Pure

University of Aberdeen

eQTL set-based association analysis identifies novel susceptibility loci for Barrett's esophagus and esophageal adenocarcinoma

Wang, Xiaoyu; Gharahkhani, Puya; Levine, David M; Fitzgerald, Rebecca C; Gockel, Ines; Corley, Douglas A; Risch, Harvey A; Bernstein, Leslie; Chow, Wong-Ho; Onstad, Lynn; Shaheen, Nicholas J; Lagergren, Jesper; Hardie, Laura J; Wu, Anna H; Pharoah, Paul D P; Liu, Geoffrey; Anderson, Lesley A; Iyer, Prasad G; Gammon, Marilie D; Caldas, Carlos; Ye, Weimin; Barr, Hugh; Moayyedi, Paul; Harrison, Rebecca; Watson, Rg Peter; Attwood, Stephen; Chegwidden, Laura; Love, Sharon B; MacDonald, David; deCaestecker, John; Prenen, Hans; Ott, Katja; Moebus, Susanne; Venerito, Marino; Lang, Hauke; Mayershofer, Rupert; Knapp, Michael; Veits, Lothar; Gerges, Christian; Weissmüller, Josef; Reeh, Matthias; Nöthen, Markus M; Izbicki, Jakob R; Manner, Hendrik; Neuhaus, Horst; Rösch, Thomas; Böhmer, Anne Christin; Hölscher, Arnulf H; Anders, Mario; Pech, Oliver; Schumacher, Brigitte; Schmidt, Claudia; Schmidt, Thomas; Noder, Tania; Lorenz, Dietmar; Vieth, Michael; May, Andrea; Hess, Timo; Kreuser, Nicole; Becker, Jessica; Ell, Christian; Tomlinson, Ian; Palles, Claire; Jankowski, Janusz A; Whiteman, David C; MacGregor, Stuart; Schumacher, Johannes; Vaughan, Thomas L; Buas, Matthew F; Dai, James Y

Published in:
Cancer Epidemiology, Biomarkers and Prevention

DOI:
10.1158/1055-9965.EPI-22-0096

Publication date:
2022

Document Version
Peer reviewed version

Link to publication

Citation for published version (APA):
Wang, X., Gharahkhani, P., Levine, D. M., Fitzgerald, R. C., Gockel, I., Corley, D. A., Risch, H. A., Bernstein, L., Chow, W-H., Onstad, L., Shaheen, N. J., Lagergren, J., Hardie, L. J., Wu, A. H., Pharoah, P. D. P., Liu, G., Anderson, L. A., Iyer, P. G., Gammon, M. D., ... Dai, J. Y. (2022). eQTL set-based association analysis identifies novel susceptibility loci for Barrett's esophagus and esophageal adenocarcinoma. *Cancer Epidemiology, Biomarkers and Prevention*, *31*(9), [cebp.0096]. https://doi.org/10.1158/1055-9965.EPI-22-0096

eQTL set-based association analysis identifies novel susceptibility loci for Barrett's esophagus and esophageal adenocarcinoma

Xiaoyu Wang[1], Puya Gharahkhani[2], David M. Levine[3], Rebecca C. Fitzgerald[4], Ines Gockel[5], Douglas A. Corley[6,7], Harvey A. Risch[8], Leslie Bernstein[9], Wong-Ho Chow[10], Lynn Onstad[1], Nicholas J. Shaheen[11], Jesper Lagergren[12,13], Laura J. Hardie[14], Anna H. Wu[15], Paul D. P. Pharoah[16,17], Geoffrey Liu[18], Lesley A. Anderson[19], Prasad G. Iyer[20], Marilie D. Gammon[21], Carlos Caldas[22], Weimin Ye[12], Hugh Barr[23], Paul Moayyedi[24], Rebecca Harrison[25], RG Peter Watson[26], Stephen Attwood[27], Laura Chegwidden[28], Sharon B. Love[29], David MacDonald[30], John deCaestecker[31], Hans Prenen[32], Katja Ott[33,34], Susanne Moebus[35], Marino Venerito[36], Hauke Lang[37], Rupert Mayershofer[38], Michael Knapp[39], Lothar Veits[40], Christian Gerges[41], Josef Weismüller[42], Matthias Reeh[43], Markus M. Nöthen[44], Jakob R. Izbicki[45], Hendrik Manner[46], Horst Neuhaus[41], Thomas Rösch[47], Anne C. Böhmer[44], Arnulf H. Hölscher[48], Mario Anders[47,49], Oliver Pech[50], Brigitte Schumacher[41,51], Claudia Schmidt[52], Thomas Schmidt[33], Tania Noder[47,] Dietmar Lorenz[53], Michael Vieth[40], Andrea May[54], Timo Hess[55], Nicole Kreuser[5], Jessica Becker[44], Christian Ell[56], Ian Tomlinson[57], Claire Palles[58], Janusz A. Jankowski[59], David C. Whiteman[60], Stuart MacGregor[2], Johannes Schumacher[55], Thomas L. Vaughan[1,61], Matthew F. Buas[62,*], James Y. Dai[1,3,*]

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[2]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

[3]Department of Biostatistics, University of Washington, School of Public Health, Seattle, Washington, USA

[4]Medical Research Council (MRC) Cancer Unit, Hutchison-MRC Research

Centre, University of Cambridge, Cambridge, UK

[5]Department of Visceral, Transplant, Thoracic and Vascular Surgery, University Hospital of Leipzig, Leipzig, Germany

[6]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

[7]San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, California, USA

[8]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA

[9]Department of Population Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, California, USA

[10]Department of Epidemiology, MD Anderson Cancer Center, Houston, Texas, USA

[11]Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

[12]Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

[13]School of Cancer and Pharmaceutical Sciences, King's College London.

[14]Division of Epidemiology, University of Leeds, Leeds, UK

[15]Department of Population and Public Health Sciences, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA

[16]Department of Oncology, University of Cambridge, Cambridge, UK

[17]Department of Public Health and Primary Care, University of Cambridge,

Cambridge, UK

[18]Pharmacogenomic Epidemiology, Ontario Cancer Institute, Toronto, Ontario, Canada

[19]Department of Epidemiology and Public Health, Queen's University of Belfast, Royal Group of Hospitals, Northern Ireland

[20]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota, USA

[21]Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA

[22]Cancer Research UK, Cambridge Institute, Cambridge, UK

[23]Department of Upper GI Surgery, Gloucestershire Royal Hospital, Gloucester, UK

[24]Farncombe Family Digestive Health Research Institute, Department of Medicine, McMaster University, Hamilton, Ontario, Canada

[25]Department of Pathology, Leicester Royal Infirmary, Leicester, UK

[26]Department of Medicine, Institute of Clinical Science, Royal Victoria Hospital, Belfast, UK

[27]Department of General Surgery, North Tyneside General Hospital, North Shields, UK

[28]University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

[29]Centre for Statistics in Medicine and Oxford Clinical Trials Research Unit, Oxford, UK

[30]Department of Oral Biological and Medical Sciences, University of British Columbia, Vancouver, British Columbia, Canada

[31]Digestive Diseases Centre, University Hospitals of Leicester, Leicester, UK

[32]Oncology Department, University Hospital Antwerp, Edegem, Belgium

[33]Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany

[34]Department of General, Visceral and Thorax Surgery, RoMed Klinikum Rosenheim, Rosenheim, Germany

[35]Institute for Urban Public Health, University Hospitals, University of Duisburg-Essen, Essen, Germany

[36]Department of Gastroenterology, Hepatology and Infectious Diseases, Otto-von-Guericke University Hospital, Magdeburg, Germany

[37]Department of General, Visceral and Transplant Surgery, University Medical Center, University of Mainz, Mainz, Germany

[38]Gastroenterologie am Burgweiher, Bonn, Germany

[39]Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany

[40]Institute of Pathology, Friedrich-Alexander-University Erlangen-Nuremberg, Klinikum Bayreuth, Bayreuth, Germany

[41]Department of Internal Medicine, Evangelisches Krankenhaus, Düsseldorf, Germany

[42]Gastroenterologische Gemeinschaftspraxis, Koblenz, Germany

[43]Department of General, Visceral and Thoracic Surgery, Asklepios Harzklinik Goslar, Goslar, Germany

[44]Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany

[45]General, Visceral and Thoracic Surgery Department and Clinic. University Medical Center Hamburg-Eppendorf. Hamburg. Germany

[46]Department of Internal Medicine II, Frankfurt Hoechst Hospital, Frankfurt, Germany

[47]Department of Interdisciplinary Endoscopy, University Hospital Hamburg-Eppendorf, Hamburg, Germany

[48]Clinic for General, Visceral and Trauma Surgery, Contilia Center for Esophageal Diseases. Elisabeth Hospital Essen, Germany

[49]Department of Gastroenterology and Interdisciplinary Endoscopy, Vivantes Wenckebach-Klinikum, Berlin, Germany

[50]Department of Gastroenterology and Interventional Endoscopy, St. John of God Hospital, Regensburg, Germany

[51]Department of Internal Medicine and Gastroenterology, Elisabeth Hospital, Essen, Germany

[52]Department of General, Visceral and Cancer Surgery, University of Cologne, Cologne, Germany

[53]Department of General and Visceral Surgery, Sana Klinikum, Offenbach, Germany

[54]Department of Gastroenterology, Oncology and Pneumology, Asklepios Paulinen Klinik, Wiesbaden, Germany

[55]Center for Human Genetics, University Hospital of Marburg, Marburg, Germany

[56]Department of Medicine II, Sana Klinikum, Offenbach, Germany

[57]Edinburgh Cancer Research Centre, IGMM, University of Edinburgh, UK

[58]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

[59]Comprehensive Clinical Trials Unit, University College London, London, UK

[60]Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, Australia.

[61]Department of Epidemiology, University of Washington, School of Public Health, Seattle, Washington, USA

[62]Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, New York 14263 USA

*Correspondence:

James Y. Dai, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Mail Stop M2-B500, Seattle, WA 98109. Telephone number: 206-667-6364. E-mail address: jdai@fredhutch.org;

Matthew F. Buas, Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263. Telephone number: 716-845-4754. E-mail address: matthew.buas@roswellpark.org

Disclosure of Interest: The authors report no conflict of interest.

Abstract

Background:

Over 20 susceptibility single-nucleotide polymorphisms (SNPs) have been identified for esophageal adenocarcinoma (EAC) and its precursor, Barrett's esophagus (BE), explaining a small portion of heritability.

Methods:

Using genetic data from 4,323 BE and 4,116 EAC patients aggregated by international consortia including the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON), we conducted a comprehensive transcriptome-wide association study (TWAS) for BE/EAC, leveraging Genotype Tissue Expression (GTEx) gene expression data from six tissue types of plausible relevance to EAC etiology: mucosa and muscularis from the esophagus, gastroesophageal (GE) junction, stomach, whole blood, and visceral adipose. Two analytical approaches were taken: standard TWAS using the predicted gene expression from local expression quantitative trait loci (eQTLs), and set-based *SKAT* association using selected eQTLs that predict the gene expression.

Results:

While the standard approach did not identify significant signals, the eQTL set-based approach identified eight novel associations, three of which were validated in independent external data (eQTL SNP sets for *EXOC3*, *ZNF641* and *HSP90AA1*).

Conclusions:

This study identified novel genetic susceptibility loci for EAC and BE using an eQTL set based genetic association approach.

Impact:

This study expanded the pool of genetic susceptibility loci for EAC and BE, suggesting the potential of the eQTL set based genetic association approach as an alternative method for TWAS analysis.

Introduction

Incidence of esophageal adenocarcinoma (EAC) has risen sharply over recent decades(1-4), and it is now the predominant subtype of esophageal cancer in the US and many other western countries. Patients diagnosed with advanced EAC have a 5-year survival rate below 20%(5-10). Progress has been made in identifying genetic and environmental risk factors for EAC and its epithelial precursor lesion, Barrett's esophagus (BE)(11). GE reflux(12,13), obesity(14,15), and tobacco smoking(16,17), collectively explain up to ~75% of cancer risk(18-20). While over 20 susceptibility single-nucleotide polymorphisms (SNPs) have been identified through genome-wide association studies (GWAS) in the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON) and related efforts(21-29), these loci explain only a small portion of overall heritability (h^2_g estimated as 0.25 for EAC; 0.35 for BE)(30), and few have been linked specifically to progression to cancer(22,29).

One of the notable methodological advances in the post-GWAS era is integrating the transcriptome into genetic association analyses(31,32). Evidence is abundant that trait-associated SNPs are more likely to be expression quantitative trait loci (eQTLs)(33), which are pervasive in the human genome(34-38). Motivated by the premise that eQTL may influence disease phenotypes by altering gene expression levels, association approaches leveraging eQTL and transcriptome data in genotype-tissue expression (GTEx)(38), namely transcriptome-wide association studies (TWAS), have become a mainstream approach in post-GWAS analyses, leading to the discovery of multiple novel susceptibility genes(39-42), for prostate(43), ovarian(44), breast(45), and colorectal cancers(46,47). Notably, the initial TWAS method - *PrediXcan(31)* - first

builds a genetic prediction model for gene expression, then assesses genetically predicted gene expression for its association with a trait of interest, much resembling the classical instrumental variable regression approach in econometrics. In the same vein, newer methods for TWAS(32,48) removed the requirement of individual-level genetic data by exploiting GWAS summary statistics and genetic correlation data from external data such as the 1000 Genomes Project. Recognizing a portion of eQTL regulation of gene expression can be conservative across tissues, new methodological development for TWAS has been focused on leveraging multiple tissues available in GTEx for improving power of genetic prediction and subsequent association(49,50). On the other hand, it has been noted with caution recently that TWAS can be prone to spurious results with expression data from non-trait-related tissues or cell types, and that the best practice may be choosing the most mechanistically related tissue(s) available(51).

In our view, the main challenge to apply TWAS to the EAC genetic research is that there is not yet a large set of BE samples, the mechanistically relevant tissue for EAC development, with both germline genotypes and transcriptome data available for eQTL mapping. Although the inherited genetic component of risk for BE largely coincides with that for EAC(30), the cellular origin of BE remains controversial, with hypotheses ranging from residual embryonic cells at the GE junction to undifferentiated gastric cells in the cardia(52-54). While the GTEx Project collected four upper gastrointestinal tract tissues, including mucosa and muscularis from the esophagus, GE junction, and stomach, a limitation of bulk RNA-sequencing data is that transcriptome profiles of rarer constituent cell types (such as progenitor cells) may not be well delineated.

In this work, we conducted a comprehensive TWAS study for BE/EAC, leveraging six GTEx (V8) tissue types of plausible relevance to EAC etiology: mucosa and muscularis from the esophagus, GE junction, stomach, visceral adipose, and whole blood. Inclusion of the latter two tissues is in recognition that tissues beyond the esophago-gastric mucosa are likely to contribute biologically to the origins of BE/EAC. Abdominal obesity is a risk factor for these conditions, which not only affects reflux severity, but also increases levels of systemic inflammation through release of secreted mediators (55-57). Chronic inflammation is considered an important driver of BE/EAC pathogenesis, and the roles and contributions of circulating and infiltrating immune cells are under active investigation (58). We selected eQTLs collectively predicting RNA-sequencing based expression and built prediction models. The eQTLs predicting protein-coding genes were assessed for gene-level associations with BE/EAC risk using a discovery dataset (BEACON/Cambridge GWAS), and top signals identified were then advanced for evaluation using an independent GWAS dataset from Bonn, Germany. We used two methods to assess gene-set associations for selected eQTL: *i*) standard *PrediXcan (31)*, computing a linear combination prediction of gene expression, and *ii*) the sequence kernel association test (*SKAT)(59)*, testing gene-set association among selected eQTLs that predict gene expressions. Originally developed for rare-variant association tests, *SKAT* was used here to assemble genetic associations from eQTL without using the prediction weights derived from an extant gene-expression dataset, e.g., GTEx. An eQTL-based aggregate association strategy has been reported previously (60), though the previous method used the sum of 1-df chi-square values for the individual eQTLs. The following rationales motivate the eQTL aggregation strategy:

i) the existing GTEx tissue types may not capture the cellular origin of BE; ii) even if the etiologically-relevant cell types were contained in one of GTEx tissues, genetically predicted gene expression derived from bulk tissue RNA-seq profiles may not adequately represent the genetic component of gene expression in rarer yet etiologically-relevant cell types within that tissue. Therefore, we hypothesize that the gene-expression prediction weights for eQTLs derived from GTEx may not be always appropriate for the targeted genetic risk prediction for BE and EAC, and we postulate that a more flexible set-based global test of selected eQTLs may improve the likelihood of capturing genetic associations with disease risk which are otherwise obscured when evaluating surrogate gene expression measures from bulk tissue.

Materials and Methods

Individual-level data and summary statistics from existing GWAS

Genome-wide association data from three genetic studies were obtained for this analysis. Given that our analytic plan encompassed a multitude of correlated analyses and included exploratory methodologic comparisons, e.g., two analysis strategies for six tissues and 3 trait comparisons of interest (BE vs control, EAC vs control and BE/EAC vs control), a discovery-validation approach was adopted to better control the potential false positive results. For the discovery set, individual-level genotype data were available from the BEACON consortium (dbGaP phs000869.v1.p1) (2,413 BE cases, 1,512 EAC cases and 6,718 control participants) and the Cambridge GWAS (873 BE cases, 995 EAC cases, and 3,408 control participants); for validation, SNP summary statistics were available from the Bonn GWAS (1,037 BE cases, 1,609 EAC cases, and 3,537 control participants). After quality control, the discovery set included 702,492

SNPs on autosomal chromosomes. An additional 4,541 controls of European ancestry were obtained from the database of Genotypes and Phenotypes (dbGaP) (phs000187.v1.p1, phs000196.v2.p1, and phs000524.v1.p1) and merged with the BEACON discovery data to increase statistical power to detect risk loci. The Michigan imputation server (61) was used to impute genotype data on chromosomes 1-22, with the most accurate and largest panel - the Haplotype Reference Consortium (HRC) (Version r1.1 2016) for European (EUR) as the population reference. Imputed genotype data included 5,312,829 SNPs with imputation quality score > 0.4, MAF> 0.05, call rate > 95% and Hardy-Weinberg equilibrium P-value > 1e-5. For the Bonn dataset, imputation was previously carried out using the 1000 Genomes Phase1 EUR reference panel, and imputed genotype data included a total of ~9 million SNPs with minor allele frequency > 0.001.

GTEx germline sequencing data and RNA-seq transcriptome data for eQTL prediction of gene expression

GTEx data (V8) from subjects of European ancestry were used in this analysis. RNA-seq gene expression data were retrieved from 6 tissues of plausible biologic relevance to EAC development (esophagus GE junction: n=275, esophagus - mucosa: n=411, esophagus - muscularis: n=385, stomach: n=260, adipose - visceral: n=393, and whole blood: n=558). Transcripts per million (TPM) data were downloaded, and the trimmed mean of M values (TMM) normalization method was implemented in *edgeR(62)*. For each gene in a tissue, gene expression values were standardized across samples. SNP genotypes were obtained from whole genome sequencing data for ~46,569,000 variants.

The expression levels for a gene were modeled using an *ElasticNet* linear model with local SNPs in a 1Mb region flanking the TSS of the gene, and covariates including the top four genotype principal components, top 15 Probabilistic Estimation of Expression Residuals (PEER) factors, sex, age, sequencing platform indicator (Illumina HiSeq 2000 or HiSeq X), and sequencing protocol indicator (PCR based or PCR-free). The elastic net model was implemented using the R package *glmnet(63)*. Highly correlated SNPs with Pearson correlation >0.9 were removed before running the elastic net model. The penalty parameter was selected by the minimum ten-fold cross validation error. The ten-fold cross-validated R^2 for genetically predicted gene expression was used to summarize the strength of genetic prediction. The distribution of R^2 for predicting gene expressions in a tissue is displayed by violin plot. Genes with estimated R^2 >0.01 (correlation >0.1) for a tissue entered subsequent genetic association analysis, using the SNPs with non-zero estimated coefficients identified as eQTLs.

eQTL set-based association analysis in the discovery set (Beacon and Cambridge individual-level data)

For each of six tissues and three trait comparisons (BE vs Control, EAC vs Control, BE/EAC combined vs Control), gene-set association analyses were conducted by the following two approaches. First, in the standard TWAS approach, predicted gene-expression from the GTEx-derived *ElasticNet* model was assessed for its association with the trait by a logistic model, adjusting for sex, age, and the top six genotype principal components. Second, the selected eQTLs from the GTEx-derived *ElasticNet* model were assessed for their collective association with the trait by *SKAT(59)*,

adjusting for the same set of covariates. Manhattan plots were drawn to show p-values for gene SNP sets by chromosome. False discovery rate (Benjamini-Hochberg FDR) was used for to account for multiple testing. For genes of interest for discovery, individual SNP-trait associations were also assessed and plotted using *LocusZoom* software. To determine whether an identified gene-set association is caused by a previously identified risk SNP in the neighborhood, a *SKAT* model was also fitted to include the known GWAS SNP in the region.

Validation of the discovered eQTL set based associations in the Bonn dataset

Gene-level eQTL SNP sets putatively associated with a trait were next evaluated using Bonn GWAS summary data. The *SKAT* association statistics for the gene sets were approximated by a score-statistic method(64), using univariate summary statistics and the genetic correlation matrix computed from European ancestry participants of the 1000 Genomes Project. For a few SNPs in the discovery set but missing in the validation set due to different imputation panels, we used the closest SNPs within 50 bp and with correlation > 0.6, whenever available, as the proxy to minimize the impact of missing SNPs. To account for multiple testing in the validation stage, the Hochberg adjusted p-value for controlling family-wise error rate was used.

Data availability

The BEACON data with supplemented controls were obtained from dbGaP (phs000869.v1.p1, phs000187.v1.p1, phs000196.v2.p1, and phs000524.v1.p1). The GTEx genotype and gene expression data were obtained from dbGaP (phs000424.v8.p2).

Results

cis-eQTL predicted gene expressions in six etiologically relevant tissues

Transcriptome data and germline whole-genome sequencing data from GTEx (V8) were analyzed for building genetic prediction models for gene expressions in each of six etiologically relevant tissues for BE/EAC: esophageal mucosa (n=411), esophageal muscularis (n=385), GE junction (n=275), stomach (n=260), adipose – visceral (n=393), and whole blood (n=558). Common SNPs (MAF>0.05) located within ± 500kb of the transcription starting site (TSS) of a gene were identified from GTEx whole-genome sequencing data and selected to predict the transcript abundance by the *ElasticNet* method. Figure 1a shows the violin plots of R^2 for genes with at least 1 SNP being selected and $R^2 \geq 0.01$ (correlation of observed and predicted gene expression ≥0.1) in the six tissues. The four tissues in the upper GI tract (esophageal mucosa and muscularis, GE junction, and stomach) have a greater number of predictable genes and higher R^2 in this subset: esophageal mucosa has the largest number of genes with $R^2 \geq 0.01$ (n=7463); GE junction has the highest median (0.037) despite the smaller sample size for junction tissue. There is substantial variability among the numbers of "genetically predictable" genes across tissues (5160 in blood ~ 7463 in esophageal mucosa), and the genes shared between tissues. The latter is exemplified by a Venn diagram in Figure 1b, which shows the overlapping set between the three esophageal tissues. Between any two tissue types, 35~45% of genes are not shared, underscoring the significance of both cross-tissue and tissue-specific genetic regulation of gene expression.

eQTL set based association analysis identified susceptibility loci for BE and EAC in BEACON/Cambridge discovery set

The selected eQTLs in the genetic prediction models were assessed for association with BE, EAC, or BE/EAC as a combined trait in BEACON/Cambridge discovery set, using two methods: 1) the standard TWAS approach, where the predicted gene expression from the *ElasticNet* model was assessed for its association with the trait in a logistic regression model; 2) the gene-set association method *SKAT*, using the selected eQTLs from the *ElasticNet* model. Across six tissues and for three trait association comparisons, there are a total of 116,853 gene set associations being tested. Because of high correlations between p-values from the same genes across different tissues and trait comparisons, the Bonferroni procedure can be overly conservative. Instead, false discovery rate (FDR) was used to adjust for multiple testing, in part because it can effectively account for correlation. Figure 2 shows the Manhattan plots of p-values for the three comparisons using the two methods (standard TWAS on the bottom of each panel and *SKAT*-eQTL on the top). No genes analyzed by the standard TWAS approach satisfied FDR<0.05 (minimum FDR=0.187, e.g., *EXOC3*, *BARX1*, and *LDAH*). In contrast, the *SKAT*-eQTL method identified a total of twenty-one genes with significant associations at FDR<0.05 (red dotted line in Figure 2), representing a mix of novel and known loci. Table 1 shows eight novel eQTL set-based associations in six loci that either have not been reported previously or are independent of the known GWAS SNP in conditional analysis. Table 2 shows thirteen loci that have been previously linked to susceptibility, containing putative genes including *LDAH*, *BARX1*, *ALDH1A2*, and *CRTC1*.

One consistent theme in Table 1 and 2 is that *SKAT*-eQTL produced uniformly smaller p-values than the standard TWAS method for these eQTL set associations, suggesting that the weighted linear combination of eQTL for predicting these gene expressions may not always be powerful to capture genetic associations. For example, *EXOC3* is located in locus 5p15.33, 220kb away from the known risk SNP rs9918259, with its flanking region containing 28 selected *cis*-eQTLs in adipose tissue, explaining 4.2% of variability in its gene expression. When assessed by *SKAT*, this set of SNPs was significantly associated with BE/EAC with p-value 8.24×10^{-6} (FDR=0.0365). The standard TWAS analysis yielded a larger p-value, 1.81×10^{-5}. Figure 3 shows the regional plots of the three novel loci that were discovered in BEACON/Cambridge discovery set and validated in Bonn data. Specifically, Figure 3a shows a cluster of *cis*-eQTLs, located in the *EXOC3* gene, that were individually associated with BE/EAC at a moderate level of significance (p-value 10^{-2} ~10^{-4}). Previous meta-analysis identified rs9918259 as a risk SNP in *CEP72/TPPP*. Adjusting for rs9918259 in the *SKAT* regression model attenuated the p-value for *EXOC3* gene set association from 8.24×10^{-6} to 2.98×10^{-3}, suggesting that the SNP set of eQTL predicting *EXOC3* gene expression may add new evidence for association at this locus.

The remaining seven eQTL sets in Table 1 are all >10 Mb away from the closest existing GWAS SNP; adjusting for existing GWAS SNPs did not reduce the gene-set association significance. Of top interest is *HSP90AA1(65,66)*, which is located on chromosome 14, with no GWAS risk SNPs previously identified. This gene was identified by association of 94 eQTLs in blood with BE. The regional plot in Figure 3e shows widespread individual eQTL associations over a 1-Mb window around this gene,

exemplifying the power of gene-set association in aggregating signals that may not reach genome-wide significance individually. Four eQTLs of *ZNF641(67)* gene were identified in esophageal junction, collectively associated with BE/EAC (p-value=3.81×10^{-6}), and also individually associated with BE/EAC with less significance (Figure 3e). Table 2 shows loci identified by eQTL gene-set association analyses which have been previously linked to risk of BE and EAC. All are located within 0.5 Mb of an existing GWAS SNP; all but three (*CRTC1, SSBP4* and *JUND*) became non-significant when adjusting for the closest GWAS SNP(s). Six eQTL sets including *CRTC1, SSBP4* and *JUND* are from 19q13.11, a locus harboring three risk SNPs in *CRTC1* that have been consistently detected in previous GWAS efforts. The top association in this locus is *THEM161A* (p=9.02×10^{-10}). Adjusting for the three known risk SNPs in the region (rs10423674, rs10419226, rs199620551) does not completely remove association significance for *CRTC1*, *JUND*, and *SSBP4*, suggesting that this region may have independent risk alleles other than the three known risk SNPs. The locus 3q27.1 contains rs9823696, the only previous risk SNP linked to EAC but not BE. Rather than *HTR3C* and *ABCC5*, the nearest genes to this GWAS variant, gene-set analysis implicated *YEATS2* and *ABCF3*. The remaining genes in Table 2 were reported in prior GWAS as candidate risk genes based on their proximity to index SNPs: *LDAH/GDF7* (2p24.1), *BARX1* (9q22.32), and *ALDH1A2*/*AQP9* (15q21.3).

Three eQTL set associations in Table 1 were replicated using Bonn GWAS data

We evaluated all association signals identified in Table 1 using the *SKAT* method and summary statistics from the Bonn GWAS (with 1000 Genomes EUR LD structure) (Table 3). Because the trait association analyses were based on HRC imputation, and

the available Bonn summary statistics were imputed based on 1000 Genomes data (V3), a small number of SNPs were missing for some of the eight genes. eQTL SNP sets for three genes had evidence of replication in the Bonn data, using the Hochberg adjusted p-value <0.05 as the threshold for confirming candidate associations: *EXOC3*, *ZNF641*, and *HSP90AA1*. In particular, *EXOC3* (p=0.0000185) and *ZNF641* (p=0.00378) remained significant even using the stringent Bonferroni correction. Twelve out of 26 eQTLs for *EXOC3* had a univariate p-value <0.05 (minimum p-value 1.69×10^{-7}). Three out of four eQTLs for *ZNF641* also have univariate p-values <0.05.

Discussion

Advanced esophageal adenocarcinoma is a deadly disease with rising incidence in Western countries. International efforts including BEACON and other European studies have identified ~20 susceptibility SNPs, though collectively these risk SNPs explain only a small portion of the genetic heritability. Polygenic risk scores (PRS) based on GWAS hits have not been able to significantly improve prediction beyond environmental risk factors. In our view, current genetic studies have reached a plateau in discovery, largely due to the rarity of the cancer and the limited available sample sizes. In this work we conducted TWAS for EAC and BE, using data from six etiologically relevant GTEx tissues. The standard TWAS method using predicted gene expression was compared to a novel approach that assessed gene-level eQTL-set associations by *SKAT*. Individual-level genetic data from BEACON/Cambridge were used as the discovery set and summary statistics of genetic data from Bonn were used as the validation set. Using a

significance threshold of FDR<0.05 in the discovery set, standard TWAS identified no associations, while the *SKAT* approach yielded 13 eQTL set associations in 11 loci. Among them, 8 eQTL set associations (5 loci) are novel findings, either representing novel susceptibility regions without previously identified risk SNPs, or in the case of *EXOC3*, a novel signal independent of known risk SNPs in the neighborhood. Among the genes from the known loci, our results suggest that there are potentially susceptibility genes at 19p13.11 independent of the three known risk SNPs. Notably, the loci identified by eQTL set associations largely did not overlap with loci previously reported for gastro-oesophageal reflux disease (GERD), a major risk factor for BE/EAC and a clinical trait by itself. For eQTLs in Tables 1 and 2, only one eQTL rs9636202 (*ISYNA1*) was found to be a GERD risk locus previously reported (68). These results represent a significant advance in identifying novel inherited genetic risk associations for BE/EAC, since the publication of the first GWAS meta-analysis in 2016 (22). While we underscore the importance of discovering new candidate risk loci for a rare cancer with limited study samples available, we also acknowledge that caution is needed in interpreting eQTL set associations. Functional laboratory studies are essential to identify causal variants and genes that are driving observed associations.

One interesting observation is that there is no finding from the standard approach, while *SKAT-eQTL* produced 13 significant eQTL set associations. All set associations in Tables 1 and 2 have smaller p-values from *SKAT-eQTL*, some substantially more significant (e.g., *HSP90AA1* and *KRTAP5-8*, for which p-values of the standard TWAS are greater than 0.05). This analysis is a single observation of applying two methods applied to GWAS datasets, therefore does not establish the power comparison between

the two methods. A formal power comparison using simulated genetic datasets is necessary and will be pursued in future work. Here we merely conjecture why an eQTL-set ensemble association may be advantageous in the context of BE/EAC studies. The prevailing theory for TWAS using imputed gene expression is that the genetic component of gene expression for genes relevant to disease etiology can be accurately predicted, using etiologically relevant tissue samples. As noted previously, however, a complicating factor is that most tissues are comprised of multiple cell types in varying abundance, and only a subset of these cell types – often representing a small fraction of overall cells – may be the most relevant to disease development. This indeed may be the case for BE/EAC, as candidates for the cell-of-origin include subpopulations of precursors in the GE junction or gastric cardia. That said, cancer development is a multi-faceted process involving not only stem-like precursor cells, but also the tissue microenvironment, comprised of and influenced by multiple constituent cell types. It is quite possible that for this reason, a more flexible gene-set association method, such as *SKAT*, may outperform standard TWAS. If this hypothesis is true, one may envision that the SKAT-eQTL association approach could be applied successfully on a wider scale and similarly help accelerate discovery of novel loci for other types of cancers.

Compared to the standard GWAS analysis for individual SNP associations, the eQTL set based aggregation approach provides better power in detecting loci that contain multiple eQTL association signals. In Supplementary Figure 1, we show the locus zoom plots for the individual SNP associations combining the discovery and validation set from the eight loci identified in Table 1. None of the SNPs in these regions reached the p-value cutoff of 5×10^{-8}. However, the set-based method can assemble and

detect multiple eQTL associations in a locus with single variants reaching moderate levels of significance.

Rarely employed in TWAS, the discovery-validation design we adopted is partly driven by the multitude of analyses we intended to conduct using two analytical approaches and 6 potentially relevant tissues, and for 3 trait comparisons of interest, as well as the lack of access to individual-level genetic data in Bonn study. The latter is a limitation of current analysis. Confirmation of discovery-stage associations, using an independent validation dataset, reduces the possibility of false positive discoveries. As noted before, though motivated by the goal of deciphering causal pathways of disease etiology, the original TWAS approach is not immune to spurious findings. This is particularly true for the *SKAT-eQTL* approach – it is essentially a gene-set association method assembling risk associations that are individually unlikely to survive the multiple-testing penalty. Three novel eQTL set associations (*EXOC3*, *SENP6*, *HSP90AA1*) were validated by the Bonn summary data, each of which has multiple individual risk SNPs with a moderate level of association (Figure 3).

We end our discussion with several other limitations and future work. First, while delivering several novel susceptibility signals, the power of our TWAS is limited by the sample size of BE/EAC cases due to its rarity. Compared to existing TWAS for other cancers, our sample size is much smaller. New population-based genetic studies are needed to improve power and further advance EAC genetic research. Second, our analyses were restricted to European ancestry participants. Although the incidence of EAC in whites is much higher than that in African Americans, future studies are needed for eQTL and TWAS in broader populations. Third, although our genetic findings are

promising, experimental studies are needed to understand the mechanisms of genetic risk predisposition.

Acknowledgements

J.Y. Dai and X. Wang were supported by the National Cancer Institute (NCI) grant R01 CA222833. M.F. Buas was supported by NCI grant R03CA223731. Computation is in part supported by National Institutes of Health grant S10OD028685 to Fred Hutch Cancer Research Center. A full list of acknowledgments for consortia data used for this study is provided in the Supplementary Data.

References

1. Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. Ann Oncol **2012**;23(12):3155-62 doi 10.1093/annonc/mds181.
2. Holmes RS, Vaughan TL. Epidemiology and pathogenesis of esophageal cancer. Semin Radiat Oncol **2007**;17(1):2-9 doi 10.1016/j.semradonc.2006.09.003.
3. Steevens J, Botterweck AA, Dirx MJ, van den Brandt PA, Schouten LJ. Trends in incidence of oesophageal and stomach cancer subtypes in Europe. Eur J Gastroenterol Hepatol **2010**;22(6):669-78 doi 10.1097/MEG.0b013e32832ca091.
4. Coleman HG, Xie SH, Lagergren J. The Epidemiology of Esophageal Adenocarcinoma. Gastroenterology **2018**;154(2):390-405 doi 10.1053/j.gastro.2017.07.046.
5. Njei B, McCarty TR, Birk JW. Trends in esophageal cancer survival in United States adults from 1973 to 2009: A SEER database analysis. J Gastroenterol Hepatol **2016**;31(6):1141-6 doi 10.1111/jgh.13289.
6. Gavin AT, Francisci S, Foschi R, Donnelly DW, Lemmens V, Brenner H*, et al.* Oesophageal cancer survival in Europe: a EUROCARE-4 study. Cancer Epidemiol **2012**;36(6):505-12 doi 10.1016/j.canep.2012.07.009.
7. Launoy G, Bossard N, Castro C, Manfredi S, Group GE-W. Trends in net survival from esophageal cancer in six European Latin countries: results from the SUDCAN population-based study. Eur J Cancer Prev **2017**;26 Trends in cancer net survival in six European Latin Countries: the SUDCAN study:S24-S31 doi 10.1097/CEJ.0000000000000308.
8. Edgren G, Adami HO, Weiderpass E, Weiderpass Vainio E, Nyrén O. A global assessment of the oesophageal adenocarcinoma epidemic. Gut **2013**;62(10):1406-14 doi 10.1136/gutjnl-2012-302412.
9. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin **2021**;71(1):7-33 doi 10.3322/caac.21654.
10. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M*, et al.* 2021 SEER Cancer Statistics Review, 1975-2018. <https://seer.cancer.gov/csr/1975_2018/>.
11. Spechler SJ, Souza RF. Barrett's esophagus. N Engl J Med **2014**;371(9):836-45 doi 10.1056/NEJMra1314704.
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):825-31 doi 10.1056/NEJM199903183401101.
13. Cook MB, Corley DA, Murray LJ, Liao LM, Kamangar F, Ye W*, et al.* Gastroesophageal reflux in relation to adenocarcinomas of the esophagus: a pooled analysis from the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON). PLoS One **2014**;9(7):e103508 doi 10.1371/journal.pone.0103508.
14. Kubo A, Cook MB, Shaheen NJ, Vaughan TL, Whiteman DC, Murray L*, et al.* Sex-specific associations between body mass index, waist circumference and the risk of Barrett's oesophagus: a pooled analysis from the international BEACON consortium. Gut **2013**;62(12):1684-91 doi 10.1136/gutjnl-2012-303753.
15. Hoyo C, Cook MB, Kamangar F, Freedman ND, Whiteman DC, Bernstein L*, et al.* Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the International BEACON Consortium. Int J Epidemiol **2012**;41(6):1706-18 doi 10.1093/ije/dys176.

16. Cook MB, Kamangar F, Whiteman DC, Freedman ND, Gammon MD, Bernstein L*, et al.* Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. J Natl Cancer Inst **2010**;102(17):1344-53 doi 10.1093/jnci/djq289.
17. Cook MB, Shaheen NJ, Anderson LA, Giffen C, Chow WH, Vaughan TL*, et al.* Cigarette smoking increases risk of Barrett's esophagus: an analysis of the Barrett's and Esophageal Adenocarcinoma Consortium. Gastroenterology **2012**;142(4):744-53 doi 10.1053/j.gastro.2011.12.049.
18. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL*, et al.* Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst **2003**;95(18):1404-13 doi 10.1093/jnci/djg047.
19. Wang SM, Katki HA, Graubard BI, Kahle LL, Chaturvedi A, Matthews CE*, et al.* Population Attributable Risks of Subtypes of Esophageal and Gastric Cancers in the United States. Am J Gastroenterol **2021**;116(9):1844-52 doi 10.14309/ajg.0000000000001355.
20. Olsen CM, Pandeya N, Green AC, Webb PM, Whiteman DC, Study AC. Population attributable fractions of adenocarcinoma of the esophagus and gastroesophageal junction. Am J Epidemiol **2011**;174(5):582-90 doi 10.1093/aje/kwr117.
21. Su Z, Gay LJ, Strange A, Palles C, Band G, Whiteman DC*, et al.* Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat Genet **2012**;44(10):1131-6 doi 10.1038/ng.2408.
22. Gharahkhani P, Fitzgerald RC, Vaughan TL, Palles C, Gockel I, Tomlinson I*, et al.* Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. Lancet Oncol **2016**;17(10):1363-73 doi 10.1016/S1470-2045(16)30240-6.
23. Buas MF, He Q, Johnson LG, Onstad L, Levine DM, Thrift AP*, et al.* Germline variation in inflammation-related pathways and risk of Barrett's oesophagus and oesophageal adenocarcinoma. Gut **2017**;66(10):1739-47 doi 10.1136/gutjnl-2016-311622.
24. Levine DM, Ek WE, Zhang R, Liu X, Onstad L, Sather C*, et al.* A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. Nat Genet **2013**;45(12):1487-93 doi 10.1038/ng.2796.
25. Palles C, Chegwidden L, Li X, Findlay JM, Farnham G, Castro Giner F*, et al.* Polymorphisms near TBX5 and GDF7 are associated with increased risk for Barrett's esophagus. Gastroenterology **2015**;148(2):367-78 doi 10.1053/j.gastro.2014.10.041.
26. Dai JY, de Dieu Tapsoba J, Buas MF, Onstad LE, Levine DM, Risch HA*, et al.* A newly identified susceptibility locus near FOXP1 modifies the association of gastroesophageal reflux with Barrett's esophagus. Cancer Epidemiol Biomarkers Prev **2015**;24(11):1739-47 doi 10.1158/1055-9965.EPI-15-0507.
27. Dai JY, Tapsoba JeD, Buas MF, Risch HA, Vaughan TL, Consortium B. Constrained Score Statistics Identify Genetic Variants Interacting with Multiple Risk Factors in Barrett's Esophagus. Am J Hum Genet **2016**;99(2):352-65 doi 10.1016/j.ajhg.2016.06.018.
28. Buas MF, Levine DM, Makar KW, Utsugi H, Onstad L, Li X*, et al.* Integrative post-genome-wide association analysis of CDKN2A and TP53 SNPs and risk of esophageal adenocarcinoma. Carcinogenesis **2014**;35(12):2740-7 doi 10.1093/carcin/bgu207.
29. Contino G, Vaughan TL, Whiteman D, Fitzgerald RC. The Evolving Genomic Landscape of Barrett's Esophagus and Esophageal Adenocarcinoma. Gastroenterology **2017**;153(3):657-73.e1 doi 10.1053/j.gastro.2017.07.007.
30. Ek WE, Levine DM, D'Amato M, Pedersen NL, Magnusson PK, Bresso F*, et al.* Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. J Natl Cancer Inst **2013**;105(22):1711-8 doi 10.1093/jnci/djt303.

31. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ*, et al.* A gene-based association method for mapping traits using reference transcriptome data. Nat Genet **2015**;47(9):1091-8 doi 10.1038/ng.3367.
32. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW*, et al.* Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet **2016**;48(3):245-52 doi 10.1038/ng.3506.
33. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet **2010**;6(4):e1000888 doi 10.1371/journal.pgen.1000888.
34. Lappalainen T, Sammeth M, Friedländer MR, t Hoen PA, Monlong J, Rivas MA*, et al.* Transcriptome and genome sequencing uncovers functional variation in humans. Nature **2013**;501(7468):506-11 doi 10.1038/nature12531.
35. Zhang X, Joehanes R, Chen BH, Huan T, Ying S, Munson PJ*, et al.* Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat Genet **2015**;47(4):345-52 doi 10.1038/ng.3220.
36. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K*, et al.* Heritability and genomics of gene expression in peripheral blood. Nat Genet **2014**;46(5):430-7 doi 10.1038/ng.2951.
37. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J*, et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet **2013**;45(10):1238-43 doi 10.1038/ng.2756.
38. Battle A, Brown CD, Engelhardt BE, Montgomery SB, Consortium G, Laboratory DtA*, et al.* Genetic effects on gene expression across human tissues. Nature **2017**;550(7675):204-13 doi 10.1038/nature24277.
39. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. Am J Hum Genet **2017**;100(3):473-87 doi 10.1016/j.ajhg.2017.01.031.
40. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE*, et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet **2016**;48(5):481-7 doi 10.1038/ng.3538.
41. Hauberg ME, Zhang W, Giambartolomei C, Franzén O, Morris DL, Vyse TJ*, et al.* Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. Am J Hum Genet **2017**;100(6):885-94 doi 10.1016/j.ajhg.2017.04.016.
42. Pavlides JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. Genome Med **2016**;8(1):84 doi 10.1186/s13073-016-0338-4.
43. Mancuso N, Gayther S, Gusev A, Zheng W, Penney KL, Kote-Jarai Z*, et al.* Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. Nat Commun **2018**;9(1):4079 doi 10.1038/s41467-018-06302-1.
44. Gusev A, Lawrenson K, Lin X, Lyra PC, Kar S, Vavra KC*, et al.* A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. Nat Genet **2019**;51(5):815-23 doi 10.1038/s41588-019-0395-x.
45. Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J*, et al.* A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet **2018**;50(7):968-78 doi 10.1038/s41588-018-0132-x.
46. Bien SA, Su YR, Conti DV, Harrison TA, Qu C, Guo X*, et al.* Genetic variant predictors of gene expression provide new insight into risk of colorectal cancer. Hum Genet **2019**;138(4):307-26 doi 10.1007/s00439-019-01989-8.

47. Su YR, Di C, Bien S, Huang L, Dong X, Abecasis G, *et al.* A Mixed-Effects Model for Powerful Association Tests in Integrative Functional Genomics. Am J Hum Genet **2018**;102(5):904-19 doi 10.1016/j.ajhg.2018.03.019.
48. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun **2018**;9(1):1825 doi 10.1038/s41467-018-03621-1.
49. Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, *et al.* A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet **2019**;51(3):568-76 doi 10.1038/s41588-019-0345-7.
50. Feng H, Mancuso N, Gusev A, Majumdar A, Major M, Pasaniuc B, *et al.* Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. PLoS Genet **2021**;17(4):e1008973 doi 10.1371/journal.pgen.1008973.
51. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, *et al.* Opportunities and challenges for transcriptome-wide association studies. Nat Genet **2019**;51(4):592-9 doi 10.1038/s41588-019-0385-z.
52. Wang X, Ouyang H, Yamamoto Y, Kumar PA, Wei TS, Dagher R, *et al.* Residual embryonic cells as precursors of a Barrett's-like metaplasia. Cell **2011**;145(7):1023-35 doi 10.1016/j.cell.2011.05.026.
53. Jiang M, Li H, Zhang Y, Yang Y, Lu R, Liu K, *et al.* Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. Nature **2017**;550(7677):529-33 doi 10.1038/nature24269.
54. Nowicki-Osuch K, Zhuang L, Jammula S, Bleaney CW, Mahbubani KT, Devonshire G, *et al.* Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. Science **2021**;373(6556):760-7 doi 10.1126/science.abd1449.
55. Ruhl CE, Everhart JE. Overweight, but not high dietary fat intake, increases risk of gastroesophageal reflux disease hospitalization: the NHANES I Epidemiologic Followup Study. First National Health and Nutrition Examination Survey. Ann Epidemiol **1999**;9(7):424-35 doi 10.1016/s1047-2797(99)00020-4.
56. Dieudonne MN, Bussiere M, Dos Santos E, Leneveu MC, Giudicelli Y, Pecquery R. Adiponectin mediates antiproliferative and apoptotic responses in human MCF7 breast cancer cells. Biochem Biophys Res Commun **2006**;345(1):271-9 doi 10.1016/j.bbrc.2006.04.076.
57. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet **2004**;363(9418):1346-53 doi 10.1016/S0140-6736(04)16044-3.
58. Reid BJ, Li X, Galipeau PC, Vaughan TL. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat Rev Cancer **2010**;10(2):87-101 doi 10.1038/nrc2773.
59. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet **2011**;89(1):82-93 doi 10.1016/j.ajhg.2011.05.029.
60. Ferreira MA, Jansen R, Willemsen G, Penninx B, Bain LM, Vicente CT, *et al.* Gene-based analysis of regulatory variants identifies 4 putative novel asthma risk genes related to nucleotide synthesis and signaling. J Allergy Clin Immunol **2017**;139(4):1148-57 doi 10.1016/j.jaci.2016.07.017.
61. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, *et al.* Next-generation genotype imputation service and methods. Nat Genet **2016**;48(10):1284-7 doi 10.1038/ng.3656.
62. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **2010**;11(3):R25 doi 10.1186/gb-2010-11-3-r25.

63. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw **2010**;33(1):1-22.
64. Hu YJ, Berndt SI, Gustafsson S, Ganna A, Hirschhorn J, North KE*, et al.* Meta-analysis of gene-level associations for rare variants based on single-variant statistics. Am J Hum Genet **2013**;93(2):236-48 doi 10.1016/j.ajhg.2013.06.011.
65. Chu SH, Liu YW, Zhang L, Liu B, Li L, Shi JZ. Regulation of survival and chemoresistance by HSP90AA1 in ovarian cancer SKOV3 cells. Mol Biol Rep **2013**;40(1):1-6 doi 10.1007/s11033-012-1930-3.
66. Eustace BK, Sakurai T, Stewart JK, Yimlamai D, Unger C, Zehetmeier C*, et al.* Functional proteomic screens reveal an essential extracellular role for hsp90 alpha in cancer cell invasiveness. Nat Cell Biol **2004**;6(6):507-14 doi 10.1038/ncb1131.
67. Qi X, Li Y, Xiao J, Yuan W, Yan Y, Wang Y*, et al.* Activation of transcriptional activities of AP-1 and SRE by a new zinc-finger protein ZNF641. Biochem Biophys Res Commun **2006**;339(4):1155-64 doi 10.1016/j.bbrc.2005.11.124.
68. Ong JS, An J, Han X, Law MH, Nandakumar P, Schumacher J*, et al.* Multitrait genetic association analysis identifies 50 new risk loci for gastro-oesophageal reflux, seven new loci for Barrett's oesophagus and provides insights into clinical heterogeneity in reflux diagnosis. Gut **2021** doi 10.1136/gutjnl-2020-323906.

Tables

Table 1. Novel eQTL set-based genetic associations with BE, EAC, or BE/EAC.

Locus	Gene	Tissue	Trait	# eQTLs	R^2	Most significant GWAS SNP	Distance (Mb[a])	TWAS-P[b]	SKAT-P[c]	FDR[d]	P-adj[e]
5p15.33	*EXOC3*	Adipose	BE/EA	28	0.042	rs9918259	0.22	1.81×10^{-5}	8.24×10^{-6}	0.037	2.98×10^{-3}
6q14.1	*SENP6*	Blood	BE/EA	51	0.046	rs76014404	13.92	2.23×10^{-3}	7.12×10^{-6}	0.036	5.32×10^{-6}
11q13.4	*KRTAP5-8*	Adipose	EA	23	0.047	rs4930068	69.26	0.27	1.41×10^{-5}	0.049	3.28×10^{-5}
12q13.11	*ZNF641*	Junction	BE/EA	4	0.014	rs1247942	65.90	5.16×10^{-6}	3.81×10^{-6}	0.025	2.26×10^{-6}
14q32.31	*HSP90AA1*	Blood	BE	94	0.040	--	--	0.66	8.49×10^{-6}	0.037	--
16q23.1	*CFDP1*	Stomach	BE/EA	5	0.011	rs1979654	11.07	3.98×10^{-3}	4.29×10^{-6}	0.026	2.67×10^{-6}
16q23.1	*CHST5*	Junction	BE	21	0.026	rs1979654	10.84	0.36	2.45×10^{-6}	0.022	3.14×10^{-6}
16q23.1	*BCAR1*	Blood	BE	335	0.021	rs1979654	11.14	0.13	6.08×10^{-6}	0.034	1.12×10^{-5}

[a] Distance between the gene and the most significant GWAS risk SNP identified from previous GWAS.

[b] P-value for association analyses in standard TWAS.

[c] P-value for eQTL gene-set SKAT association.

[d] FDR based on the pooled set of p-values for eQTL gene-set SKAT associations across three trait comparisons and six tissues.

[e] P-value derived from the SKAT model adjusting for GWAS risk SNPs.

Table 2. eQTL set-based genetic associations with BE, EAC, or BE/EAC at thirteen risk regions reported in prior GWAS.

Locus	Gene	Tissue	Trait	# eQTLs	R^2	Most significant GWAS SNP	Distance (Mb[a])	TWAS-P[b]	SKAT-P[c]	FDR[d]	P-adj[e]
2p24.1	*LDAH*	Junction	BE/EA	25	0.179	rs7255	0.005	1.94×10^{-4}	7.11×10^{-8}	2.37×10^{-3}	0.429
2p24.1	*GDF7*	Blood	BE/EA	15	0.079	rs3072	0.012	0.146	3.38×10^{-7}	5.67×10^{-3}	0.641
3q27.1	*YEATS2*	Stomach	EA	6	0.018	rs9823696	0.368	0.160	5.18×10^{-6}	0.030	0.942
3q27.1	*ABCF3*	Adipose	EA	43	0.063	rs9823696	0.120	0.223	1.45×10^{-5}	0.049	0.339
9q22.32	*BARX1*	Adipose	BE/EA	5	0.028	rs11789015	0.002	4.35×10^{-6}	3.03×10^{-6}	0.022	0.386
15q21.3	*ALDH1A2*	Adipose	BE/EA	19	0.023	rs66725070	0.022	3.58×10^{-3}	7.87×10^{-7}	0.012	0.506
15q21.3	*AQP9*	Blood	BE/EA	206	0.018	rs2464469	0.068	0.573	8.08×10^{-6}	0.037	0.054
19p13.11	*JUND*	Blood	BE/EA	347	0.011	rs10419226	0.413	0.221	2.62×10^{-6}	0.022	0.041
19p13.11	*SSBP4*	Blood	BE	44	0.011	rs10419226	0.273	0.904	1.82×10^{-6}	0.019	0.006
19p13.11	*ISYNA1*	Blood	BE/EA	206	0.028	rs10419226	0.258	0.679	1.35×10^{-5}	0.046	0.085
19p13.11	*KLHL26*	Blood	BE/EA	32	0.015	rs10419226	0.055	0.212	3.21×10^{-7}	5.67×10^{-3}	0.428
19p13.11	*CRTC1*	Adipose	BE/EA	21	0.061	rs10419226	0.009	0.130	6.61×10^{-8}	2.37×10^{-3}	0.002
19p13.11	*TMEM161A*	Mucosa	BE/EA	4	0.013	rs10423674	0.412	4.25×10^{-3}	9.02×10^{-10}	1.07×10^{-4}	0.203

[a] Distance between the gene and the most significant GWAS risk SNP identified from previous GWAS

[b] P-value derived from association analyses in standard-TWAS

[c] P-value derived from association analyses in SKAT-TWAS

[d] FDR based on p-value derived from association analyses in SKAT-TWAS

[e] P-value derived from association analyses in SKAT-TWAS after adjusting for GWAS risk SNP

Table 3. Association of eQTLs for the eight genes in Table 1 with BE, EAC or BE/EAC risk in Bonn GWAS.

Locus	Gene	#eQTLs	Trait	#snps in Bonn	#snps (P<0.05)[a]	MinP[b]	Bonn_P[c]	Hochberg_P
5p15.33	*EXOC3*	28	BE/EA	26	12	1.69×10^{-7}	1.85×10^{-5}	1.48×10^{-4}
6q14.1	*SENP6*	51	BE/EA	48	3	3.55×10^{-4}	5.01×10^{-2}	0.189
11q13.4	*KRTAP5-8*	23	EA	20	2	8.02×10^{-4}	0.162	0.324
12q13.11	*ZNF641*	4	BE/EA	4	3	4.26×10^{-3}	3.78×10^{-3}	0.026
14q32.31	*HSP90AA1*	94	BE	89	14	0.012	8.28×10^{-3}	0.049
16q23.1	*CFDP1*	5	BE/EA	5	1	0.028	0.053	0.189
16q23.1	*CHST5*	21	BE	20	2	9.81×10^{-3}	0.063	0.189
16q23.1	*BCAR1*	335	BE	328	11	4.63×10^{-3}	0.464	0.464

[a] Number of SNPs with univariate p-value< 0.05 in Bonn data.

[b] Minimum univariate p-value in Bonn data.

[c] Validation SKAT p-value in Bonn data.

Figure Legends

Figure 1. Predictive eQTL models for gene expression across 6 tissues in GTEx. (a) Violin plots for R^2 estimates for genes with $R^2 \geq 0.01$. (b) Venn diagram of genes with $R^2 \geq 0.01$ in three esophageal tissues in GTEx.

Figure 2. Manhattan plots for p-values derived from two methods: eQTL set-based association by SKAT (top); standard TWAS using predicted gene expression (bottom). (a) BE versus control. (b) EA versus control. (c) BE/EA vs control. The line for FDR=0.05 is based on all three trait associations.

Figure 3. Regional plots for novel loci that were discovered in Beacon/Cambridge discovery set and validated in Bonn data. (a) eQTLs of *EXOC3* in discovery. (b) eQTLs of *EXOC3* in validation. (c) eQTLs of *ZNF641* in discovery. (d) eQTLs of *ZNF641* in validation. (e) eQTLs of *HSP90AA1* in discovery. (f) eQTLs of *HSP90AA1* in validation.

Figure 1

a

393 samples 5142 genes | 558 samples 5289 genes | 275 samples 4804 genes | 411 samples 6306 genes | 385 samples 5870 genes | 260 samples 3988 genes

Heritability

0 0.2 0.4 0.6 0.8 1

adipose blood junction mucosa muscularis stomach

b

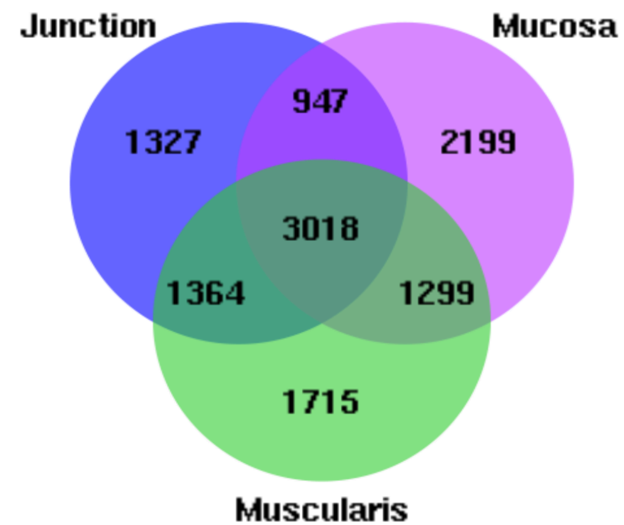
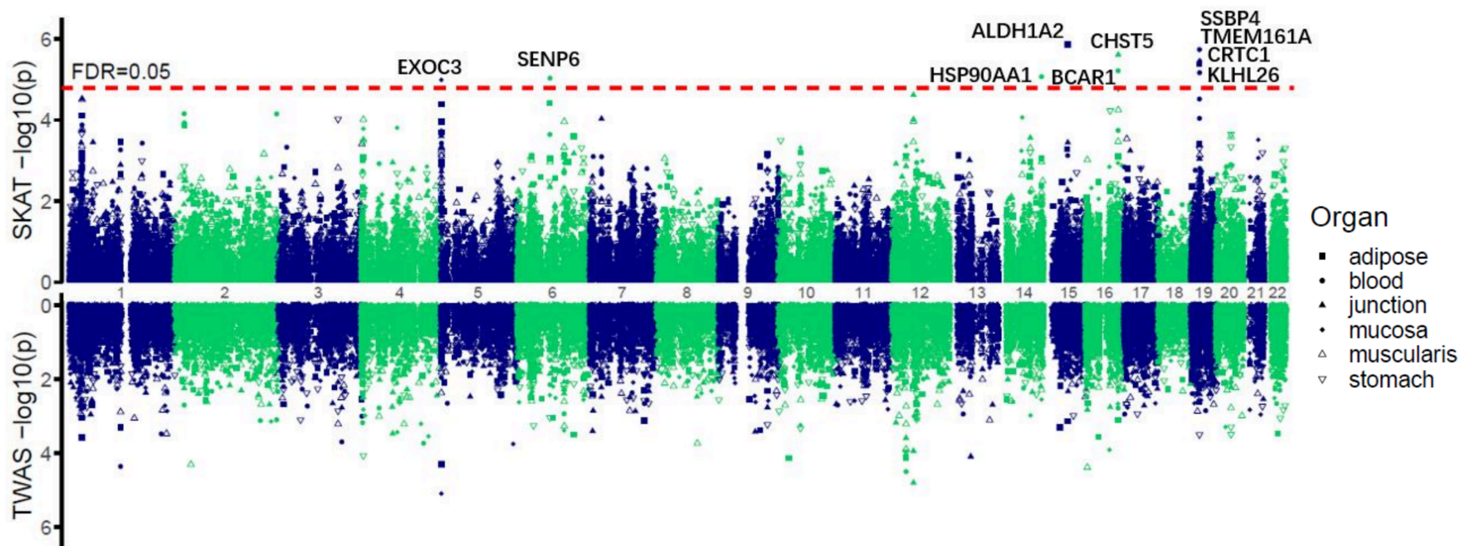
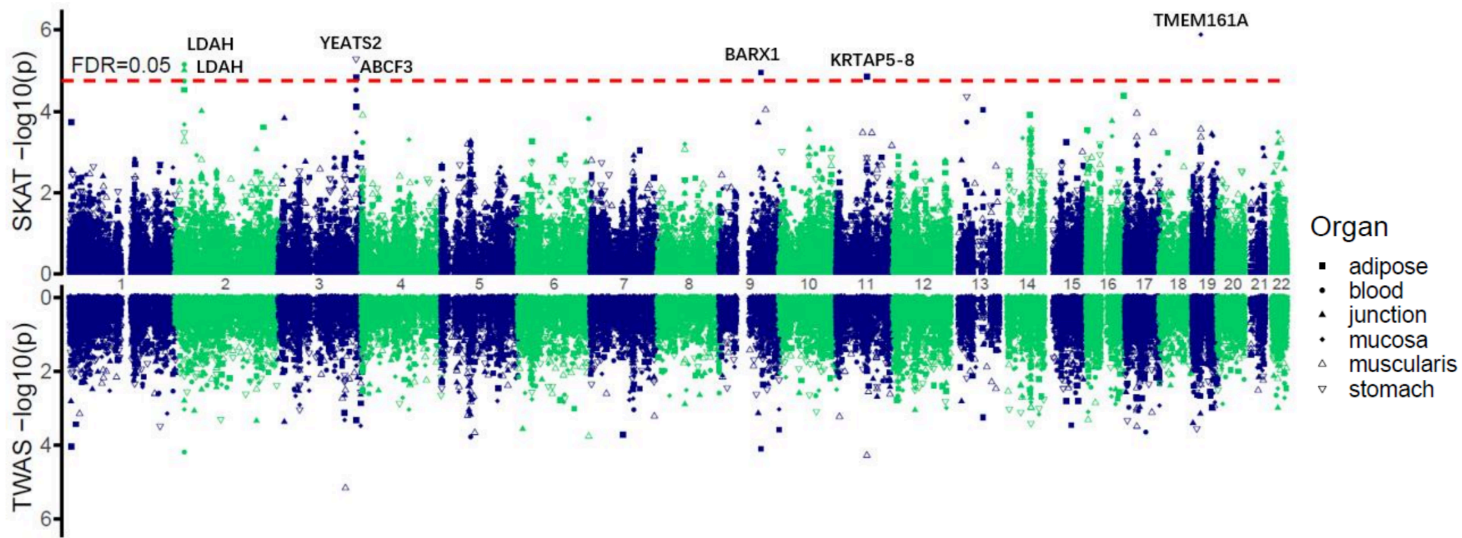


Figure 2

a



b



c

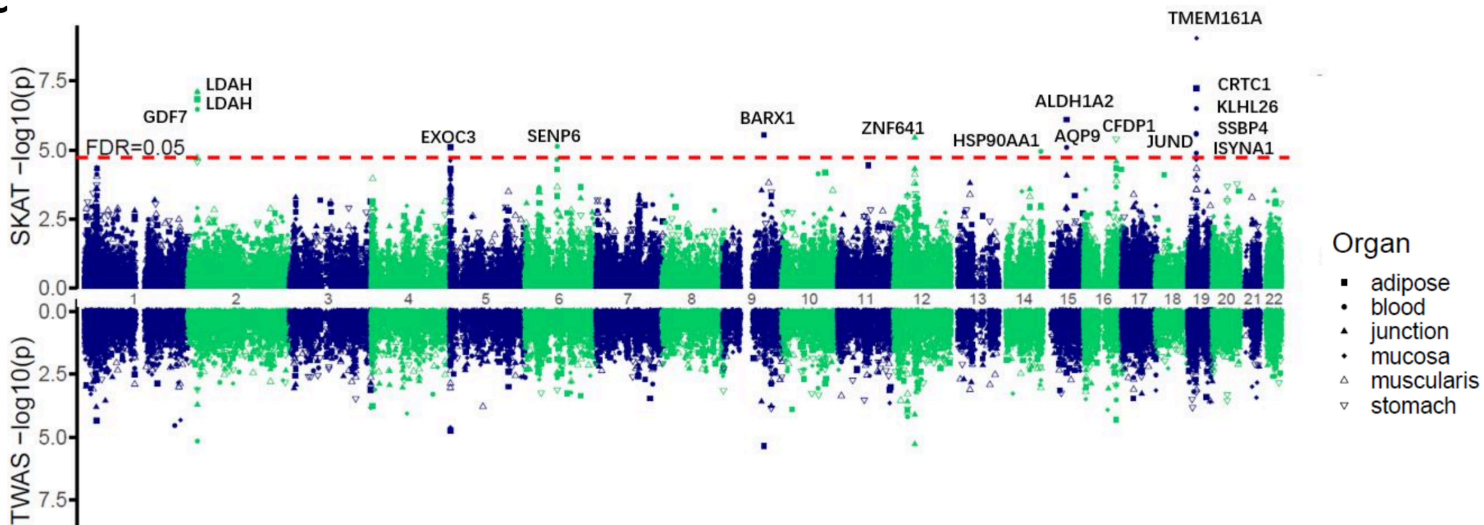


Figure 3