A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett’s esophagus

David M Levine¹, Veronica E Ek², Rui Zhang¹, Xinxue Liu³, Lynn Onstad⁴, Cassandra Sather⁵, Pierre Lao-Sirieix⁶, Marilie D Gammon⁶, Douglas A Corley⁷,⁸, Nicholas J Shaheen⁹, Nigel C Bird¹⁰, Laura J Hardie¹¹, Liam J Murray¹², Brian J Reid¹³,¹⁴, Wong-Ho Chow¹⁴, Harvey A Risch¹⁵, Olof Nyrén¹⁶, Weimin Ye¹⁶, Geoffrey Liu¹⁷, Yvonne Romero¹⁸,¹⁹, Leslie Bernstein²⁰, Anna H Wu²¹, Alan G Casson²², Stephen J Chanock²³, Patricia Harrington²⁴–²⁶, Isabel Caldas²⁶,²⁷, Irene Debiram-Beecham³, Carlos Caldas²⁶,²⁷, Nicholas K Hayward²⁸, Paul D Pharoah²⁴–²⁶, Rebecca C Fitzgerald³, Stuart MacGregor², David C Whiteman²⁹ & Thomas L Vaughan³⁰

Esophageal adenocarcinoma is a cancer with rising incidence and poor survival. Most such cancers arise in a specialized intestinal metaplastic epithelium, which is diagnostic of Barrett’s esophagus. In a genome-wide association study, we compared esophageal adenocarcinoma cases (n = 2,390) and individuals with precancerous Barrett’s esophagus (n = 3,175) with 10,120 controls in 2 phases. For the combined case group, we identified three new associations. The first is at 19p13 (rs10419226: \( P = 3.6 \times 10^{-10} \)) in CRTC1 (encoding CREB-regulated transcription coactivator), whose aberrant activation has been associated with oncogenic activity. A second is at 9q22 (rs11789015: \( P = 1.0 \times 10^{-6} \)) in BARX1, which encodes a transcription factor important in esophageal specification. A third is at 3p14 (rs2687201: \( P = 5.5 \times 10^{-7} \)) near the transcription factor FOXP1, which regulates esophageal development. We also refine a previously reported association with Barrett’s esophagus near the putative tumor suppressor gene FOXP1 at 16q24 and extend our findings to now include esophageal adenocarcinoma.

A genetic component to the development of Barrett’s esophagus and esophageal adenocarcinoma has long been suspected on the basis of previous studies in unrelated individuals and familial disease clusters. The current study leverages the resources of the Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON) and combines high-quality, largely population-based epidemiological studies of esophageal adenocarcinoma and Barrett’s esophagus conducted over a period of two decades.

For the discovery phase, we used 1,516 esophageal adenocarcinoma cases, 2,416 Barrett’s esophagus cases and 3,209 controls, all of European ancestry, after rigorous quality control procedures were applied to the genotyping data (Online Methods). The cases and 2,187 of the controls were collected by investigators in BEACON from cohort and case-control studies conducted in Western Europe, Australia and North America. An additional 1,022 cancer-free controls were obtained from a study of melanoma and were included to increase statistical power. All cases were histologically confirmed. The distribution of samples by study is given in Supplementary Table 1, and demographic characteristics of samples are given in Table 1.
Table 1  Top five newly identified SNPs associated with Barrett’s esophagus and esophageal adenocarcinoma

| SNP | Chr. | Position  | Nearest gene | Cases/controls |
|-----|------|-----------|--------------|---------------|
|     |      |           |              | T/G           |
| rs2687201 | 3 | 70928930  | FOXP1        | 2.00 × 10⁻⁶   |
| rs11789015 | 9 | 96716028  | BARX1        | 5.08 × 10⁻⁶   |
| rs6479527  | 9 | 96858411  | PTPOC1       | 4.74 × 10⁻⁵   |
| rs10419226 | 19| 18803172  | CRTC1        | 5.54 × 10⁻⁸   |
| rs10423674 | 19| 18817903  | CRTC1        | 1.92 × 10⁻⁶   |

Shown are the discovery, replication and meta-analysis results for Barrett’s esophagus (BE) versus controls, esophageal adenocarcinoma (EA) versus controls and combined Barrett’s esophagus and esophageal adenocarcinoma (BE + EA) versus controls. Included for each SNP are the association P value, the OR and 95% CI for the minor allele and the frequency of the minor allele in cases and controls. The slight variation in the number of discovery controls reflects the use of only unrelated samples for analysis, although six two-person families are present in the data set. Chr., chromosome. *Coded allele.

Supplementary Table 2. All samples were genotyped on the Illumina HumanOmni1-Quad platform.

We performed association analyses on the 922,031 autosomal and X-chromosome SNPs that passed quality control using an additive logistic regression model implemented in GWAStools11, including as covariates age, sex and the first four eigenvectors from principal-component analysis (PCA). To assess variants not present on the Illumina HumanOmni1-Quad chip, we performed imputation for X-chromosome SNPs that passed quality control using an additive logistic regression model implemented in GWASTools11, including as covariates age, sex and the first four eigenvectors from principal-component analysis (PCA). To assess variants not present on the Illumina HumanOmni1-Quad chip, we performed imputation for X-chromosome SNPs that passed quality control using an additive logistic regression model implemented in GWASTools11.

We selected 94 associated (P < 1 × 10⁻⁸) SNPs for replication. Of these, 87 were genotyped in 874 histologically confirmed esophageal adenocarcinoma cases from the Stomach and Oesophageal Cancer Study (SOGS), 759 were histologically confirmed Barrett’s esophagus cases from the UK Barrett’s Esophagus Gene Study (UK Gene Study) and 6,911 were controls, of which 1,711 were from the SEARCH Study and 5,200 were from the Wellcome Trust Case Control Consortium 2 (WTCCC2)12. All SOGS, UK Gene Study and SEARCH samples self-identified as Caucasian and were genotyped on a Fluidigm 96.96 Dynamic Array IFC. WTCCC2 subjects were of European ancestry, as determined by projection onto the first two principal components of a PCA of HapMap individuals, and were genotyped on a custom version of the Illumina Human1.2M-Duo array. Replication analysis was carried out using an additive logistic regression model with sex as a covariate. METAL software13 was used for meta-analysis of the discovery and replication data sets.

The three loci that reached genome-wide significance (P < 5 × 10⁻⁸) in the combined case group meta-analysis are given in Table 1, and results for all replicated SNPs are given in Supplementary Table 3. The most strongly associated SNP for each of the three loci had similar odds ratios (ORs) for Barrett’s esophagus and esophageal adenocarcinoma. None of the top imputed SNPs showed substantially stronger association than the genotyped SNPs.
Table 2  SNPs in region 16q24 on chromosome 16 near the FOXF1 gene associated with Barrett's esophagus and esophageal adenocarcinoma

| SNP             | Position | Discovery (BE) | OR (95% CI) | MAF case/control | Replication (BE) | OR (95% CI) | MAF case/control | Meta-analysis (BE) | OR (95% CI) | MAF case/control | Discovery (EA) | OR (95% CI) | MAF case/control | Replication (EA) | OR (95% CI) | MAF case/control | Meta-analysis (EA) | OR (95% CI) | MAF case/control |
|-----------------|----------|---------------|-------------|------------------|------------------|-------------|------------------|----------------------|-------------|------------------|---------------|-------------|------------------|------------------|-------------|------------------|----------------------|-------------|------------------|
| rs1490865       | G/A      | 7.00 × 10⁻²   | 1.22 × 10⁻⁴ | 1.69 × 10⁻⁴      | 3.14 × 10⁻⁶      | 2.88 × 10⁻⁶  | 1.20             | 1.11–1.30           | (1.11–1.30) | 0.504/0.465 | 1.37/1.19 | 0.87/0.63       | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |
| rs3111601       | G/A      | 6.84 × 10⁻⁴   | 1.20          | 1.11–1.30 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |
| rs9936833       | A/C      | 1.29 × 10⁻²   | 1.17          | 1.11–1.30 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |
| rs1728400       | T/G      | 2.88 × 10⁻⁶   | 1.20          | 1.11–1.30 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |
| rs3950627       | G/A      | 3.65 × 10⁻⁴   | 1.20          | 1.11–1.30 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |
| rs2178146       | A/G      | 1.04          | (0.93–1.17)  | 0.332/0.303 | 0.397/0.366   | 0.473/0.434  | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |
| rs13332095      | A/G      | 2.24 × 10⁻⁰   | 1.11          | 1.11–1.30 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 | 1.37/0.88 | 0.504/0.465 |

Shown are the discovery, replication and meta-analysis results for SNPs 100 kb up- or downstream of rs9396833, a previously identified susceptibility locus for Barrett’s esophagus14. Data are shown for rs9396833, SNPs that had more significant P values than rs9396833 in the combined Barrett’s esophagus and esophageal adenocarcinoma data set (P_discovery (BE + EA)) and SNPs that were the most significant SNP in conditional analysis in this region after fitting other SNPs (Table 3). Chr., chromosome.

*Bolded alleles.

The most significantly associated locus was found at 19p13 (Fig. 2a): rs10419226, P_meta (BE + EA) = 3.55 × 10⁻¹⁰, OR = 1.18, 95% confidence interval (CI) = 1.12–1.24. Five imputed SNPs in high linkage disequilibrium (LD) with rs10419226 (r² > 0.85) had associations that reached genome-wide significance in the combined discovery data set, and two were also significant in the Barrett’s esophagus discovery set (Supplementary Table 4). A second significantly associated locus was identified at 9q22.32 (Fig. 2b): rs11789015, P_meta (BE + EA) = 1.02 × 10⁻⁵, OR = 0.83, CI = 0.79–0.88. The third locus with genome-wide significant association was found at 3p13 (rs2687201) near FOXP1 (Fig. 2c): P_meta (BE + EA) = 5.47 × 10⁻⁹, OR = 1.18, CI = 1.12–1.25.

A previous study of Barrett’s esophagus identified the rs9936833 SNP, near the putative tumor suppressor gene FOXP1 (ref. 14). A subset of the BEACON samples from the present study (2,398 Barrett’s esophagus cases and 2,167 controls) was used in the replication analysis of rs9936833 (P = 5.13 × 10⁻⁴, OR = 1.16, CI = 1.07–1.27). With the additional samples used here (18 Barrett’s esophagus cases and 1,042 controls), the P value for association was more significant (P_discovery (BE) = 1.69 × 10⁻⁴), but there was no change in the odds ratio (OR = 1.16, CI = 1.07–1.26). This SNP was associated with esophageal adenocarcinoma (P_discovery (EA) = 2.06 × 10⁻³, OR = 1.16, CI = 1.05–1.27) (Table 2).

Examining the regional association plot for chromosome 16 near rs9936833 for the combined data (Fig. 2d), we identified four nearby SNPs that had more significant P values than rs9936833 in the combined Barrett’s esophagus and esophageal adenocarcinoma data set (P_discovery (BE + EA) and SNPs that were the most significant SNP in conditional analysis in this region after fitting other SNPs (Table 3).
The nearest gene to the peak SNP on chromosome 3 (rs2687201) is FOXP1. The transcription factors FOXP1 and FOXP2 cooperatively regulate lung and esophageus development, and FOXP1 is a therapeutic target in cancer27,28. The FOX family is overexpressed in esophageal cancer14. There are several SNPs in high LD ($r^2 > 0.8$) with rs2687201 that lie within enhancer histone marks. One of these, rs7626449, is at a site where there is also evidence from DNase I hypersensitive site sequencing of transcription factor binding in esophageal epithelial cells29.

The rs11789015 SNP lies in an intron of BARX1, which encodes a homeobox transcription factor whose homolog is known to be involved in esophageal and trachea differentiation in developing mouse embryos and is associated with the downregulation of Wnt pathway activity in stomach morphogenesis and specification30. The BARX1 promoter region is hypermethylated in gastric cancer cell lines and patient samples, BARX1 mRNA expression is reduced in gastric cancer tissues and cell lines (S.J. Meltzer, personal communication), rs11789015 lies in a region where histone marks denote likely promoter activity, rs11789015 also alters a known regulatory motif for the transcription factor FOXP1. A correlated SNP, rs62574346 ($r^2 = 0.97$ with rs11789015), resides where there is also evidence from DNase I hypersensitive site sequencing of transcription factor binding in esophageal epithelial cells30.

A subset of the BEACON data presented here (Supplementary Fig. 3a) formed part of the replication arm of a recent Barrett’s esophagus genome-wide association study (GWAS)14. A primary finding from that work was an association with Barrett’s esophagus at 16q24 involving rs9936833. Here we found clear evidence that this locus is also associated with esophageal adenocarcinoma ($P_{\text{discovery (EA)}} = 2.06 \times 10^{-3}$, OR = 1.16, CI = 1.05–1.27), as did a recent small study (316 esophageal adenocarcinoma cases and 602 controls; OR = 1.21, CI = 0.99–1.47)31. Two other SNPs near rs9936833, rs2178146 and rs3111601, have stronger and more significant associations in esophageal adenocarcinoma cases (Table 2 and Supplementary Fig. 3b).

Because the size and direction of the effects for the Barrett’s esophagus–associated SNPs at 16q24 were similar in esophageal adenocarcinoma, we used the combined Barrett’s esophagus and adenocarcinoma data to identify other SNPs that are more significantly associated at 16q24 than rs9936833 (Table 2). One of these was rs3111601, which was
in high LD \((r^2 = 0.75)\) with rs9936833. All of the SNPs in high LD with rs3111601 were intergenic, although rs1979654 \((r^2 = 0.64\) with rs3111601) stood out as having excellent regulatory potential across a wide range of cell types and is likely to affect protein binding, chromatin structure and histone modification.\(^9\) There was evidence for additional independent signals in the region at rs3950627 (38 kb closer to FOXI1) and rs2178146 (64 kb closer to FOXI1), as both had similar association \(P\) values to that of rs3111601 (Fig. 2d and Table 2).

**Table 3** | Stepwise conditional analysis to test for independent SNP signals in the 16q24 region

| SNP(s)       | rs3950627 | rs3950627 + rs2178146 | rs3950627 + rs2178146 + rs3111601 + rs1490865 |
|--------------|-----------|-----------------------|-----------------------------------------------|
| rs3950627    | 1.18      | 1.18                  | 1.18                                          |
| rs2178146    | 0.84      | 0.88                  | 0.88                                          |
| rs3111601    | 1.19      | 1.13                  | 1.13                                          |
| rs1490865    | 1.05      | 1.10                  | 1.10                                          |
| rs3332095    | 1.21      | 1.20                  | 1.20                                          |
| rs9936833    | 1.16      | 1.10                  | 1.10                                          |

Each SNP is the most significant SNP 100 kb up- or downstream of rs9936833 after fitting all SNPs in the rows above as additional covariates in the same logistic regression model used in the primary analysis. Starting with the most significant SNP in this region, rs3950627, this methodology fit four other SNPs (rows 2–5) before stopping when the \(P\) value of the most significant remaining SNP was >0.01. The last row is the \(P\) value of an association test of rs9936833 in each of these models.
but were in only modest LD \( (r^2 < 0.2) \) with rs3111601. Neither is a good regulatory candidate, although rs8045253 (imputation association \( P_{\text{Discovery}} (\text{BE} + \text{EA}) = 8.04 \times 10^{-5} \)), \( r^2 = 0.63 \) with rs3950627, changes a motif for the transcription factor FOXP1. Alteration to the way in which FOXP1 binds to this region is particularly interesting in light of our association findings on chromosome 3.

In summary, we report the first genome-wide association study of esophageal adenocarcinoma and the first to examine this cancer together with its precancerous lesion, Barrett’s esophagus. Consistent with our findings showing extensive polygenic overlap between esophageal adenocarcinoma and Barrett’s esophagus\(^{15} \), our most significant results were for cancer and pre-cancer combined. Together, these findings suggest that much of the genetic basis for esophageal adenocarcinoma lies in the development of Barrett’s esophagus, rather than in progression from Barrett’s esophagus to esophageal adenocarcinoma. We found three new genome-wide significant loci for esophageal adenocarcinoma and Barrett’s esophagus combined and extended existing findings at the FOX1 and HLA loci. One of the newly associated regions is on chromosome 3p13 near FOX1, a gene encoding a transcription factor that regulates esophageal development. Interestingly, two of the other regions (BARX1 at 9q22.32 and FOX1 at 16q24.1) contain tumor-associated SNPs that disrupt binding of FOXP1. Further dissection of these loci is likely to lead to insights into the etiology of this rapidly fatal cancer.

**URLs**  
BEACON, http://beacon.itvet.net; eQTL resources, http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/; GWASTools, http://www.bioconductor.org/packages/devel/bioc/html/GWASTools.html;  
HaploReg, http://www.broadinstitute.org/mammals/haploreg; MaCH, http://www.sph.umich.edu/csg/abecasis/mach/index.html;  
METAL, http://www.sph.umich.edu/csg/abecasis/metal/index.html; R, http://www.R-project.org/; RegulomeDB, http://www.regulomedb.org/index; SNPRelate, http://cran.r-project.org/web/packages/SNPRelate/index.html; Wellcome Trust Case Control Consortium, http://www.wtccc.org.uk/; UCSC Genome Browser, http://genome.ucsc.edu.

**METHODS**

Methods and any associated references are available in the online version of the paper.  
*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

**ACKNOWLEDGMENTS**

We thank P. Christopherson, P. Hansen, L. Nolan and T. Watson for their efforts in project management and the organization of biospecimens and data, and we thank Cathy Laurie, Cecelia Laurie and B. Weir for helpful suggestions regarding analyses.

This work was primarily funded by a US National Institutes of Health (NIH) grant (R01CA136725) to T.L.V. and D.C.W. T.L.V. is also supported by grant R01CA133996 and the UTMADCC NIH SPORE in Melanoma 2P50CA093459, as well as by the Marit Peterson Fund for Melanoma Research. CIDR is supported by contract HHSN268200702696C from the US NIH.

The study was generated by the Welcome Trust Case Control Consortium. Funding for the project was provided by the Wellcome Trust under award 076113. A full list of the investigators who contributed to the generation of the data is available from the website (see URLs).

The Romero Registry is supported in part by the American Digestive Health Foundation Endoscopic Research Award, the American College of Gastroenterology Junior Faculty Development Award, the Glaxo Wellcome, Inc., Institute for Digestive Health Clinical Research Award and the Miles and Shirley Fiterman Center for Digestive Diseases at the Mayo Clinic (Rochester, Minnesota). The Romero Registry also receives charitable gifts from five industry partners (Afferm, AstraZeneca, Santarus, Takeda and Wyeth). The US Multicenter Study was funded by U01CA57949, U01CA57983 and U01CA57923 from the US NIH.

Members of the Romero Registry: Sami R. Achem (Mayo Clinic, Jacksonville, Florida, USA), Yvonne Romero (Mayo Clinic, Rochester, Minnesota, USA), David A. Ahlquist (Mayo Clinic, Rochester, Minnesota, USA), Steven R. Alberts (Mayo Clinic, Rochester, Minnesota, USA), Jeffrey A. Alexander (Mayo Clinic, Rochester, Minnesota, USA), Mark S. Allen (Mayo Clinic, Rochester, Minnesota, USA), Aminda S. Arora (Mayo Clinic, Rochester, Minnesota, USA), Jonathan B. Ashman (Mayo Clinic, Phoenix, Arizona, USA), Pamela J. Atherton (Mayo Clinic, Rochester, Minnesota, USA), Lisa A. Boardman (Mayo Clinic, Rochester, Minnesota, USA), Ernest P. Bours (Mayo Clinic, Rochester, Florida, USA), Vicki A. Bryhn (Mayo Clinic, Rochester, Minnesota, USA), Patrick A. Burch (Mayo Clinic, Rochester, Minnesota, USA), George E. Burdick (Mayo Clinic, Rochester, Minnesota, USA), Nareti S. Butt (Mayo Clinic, Rochester, Minnesota, USA), John K. Camoriano (Mayo Clinic, Phoenix, Arizona, USA), John R. Cancemi (Mayo Clinic, Jacksonville, Florida, USA), Stephen D. Cassivi (Mayo Clinic, Rochester, Minnesota, USA), Frances K. Cayr (Mayo Clinic, Jacksonville, Florida, USA), Amy C. Clayton (Mayo Clinic, Rochester, Minnesota, USA), Michael D. Crowell (Mayo Clinic, Phoenix, Arizona, USA), Julie M. Cunningham (Mayo Clinic, Rochester, Minnesota, USA), Mariza de Andrade (Mayo Clinic, Rochester, Minnesota, USA), Piet de Groen (Mayo Clinic, Rochester, Minnesota, USA), Giovanni De Petris (Mayo Clinic, Phoenix, Arizona, USA), Claude Deschamps (Mayo Clinic, Rochester, Minnesota, USA), Kenneth R. DeVault (Mayo Clinic, Jacksonville, Florida, USA), Robert B. Diasio (Mayo Clinic, Rochester, Minnesota, USA), John K. DiBaise (Mayo Clinic, Phoenix, Arizona, USA), Eric S. Edell (Mayo Clinic, Rochester, Minnesota, USA), Sharon Elcombe (Mayo Clinic, Rochester, Minnesota, USA), Charles Erlichman (Mayo Clinic, Rochester, Minnesota, USA), Douglas O. Faigl (Mayo Clinic, Phoenix, Arizona, USA), Tom R. Fitch (Mayo Clinic, Phoenix, Arizona, USA), David E. Fleischer (Mayo Clinic, Rochester, Minnesota, USA), Jean C. Fox (Mayo Clinic, Rochester, Minnesota, USA), Prasad A. Iyer (Mayo Clinic, Rochester, Minnesota, USA), Amy E. Fox-Ofenstein (Mayo Clinic, Rochester, Minnesota, USA), Dawn Francis (Mayo Clinic, Rochester, Minnesota, USA), Mary F. Fredericksen (Mayo Clinic, Rochester, Minnesota, USA), Evanthia Galanis (Mayo Clinic, Rochester, Minnesota, USA), Debra M. Geno (Mayo Clinic, Rochester, Minnesota, USA), Axel Grothey (Mayo Clinic, Rochester, Minnesota, USA), Michael G. Haddox (Mayo Clinic, Rochester, Minnesota, USA), Kevin C. Halling (Mayo Clinic, Rochester, Minnesota, USA), Dennis M. Harmo (Mayo Clinic, Jacksonville, Florida, USA), Tracy W. Halton (Mayo Clinic, Jacksonville, Florida, USA), Timothy Hobday (Mayo Clinic, Rochester, Minnesota, USA), Lesley A. Houghton (Mayo Clinic, Jacksonville, Florida, USA), Prasad A. Iyer (Mayo Clinic, Rochester, Minnesota, USA), Dawn E. Jaroszewski (Mayo Clinic, Phoenix, Arizona, USA), Aminah Jatoi (Mayo Clinic, Rochester, Minnesota, USA), Robert B. Jenkins (Mayo Clinic, Rochester, Minnesota, USA), Elizabeth A. Johnson (Mayo Clinic, Jacksonville, Florida, USA), Rajini Katipamula (Mayo Clinic, Rochester, Minnesota, USA), David A. Katska (Mayo Clinic, Rochester, Minnesota, USA), Sharon F. Kauffman (Mayo Clinic, Rochester, Minnesota, USA), Andrew P. Keaveny (Mayo Clinic, Jacksonville, Florida, USA), Daniel A. Keller (Mayo Clinic, Rochester, Minnesota, USA), George P. Kim (Mayo Clinic, Jacksonville, Florida, USA), Benjamin R. Kipp (Mayo Clinic, Rochester, Minnesota, USA), Dora M. Lam-Himlin (Mayo Clinic, Rochester, Minnesota, USA), Stephen M. LaQuaglia (Mayo Clinic, Rochester, Minnesota, USA), Louis Lanza (Mayo Clinic, Phoenix, Arizona, USA), Shauna N. Legrand (Mayo Clinic, Jacksonville, Florida, USA), Paul J. Limburg (Mayo Clinic, Rochester, Minnesota, USA), Wilma L. Lingle (Mayo Clinic, Rochester, Minnesota, USA), Wanguo Liu (Mayo Clinic, Jacksonville, Florida, USA), Haiming Liu (Mayo Clinic, Jacksonville, Florida, USA), Lian Liu (Mayo Clinic, Phoenix, Arizona, USA), David W. Longley (Mayo Clinic, Jacksonville, Florida, USA), John D. Longley (Mayo Clinic, Jacksonville, Florida, USA), Yvonne L. Longley (Mayo Clinic, Jacksonville, Florida, USA), Michael L. Longley (Mayo Clinic, Jacksonville, Florida, USA), Amy A. Loague (Mayo Clinic, Jacksonville, Florida, USA), J. Jeffrey Louie (Mayo Clinic, Jacksonville, Florida, USA), Alan T. Macdonald (Mayo Clinic, Jacksonville, Florida, USA), Peter D. MacIsaac (Mayo Clinic, Jacksonville, Florida, USA), Mary B. Fredericksen (Mayo Clinic, Rochester, Minnesota, USA), Elizabeth A. Johnson (Mayo Clinic, Jacksonville, Florida, USA), Mark S. Allen (Mayo Clinic, Rochester, Minnesota, USA), Andrew J. Antonescu (Mayo Clinic, Jacksonville, Florida, USA), Robert B. Diasio (Mayo Clinic, Rochester, Minnesota, USA), Kevin C. Halling (Mayo Clinic, Rochester, Minnesota, USA), Prasad A. Iyer (Mayo Clinic, Rochester, Minnesota, USA), Amy E. Fox-Ofenstein (Mayo Clinic, Rochester, Minnesota, USA), Dawn Francis (Mayo Clinic, Rochester, Minnesota, USA), Mary F. Fredericksen (Mayo Clinic, Rochester, Minnesota, USA), Evanthia Galanis (Mayo Clinic, Rochester, Minnesota, USA), Debra M. Geno (Mayo Clinic, Rochester, Minnesota, USA), Axel Grothey (Mayo Clinic, Rochester, Minnesota, USA), Michael G. Haddox (Mayo Clinic, Rochester, Minnesota, USA), Kevin C. Halling (Mayo Clinic, Rochester, Minnesota, USA), Dennis M. Harmo (Mayo Clinic, Jacksonville, Florida, USA), Tracy W. Halton (Mayo Clinic, Jacksonville, Florida, USA), Timothy Hobday (Mayo Clinic, Rochester, Minnesota, USA), Lesley A. Houghton (Mayo Clinic, Jacksonville, Florida, USA), Prasad A. Iyer (Mayo Clinic, Rochester, Minnesota, USA), Dawn E. Jaroszewski (Mayo Clinic, Phoenix, Arizona, USA), Aminah Jatoi (Mayo Clinic, Rochester, Minnesota, USA), Robert B. Jenkins (Mayo Clinic, Rochester, Minnesota, USA), Elizabeth A. Johnson (Mayo Clinic, Jacksonville, Florida, USA), Rajini Katipamula (Mayo Clinic, Rochester, Minnesota, USA), David A. Katska (Mayo Clinic, Rochester, Minnesota, USA), Sharon F. Kauffman (Mayo Clinic, Rochester, Minnesota, USA), Andrew P. Keaveny (Mayo Clinic, Jacksonville, Florida, USA), Daniel A. Keller (Mayo Clinic, Rochester, Minnesota, USA), George P. Kim (Mayo Clinic, Jacksonville, Florida, USA), Benjamin R. Kipp (Mayo Clinic, Rochester, Minnesota, USA), Dora M. Lam-Himlin (Mayo Clinic, Rochester, Minnesota, USA), Stephen M. LaQuaglia (Mayo Clinic, Rochester, Minnesota, USA), Louis Lanza (Mayo Clinic, Phoenix, Arizona, USA), Shauna N. Legrand (Mayo Clinic, Jacksonville, Florida, USA), Paul J. Limburg (Mayo Clinic, Rochester, Minnesota, USA), Wilma L. Lingle (Mayo Clinic, Rochester, Minnesota, USA), Wanguo Liu (Mayo Clinic, Jiangxi,
The authors declare no competing financial interests.

D.C.W. drafted the manuscript. All authors critically revised the manuscript for L.O., P.D.P. and R.Z. analyzed data. W.E.E., R.C.F., D.M.L., S.M., L.O., T.L.V. and D.C.W., A.H.W. and W.Y. recruited participants. P.H., P.L.-S. and C.S. processed I.D.-B., R.C.F., M.D.G., L.J.H., N.K.H., G.L., L.J.M., O.N., B.J.R., Y.R., N.J.S., T.L.V., W.Y. planned the study. P.D.P., T.L.V. and D.C.W. obtained funding. C.C., R.C.F., G.L., L.J.M., O.N., P.D.P., B.J.R., H.A.R., Y.R., N.J.S., T.L.V., D.C.W., A.H.W. and W.Y. jointly led the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jochem, V.J., Fuest, P.A. & Fromkes, J.J. Familial Barrett’s esophagus associated with adenosarcoma. Gastroenterology 102, 1400–1402 (1992).
2. Crabb, D.W. et al. Familial gastroesophageal reflux and development of Barrett’s esophagus. Ann. Intern. Med. 103, 52–54 (1985).
3. Poynton, A.R., Walsh, T.N., O’ Sullivan, G. & Hennessey, T.P. Carcinoma arising in familial Barrett’s esophagus. Am. J. Gastroenterol. 91, 1855–1856 (1996).
4. Romero, Y. et al. Familial aggregation of gastroesophageal reflux in patients with Barrett’s esophagus and esophageal adenocarcinoma. Gastroenterology 113, 1449–1456 (1997).
5. Melzer, L., Shiroyerman, R., Appelman, Z. & Kashian, H. Familial Barrett’s oesophagus: another risk factor? Scand. J. Gastroenterol. 40, 1127–1128 (2005).
6. Groves, C., Jankowski, J., Barker, F. & Holdstock, G. A family history of Barrett’s oesophagus: another risk factor? Am. J. Gastroenterol. 94, 1172–1178 (1999).
7. Chak, A. et al. Identification of Barrett’s esophagus in relatives by endoscopic screening. Am. J. Gastroenterol. 99, 2107–2114 (2004).
8. Fitzgerald, R.C. Complex diseases in gastroenterology and hepatology: GERD, Barrett’s, and esophageal adenocarcinoma. Clin. Gastroenterol. Hepatol. 3, 529–537 (2005).
9. Amos, C.I. et al. Genome-wide association study identifies novel loci predisposing to cutaneous melanoma. Hum. Mol. Genet. 20, 5012–5023 (2011).
10. Gogarten, S.M. et al. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics 28, 3329–3331 (2012).
11. Strain, A. et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat. Genet. 42, 985–990 (2010).
12. Miller, C.J., Li, Y. & Abecasis, G.R. META: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
13. Su, Z. et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett’s esophagus. Nat. Genet. 44, 1131–1136 (2012).
14. Eck, W.E. et al. Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett’s esophagus and gastroesophageal reflux. J. Natl. Cancer Inst. (in the press).
15. Gu, Y. et al. Altered LKB1/CREB-regulated transcription co-activator (CRTC) signaling axis promotes esophageal cancer cell migration and invasion. Oncogene 31, 469–479 (2012).
16. Montgomery, S.B. et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773–777 (2010).
17. Cheung, L.W.T. et al. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. Cancer Discov. 1, 170–185 (2011).
18. Cortés, I. et al. p53 phosphoisoform 3-kinase subunit regulates tumor progression. Proc. Natl. Acad. Sci. USA 109, 11318–11323 (2012).
19. Venturi, S. & Venturi, M. Idiocrine in evolution of salivary glands and in oral health. Nutr. Health 20, 119–134 (2009).
20. Jankowski, J., Hopwood, D. & Wormley, K.G. Expression of epidermal growth factor, transforming growth factor α and their receptor in gastro-oesophageal diseases. Dig. Dis. 11, 1–11 (1993).
21. Menke, V. et al. Functional single-nucleotide polymorphism of epidermal growth factor is associated with the development of Barrett’s esophagus and esophageal adenocarcinoma. J. Hum. Genet. 57, 26–32 (2012).
22. Lanuti, M. et al. A functional epidermal growth factor (EGF) polymorphism, EGF serum levels, and esophageal adenocarcinoma risk and outcome. Clin. Cancer Res. 14, 3216–3222 (2008).
23. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
24. Elks, C.E. et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat. Genet. 42, 1077–1085 (2010).
25. Albanjosa, J.J. et al. The Creb1 coactivator Crtc1 is required for energy balance and fertility. Nat. Med. 14, 1112–1117 (2008).
26. Koon, H.R., Ippolito, G.C., Banham, A.H. & Tucker, P.W. FOXF1: a potential therapeutic target in cancer. Expert Opin. Ther. Targets 11, 955–965 (2007).
27. Shu, W. et al. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. Development 134, 1991–2000 (2007).
28. Boyle, A.P. et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22, 1790–1797 (2012).
29. Woo, J., Miletich, I., Kim, B.-M., Sharpe, P.T. & Shivdasani, R.A. Barx1-mediated inhibition of Wnt signaling in the mouse thoracic foregut controls trache-esophageal septation and epithelial differentiation. PLoS ONE 6, e22493 (2011).
30. Dura, P. et al. Barrett associated MHC and FOXF1 variants also increase esophageal carcinoma risk. Int. J. Cancer 133, 1751–1755 (2013).
ONLINE METHODS

Discovery. Study subjects. Cases of Barrett's esophagus and esophageal adenocarcinoma, together with associated population controls, were collected by investigators in the BEAGESS consortium. A subset of these individuals with European ancestry from epidemiological studies conducted in Western Europe, Australia and North America over the past 20 years was used in the Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study (BEAGESS). To increase the statistical power of the study, we included additional controls from a hospital-based case-control study of melanoma. These controls (MD Anderson controls) were cancer-free friends or acquaintances of European ancestry who had accompanied patients with melanoma to their clinical visits at the MD Anderson Cancer Center in Houston, Texas. The distribution of samples by study is given in Supplementary Table 1.

Histological confirmation of esophageal adenocarcinoma was carried out for all esophageal adenocarcinoma studies. Similarly, Barrett’s esophagus was histologically confirmed via identification of goblet cells in metaplastic columnar epithelium in a biopsy taken from the esophagus. Age, sex (Supplementary Table 2) and other esophageal adenocarcinoma/Barrett’s esophagus risk factors were collected by all of the included studies via standardized questionnaires, usually through personal interviews. All recruited participants gave informed consent, and this project was approved by the ethics boards of each participating institution.

Genotyping. BEAGESS specimens were shipped to the Fred Hutchinson Cancer Research Center (Seattle, Washington) where they were processed and genotyped in three batches. In each batch, samples on each genotyping plate were stratified and balanced according to case-control status, study and sex, with samples assigned to plates randomly within these strata. Genotyping of DNA from buffy coat or whole blood was performed using the Illumina HumanOmni1-Quad platform. MD Anderson controls were genotyped using the Illumina HumanOmni1-Quad platform at the Johns Hopkins University Center for Inherited Disease Research (CIDR). SNP annotations were based on version H of the Illumina product files and corresponded to the Genome Reference Consortium GRCh37 release.

Quality control. Quality assurance and quality control for the BEAGESS and MD Anderson data sets were carried out independently by the Genetics Coordinating Center at the University of Washington following standard procedures. Quality assurance and quality control for the MD Anderson data set was described previously. BEAGESS samples with call rate < 95%, admixture of more than one DNA source or unexpected relatedness (including unexpected duplicates) or misannotated sex that could not be explained were removed from the data set. We looked for batch and plate effects using intensity data and allelic frequency and checked for case-control associations with different experimental factors. No important batch or plate effects or case-control associations with experimental factors were found. We used heterozygosity, sex chromosome intensity data, identity-by-descent (IBD) analysis and visualization of B allele frequency (BAF) and log R ratio (LRR) plots to identify samples that had one or more of misannotated sex, unexpected relatedness or were sample mixtures. Two additional sample mixtures were removed from the data set. After further sample filtering to retain only unrelated samples with European ancestry (see next section), 2,416 Barrett’s esophagus cases, 1,516 esophageal adenocarcinoma cases and 2,187 controls remained. These samples were combined with 1,022 controls of European ancestry from the MD Anderson data set for discovery analysis.

SNPs were clustered using Illumina’s GenomeStudio software, and SNP clusters were defined using all samples with call rate > 95%. SNPs that had either GenTrain score < 0.6 or cluster separation value < 0.4 had their genotypes set to missing. Additionally, we filtered out SNPs that were intensity only, had missing call rate > 5%, had Hardy–Weinberg equilibrium P value over controls ≥ 10−4, had one or more discordances among any of the duplicate pairs or had a mendelian error in either BEAGESS families or HapMap trios. These filters were combined, with similar filters calculated for the MD Anderson data set.

Additionally, we removed a further 344 SNPs that were discordant in the same HapMap control samples (n = 3) run in both the BEAGESS and MD Anderson data sets. After quality assurance and quality control, a total of 926,923 SNPs remained for analysis.

PCA. We performed PCA as a two-step process using SNPRelate software. First, we used PCA to define a homogeneous set of samples with European ancestry in the BEAGESS data set. We did this by running PCA on a set of 6,248 unrelated subjects (with the exception of six two-person families) each of whom was an esophageal adenocarcinoma case, a Barrett’s esophagus case or a control. A majority of these subjects (~98%) self-identified as ‘white’, and a scatter plot of all subjects along the axes of the first two eigenvectors showed that the majority of samples formed a tight cluster (Supplementary Fig. 4). Therefore, we computed the means and standard deviations of the first two eigenvectors and defined any sample that fell within a 2-s.d. rectangle of both eigenvector means to be of homogeneous European ancestry (n = 6,125).

Second, we ran PCA on the combined data set (n = 7,147) consisting of the BEAGESS samples of European ancestry (n = 6,125) and the similarly defined set of MD Anderson controls (n = 1,022). The intent was to identify eigenvectors to include as covariates in our model to adjust for population differences that were present in the remaining European-ancestry-only samples. For this analysis, we selected 65,774 SNPs that were non-monomorphic, autosomal, passed quality control, had missing call rate < 5%, had minor allele frequency > 5%, did not have LD value > 0.2 between any two SNPs in a sliding window of 500,000 bases and were not in the LCT gene (2q21), HLA region or polymorphic regions on chromosomes 8 (8p23) and 17 (17q21.31). We included the first four eigenvectors as covariates in the association test model to account for population stratification by ancestry because they were significantly correlated with case-control status, and a scatter plot showed that the variance accounted for by each eigenvector flattened out after these four eigenvectors were included (data not shown). To check that only genome-wide variation was detected, we computed the absolute value of the correlation coefficient of each eigenvector against the genotypes for each SNP. We observed one small region of high correlation (p = 0.4) between the first eigenvector and chromosome 2, which may indicate long-range LD with the LCT gene.

Statistical analysis. After excluding six related samples and six other samples that had missing call rate > 2%, we ran a case-control analysis of the remaining 7,135 samples: 3,928 cases (1,514 esophageal adenocarcinoma and 2,414 Barrett’s esophagus) versus 3,207 controls. We used an additive logistic regression model with case status regressed on each SNP’s genotype score (coded as 0, 1 or 2 for BB, AB and AA, respectively), including as covariates age, sex and the first four PCA eigenvectors, to compute the OR and 95% CI values relating risk of esophageal adenocarcinoma or Barrett’s esophagus to a given SNP variant. To test SNPs on the X chromosome, male genotypes were coded as 0 and 2, and female genotypes were coded as 0, 1 and 2. After filtering out SNPs that did not pass quality control and SNPs with minor allele frequency < 1%, the A value was 1.04. The quantile-quantile plot is shown in Supplementary Figure 2. We also compared Barrett’s esophagus and esophageal adenocarcinoma cases separately against the controls using the same model. The corresponding Manhattan and quantile-quantile plots are shown in Supplementary Figures 1 and 2, respectively. Analysis was carried out in the R statistical programming language using the Bioconductor packages GWASTools and SNPRelate.

Using the combined Barrett’s esophagus and esophageal adenocarcinoma discovery data set, we performed a stepwise series of nested logistic regression analyses to test the independence of the associations at 16q24 near rs9936833. We used the same logistic regression model and covariates as in our primary analysis and also fitted rs3950627 as a covariate because it was the most significant SNP 100 kb up- or downstream of rs9936833. This conditional analysis identified rs2178146 as the most significant SNP 100 kb up- or downstream of rs9936833. We repeated this analysis four more times, identifying and adding to each successive model rs2178146, rs3111601, rs1490865 and rs13332095, respectively, stopping when the P value of the most significant remaining SNP was > 0.01 (Table 3).

Imputation. To assess the impact of variants not present on the Illumina HumanOmni1-Quad chip, we imputed genotypes using MaCH software and a European reference panel from the 1000 Genomes project. At each region in Tables 1 and 2, imputation was carried out in two steps. First, haplotypes were estimated in a prephasing step. Second, missing alleles for additional SNPs were imputed onto these phased haplotypes using Minimap2 and a publicly available reference panel of haplotypes from populations of European ancestry. SNPs with very different allele frequencies (χ² statistic > 40 in a test for difference in allele frequency) in the BEACON data and the reference panel
were removed before the second step. SNPs with MACH imputation $r^2 < 0.3$ (a measure of imputation quality) and SNPs with minor allele frequency <1% were also removed. Association analysis between imputed SNPs and disease status was performed using the same regression model as for genotyped SNPs but with dosage probabilities as predictors instead of the actual genotype calls. All association tests were two-sided. LD calculations ($r^2$) were performed with the discovery data when the two SNPs being compared were both genotyped; otherwise, samples of European ancestry from Phase 1 of the 1000 Genomes Project were used.

**Replication. SNP selection.** We selected SNPs after quality assurance and quality control for replication that had P value from discovery <1 × 10^{-4} and minor allele frequency >1%. This yielded 406 SNPs: 179 from Barrett’s esophagus cases, esophageal adenocarcinoma cases and controls, 105 from esophageal adenocarcinoma cases and controls and 122 from Barrett’s esophagus versus controls; 321 were unique. A subset of these SNPs (n = 111) was selected via LD pruning with PLINK using the command clump; if a SNP had LD > 0.5 with any other SNP(s) in the list, the SNP with the lowest P value was selected for replication. For each of the ten SNPs on this list with the smallest P value, we selected an additional ‘proxy’ SNP to include in case the top SNP was not successfully genotyped in the replication set. These proxy SNPs were in high LD with the top SNP but had less significant P values. We visually examined cluster plots of all SNPs and kept only those that were highly significant SNPs. SNPs were ranked ordered by P value, and replication was attempted for the top 94.

**Study subjects.** The replication cohort consisted of Barrett’s esophagus cases, esophageal adenocarcinoma cases and controls. Barrett’s esophagus cases were identified at endoscopy with a confirmed histopathological diagnosis of intestinal metaplasia from the UK Barrett’s Esophagus Gene Study. Esophageal adenocarcinoma cases were selected from SOCS and had an International Classification of Diseases coding of malignant neoplasm of the esophagus (C15) and a pathological diagnosis of adenocarcinoma. One set of controls came from the SEARCH study, which ascertains eligible cases of breast, ovarian, prostate, colorectal, melanoma and endometrial cancer from the UK Eastern Cancer Registration and Information Centre. Controls were ascertained by frequency matching on the basis of age (5-year age bands) and sex to the esophageal adenocarcinoma and Barrett’s esophagus cases, excluding individuals with a past history of cancer (excluding non-melanoma skin cancer). All recruited participants gave informed consent, and the studies have been approved by the relevant institutional ethics review board. The other set of controls was from WTCCC2.

**Genotyping.** Barrett’s esophagus cases, esophageal adenocarcinoma cases and SEARCH controls were genotyped using the Fluidigm high-throughput platform and Fluidigm 96.96 Dynamic Arrays according to the manufacturer’s instructions and were read using the Fluidigm EP1. Each array is capable of running 96 samples against 96 SNP assays. Cases and controls were plated in sets of 96 samples and combined into 384-well arrays for genotyping, with the cases and controls mixed on each 384-well plate. Genotypes were automatically called using BioMark Genotyping Analysis software, but all cluster plots were also checked manually and adjusted as needed. The WTCCC2 controls were genotyped on a custom version of the Illumina Human1.2M-Duo array.

**Quality control.** We filtered out Barrett’s esophagus cases, esophageal adenocarcinoma cases and SEARCH controls that had low call rate, had inconsistent sex, were duplicates or had self-reported ancestry of ‘non-white’ or ‘missing’. This left 759 Barrett’s esophagus cases, 874 esophageal adenocarcinoma cases and 1,711 SEARCH controls. We excluded SNPs with missing call rate ≥25%, with significant differential missing call rates in cases and controls (P < 5 × 10^{-4}), with low minor allele frequency (defined as <1%), and with significant departure from Hardy-Weinberg equilibrium (P < 0.0005), retaining 87 SNPs after quality control. We applied standard sample and SNP exclusion criteria to the WTCCC2 controls, retaining 5,190 samples of European ancestry after quality control. There were 67 SNPs after quality control in the WTCCC2 controls in common with the 87 SNPs after quality control in the Barrett’s esophagus cases, esophageal adenocarcinoma cases and SEARCH controls.

**Statistical analysis.** Each of the 87 SNPs was run using an additive logistic regression model with case status regressed on SNP genotype and including sex as a covariate. The analysis focus was on the comparison of Barrett’s esophagus and esophageal adenocarcinoma cases against controls, but we also ran each case type separately against the controls. The final data set used for replication consisted of 759 Barrett’s esophagus cases and 874 esophageal adenocarcinoma cases. For 67 SNPs, the control set consisted of 6,911 samples: 1,711 SEARCH controls and 5,200 WTCCC2 controls. For 20 SNPs that were not genotyped in the WTCCC2 data, only the 1,711 SEARCH controls were used. The R statistical programming language was used for all analyses.

**Meta-analysis.** We used the inverse variance–based method in METAL software to perform a meta-analysis of the discovery and replication data sets. This approach weights the effect size estimates (β coefficients from the discovery and replication regression models) by their standard error estimates and calculates an overall z score and P value. This analysis was carried out separately for each sample set.

**Bioinformatics and functional genomics.** Each region of interest was interrogated using the tools eQTL browser, HaploReg, RegulomeDB and the UCSC Genome Browser.

32. Laurie, C.C. et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34**, 591–602 (2010).
33. Peiffer, D.A. et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. * Genome Res.* **16**, 1136–1148 (2006).
34. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. * Nat. Genet.* **38**, 904–909 (2006).
35. Novembre, J. et al. Genes mirror geography within Europe. * Nature* **456**, 98–101 (2008).
36. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2006).
37. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. * Bioinformatics* **28**, 3326–3328 (2012).
38. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
39. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).