. *Cacriogenesis.* 2005;26(9):1536-1541.

6. Veugelers P, Porter G, Guernsey D, Casson A. Obesity and lifestyle risk factors for gastroesophageal reflux disease, Barrett esophagus and esophageal adenocarcinoma. *Dis Esophagus.* 2006;19(5):321-328.

7. Engel L, Chow W, Vaughan T, et al. Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst.* 2003;95(18):1404-1413.

8. Chow W, Blot W, Vaughan T, et al. Body mass index and risk of adenocarcinomas of the esophagus and gastric cardia. *J Natl Cancer Inst.* 1998;90(2):150-155.

9. Gammon M, Schoenberg J, Ahsan H, et al. Tobacco, alcohol, and socioeconomic status and adenocarcinomas of the esophagus and gastric cardia. *J Natl Cancer Inst.* 1997;89(17):1277-1284.

10. Anderson L, Johnston B, Watson R, et al. Nonsteroidal anti-inflammatory drugs and the esophageal inflammation-metaplasia-adenocarcinoma sequence. *Cancer Res.* 2006;66(9):4975-4982.

11. Corley D, Kerlikowske K, Verma R, Buffler P. Protective association of aspirin/NSAIDs and esophageal cancer: a systematic review and meta-analysis. *Gastroenterol.* 2003;124(1):47-56.

12. Lagergren J, Bergström R, Lindgren A, Nyrén O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med.* 1999;340(11):825-831.

13. Martino Ed, Hardie L, Wild C, et al. The NAD(P)H:quinone oxidoreductase I C609T polymorphism modifies the risk of Barrett esophagus and esophageal adenocarcinoma. *Genet Med.* 2007;9(6):341-347.

14. Zhai R, Liu G, Asomaning K, et al. Genetic polymorphisms of VEGF, interactions with cigarette smoking exposure, and esophageal adenocarcinoma risk. *Cacriogenesis.* 2008;29(12):2330-2334.

15. Miller RC, Atherton PJ, Kabat BF, et al. Marital status and quality of life in patients with esophageal cancer or Barrett esophagus: A Mayo Clinic Esophageal Adenocarcinoma and Barrett's Esophagus Registry Study. *Dig Dis Sci* 2010;55(10):2860-2868.

16. Gatenby PA, Caygill CP, Watson A, Murray L, Romero Y. Barrett's esophagus registries. *Ann N Y Acad Sci* 2011;1232(1):405-410.

17. Stauder MC, Romero Y, Kabat B, et al. Overall survival and self-reported fatigue in patients with esophageal cancer. *Support Care Cancer.* 2012.

18. Shaheen N, Green B, Medapalli R, et al. The perception of cancer risk in patients with prevalent Barrett's esophagus enrolled in an endoscopic surveillance program. *Gastroenterol.* 2005;129(2):429-436.

19. Edelstein Z, Farrow D, Bronner M, Rosen S, Vaughan T. Central adiposity and risk of Barrett's esophagus. *Gastroenterol.* 2007;133(2):403-411.

20. Galipeau P, Li X, Blount P, et al. NSAIDs modulate CDKN2A, TP53, and DNA content risk for progression to esophageal adenocarcinoma. *PLoS Med.* 2007;4(2):e67.

21. Lin D. Evaluating statistical significance in two-stage genomewide association studies. *Am J Hum Genet.* 2006;78:505-509.