Human Molecular Genetics, 2022, Vol. 31, No. 3 410–422

https://doi.org/10.1093/hmg/ddab259
Advance Access Publication Date: 9 September 2021
General Article

GENERAL ARTICLE

Prioritization and functional analysis of GWAS risk loci for Barrett's esophagus and esophageal adenocarcinoma

Jianhong Chen[1,†], Mourad Wagdy Ali[2,†], Li Yan[3,†], Shruti G. Dighe[1], James Y. Dai[4], Thomas L. Vaughan[4,5], Graham Casey[2,§,*] and Matthew F. Buas[1,§,*]

[1]Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA, [2]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA, [3]Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA, [4]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA and [5]Department of Epidemiology, University of Washington, School of Public Health, Seattle, WA, USA

*To whom correspondence should be addressed. Tel: 716-845-4754; Email: matthew.buas@roswellpark.org (M.F.B) and Tel: 434-297-4433; Email: gcasey@virginia.edu (G.C.)

Abstract

Genome-wide association studies (GWAS) have identified ~20 genetic susceptibility loci for esophageal adenocarcinoma (EAC), and its precursor, Barrett's esophagus (BE). Despite such advances, functional/causal variants and gene targets at these loci remain undefined, hindering clinical translation. A key challenge is that most causal variants map to non-coding regulatory regions such as enhancers, and typically, numerous potential candidate variants at GWAS loci require testing. We developed a systematic informatics pipeline for prioritizing candidate functional variants via integrative functional potential scores (FPS) consolidated from multi-omics annotations, and used this pipeline to identify two high-scoring variants for experimental interrogation: chr9q22.32/rs11789015 and chr19p13.11/rs10423674. Minimal candidate enhancer regions spanning these variants were evaluated using luciferase reporter assays in two EAC cell lines. One of the two variants tested (rs10423674) exhibited allele-specific enhancer activity. CRISPR-mediated deletion of the putative enhancer region in EAC cell lines correlated with reduced expression of two genes—CREB-regulated transcription coactivator 1 (*CRTC1*) and Cartilage oligomeric matrix protein (*COMP*); expression of five other genes remained unchanged (*CRLF1*, *KLHL26*, *TMEM59L*, *UBA52*, *RFXANK*). Expression quantitative trait locus mapping indicated that rs10423674 genotype correlated with *CRTC1* and *COMP* expression in normal esophagus. This study represents the first experimental effort to bridge GWAS associations to biology in BE/EAC and supports the utility of FPS to guide variant prioritization. Our findings reveal a functional variant and candidate risk enhancer at chr19p13.11 and implicate *CRTC1* and *COMP* as putative gene targets, suggesting that altered expression of these genes may underlie the BE/EAC risk association.

[†]Co-first authors
[§]Co-senior authors

Received: June 15, 2021. **Revised:** August 17, 2021. **Accepted:** August 30, 2021

Introduction

Esophageal adenocarcinoma (EAC) is the predominant subtype of esophageal cancer in Western nations and accounts for >10 000 deaths annually in the USA (1–3). Median survival remains less than one year (4). EAC arises from an epithelial precursor lesion, Barrett's esophagus (BE) (5,6), in which the normal stratified squamous epithelium of the lower esophagus is replaced by an intestinalized columnar epithelium. Several strong epidemiologic risk factors for BE/EAC were identified over two decades ago (7–13), including reflux symptoms, obesity and smoking, but an important etiologic role for inherited genetics has only more recently emerged (14).

Genome-wide association studies (GWAS) conducted by the Barrett's and Esophageal Adenocarcinoma Consortium (BEACON) and other collaborating international consortia have now collectively discovered 17 germline susceptibility loci for BE/EAC (15–18), attributing a substantial component of overall disease risk to common variant heritability (h^2_g 25–35%) (19). These studies revealed strong genetic correlation and significant polygenic overlap between BE and EAC, with all but one of the known risk loci common to both traits (15–18). Gene–environment (G × E) interaction scans identified two more loci which appear to modify the degree of risk associated with reflux symptoms (20,21), while targeted post-GWAS analyses have reported several additional candidate risk genes and pathways (22–24).

Virtually no progress has yet been made, however, in the challenging next phase of identifying the functional variants, gene targets and causal mechanisms which underlie risk associations. Consistent with results across GWAS of many complex traits, all genome-wide-significant ('index') BE/EAC variants are located in non-coding regions; 16 of 17 are intergenic or intronic, with one variant mapping to an untranslated region (UTR). Moving from association to biology is complicated by the fact that GWAS index variants themselves are rarely functional and are usually in linkage disequilibrium (LD) with tens or hundreds of other variants, many of which reside in non-coding sequences and lack obvious biological function. GWAS variants are enriched in transcriptional regulatory elements (e.g. enhancers/promoters) (25,26), and increasing evidence suggests that most risk-associated functional/causal variants act through regulatory effects leading to alterations in gene expression (27–30). The greatest challenge to progress in identifying causal variants is that a large number of candidates need to be screened.

In this report, we describe the first systematic effort to bridge BE/EAC GWAS associations to functional variants, enhancers and risk genes—a critical step toward advancing the development of new molecular targets for prevention and therapy. In the first phase of the study, we established a novel and generalizable informatics pipeline for prioritizing candidate risk variants through 'functional potential scores' (FPS). These scores were assembled by consolidating multiple and complementary variant annotations derived from sequence-based features, tissue-specific chromatin profiles and regulatory/expression quantitative trait locus (eQTL) relationships. Next, we leveraged our scoring system to guide the prioritization of a small number of GWAS loci for experimental interrogation. Using luciferase reporter enhancer activity assays, CRISPR genome editing and gene expression profiling, we identified and characterized a functional risk variant (rs10423674) and putative risk enhancer located on chr19p13.11. Further support for candidate target genes at this locus was obtained via integration of publicly available genotype-expression correlations (eQTLs) from normal human tissues and colocalization statistical analyses.

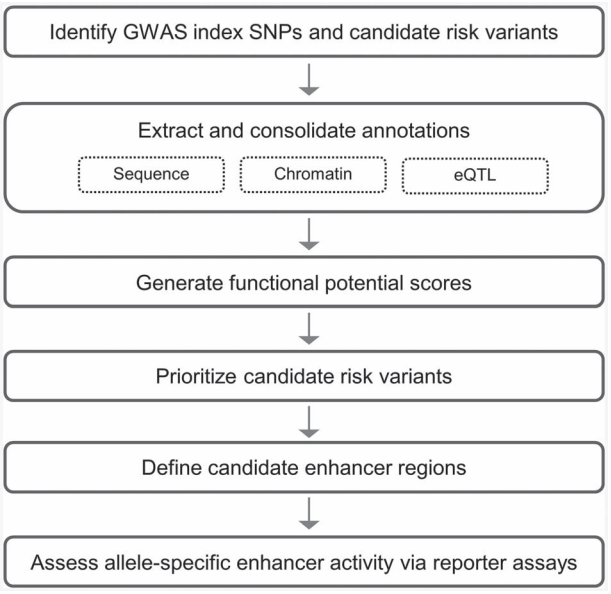

Figure 1. Functional analytic pipeline schematic summary.

Results

Assembly of FPS for candidate risk variants

We assembled a catalogue of all reported BE/EAC GWAS index variants ($P < 5 \times 10^{-8}$) (15–18,20,21). At each risk locus, we defined an associated variant set (AVS), comprised of the index variant(s) and any variants in moderate LD ($r^2 > 0.6$) (Table 1). These sets ranged in size from 3 to 297 variants (median 72, mean 82); the corresponding genomic regions encompassing each AVS ranged from 3.7 to 994 kb (median 69 kb, mean 211 kb).

To aid in the prioritization of GWAS loci and candidate risk variants for experimental follow-up, we constructed a systematic informatics pipeline for consolidating variant functional annotations into FPS. Our pipeline integrates diverse annotations from multiple resources and summarizes the relative estimated functional potential of individual variants under consideration (Fig. 1). This framework allows for selective inclusion and variable weighting of specific features and can be biologically tailored to different traits based on differential inclusion and/or weighting of tissue-specific annotations. Given that BE/EAC arises in the lower esophagus/GE junction (1), we assigned the most weight to annotations specific to, or present in, tissues of the upper gastrointestinal (GI) tract—esophagus and stomach. Lesser weight was assigned for annotations confined to non-GI tissues (Supplementary Material, Table S3). We used our pipeline to derive FPS for the 31 BE/EAC GWAS index variants and >1500 variants in moderate LD. The resulting scoring distribution across all 1565 variants was right-skewed—most variants were relatively low-scoring (FPS < 50: 70%), while progressively smaller fractions of variants accumulated increasing weight from multiple annotations (higher FPS), suggesting stronger functional potential (Fig. 2A). For each of the 19 GWAS loci, we generated a dot plot to visualize the FPS of individual variants in the AVS (Fig. 2B).

Locus prioritization and selection of candidate functional variants at prioritized risk loci

We used these scores to identify a subset of the 19 GWAS risk loci harboring variants with high estimated functional potential. Loci

Table 1. Genome-wide-significant ('index') GWAS variants reported for BE/EAC. Identifiers and genomic locations of 31 index variants, distributed across 19 susceptibility regions. The AVS for each region was defined as all variants in moderate-to-strong LD ($r^2 > 0.6$) with the index variant(s). The size of each region (in kilobases) was defined as the minimum genomic distance covering all SNPs in the AVS

Region	Variant	Chr:position	Nearest gene	AVS (*N*)	AVS (kb)
1	rs3072	2:20878406	*GDF7/C2orf43*	6	3.7
	rs7255	2:20878820			
2	rs139606545	2:200045039	*SATB2*	162	312.7
3	rs2687201	3:70928930	*FOXP1*	91	123.2
	rs2687202	3:70929983			
4	rs9823696	3:183783353	*HTR3C*	41	42.7
5	rs9918259	5:663092	*TPPP/CEP72*	6	5.3
	rs75783973	5:668309			
6	rs9257809	6:29356331	*MHC*	150	994.0
7	rs62423175	6:62195368	*KHDRBS2*	297	780.8
	rs76014404	6:62391538			
8	rs12207195	6:160974578	*LPA*	3	13.6
9	rs11765529	7:52922213	*POM121L12*	121	39.7
10	rs2188554	7:117040117	*CFTR/ASZ1*	76	403.6
	rs17451754	7:117256712			
11	rs17749155	8:10068073	*MSRA*	8	4.6
12	rs2409797	8:11433780	*LINC00208/BLK*	130	720.9
	rs10108511	8:11435516			
13	rs11789015	9:96716028	*BARX1*	75	187.0
14	rs7852462	9:100310501	*TMOD1*	8	5.7
15	rs4930068	11:2297615	*ASCL2*	72	68.8
16	rs1247942	12:114673723	*TBX5*	26	26.7
	rs2701108	12:114674261			
17	rs3784262	15:58253106	*ALDH1A2*	231	183.9
	rs66725070	15:58267416			
	rs2464469	15:58362025			
18	rs1979654	16:86396835	*FOXF1*	23	10.0
	rs9936833	16:86403118			
19	rs10419226	19:18803172	*CRTC1*	39	84.7
	rs199620551	19:18804294			
	rs10423674	19:18817903			

were rank ordered by maximum FPS in the corresponding AVS, and two top-scoring regions were selected for further study: R13 (9q22.32/*BARX1*) and R19 (19p13.11/*CRTC1*). At each locus, we further examined the highest scoring candidate functional variants included the AVS. By jointly considering FPS (and component scores) with r^2 and genomic distance to the index variant(s) (Supplementary Material, Fig. S2), we selected one highly ranked variant from each locus for subsequent experimental testing (rs11789015/9q22.32 and rs10423674/19p13.11).

Allele-specific activity of putative enhancer containing rs10423674

Candidate enhancer regions (CER) at each locus were defined as minimal genomic segments (~1 kb) spanning the prioritized candidate functional/causal variant (Fig. 3). We performed luciferase reporter assays in EAC cell lines (OE19 and OE13) to assess enhancer activity of these selected fragments and determine the effect of the candidate functional variants on such activity. Both candidate enhancers tested exhibited enhancer activity in these assays, and one of the two (CER2/rs10423674/19p13.11) showed allele-specific effects, in both cell lines (Fig. 4). These results support the utility of our FPS-guided approach for prioritizing GWAS risk loci and candidate functional variants for downstream laboratory-based interrogation.

CRISPR-mediated enhancer deletion correlates with alterations in gene expression

To begin evaluating the biological effects of the putative risk enhancer at 19p13.11 (CER2/rs10423674), we used CRISPR genome editing to delete this enhancer region in OE19 and OE13 cells (Fig. 5). We selected the five genes located within ~100 kb of rs10423674 as candidate targets for initial screening—*CRTC1*, *COMP*, *CRLF1*, *TMEM95L*, *KLHL26*—and assessed gene expression levels using TaqMan assays. The deletion correlated with statistically significant reductions in transcript levels of two of the five genes tested—*CRTC1* and *COMP*; expression of the other three genes was not affected (Fig. 6). We also investigated the potential for longer range transcriptional regulation in exploratory studies. Two additional genes located at greater distances from rs10423674 were selected as candidates: *UBA52* (~150 kb 5′) and *RFXANK* (~500 kb 3′). We did not observe significant expression changes for either gene in enhancer-deleted OE19 or OE13 cells (Supplementary Material, Fig. S3).

eQTL colocalization analysis and Hi-C chromatin interactions

To obtain complementary support for candidate target genes of CER2/rs10423674, we first analyzed publicly available eQTL data. eQTLs from the GTEx repository (31) revealed evidence for associations between rs10423674 and RNA expression levels of *CRTC1* and *COMP* in esophageal tissues (Table 2). We performed

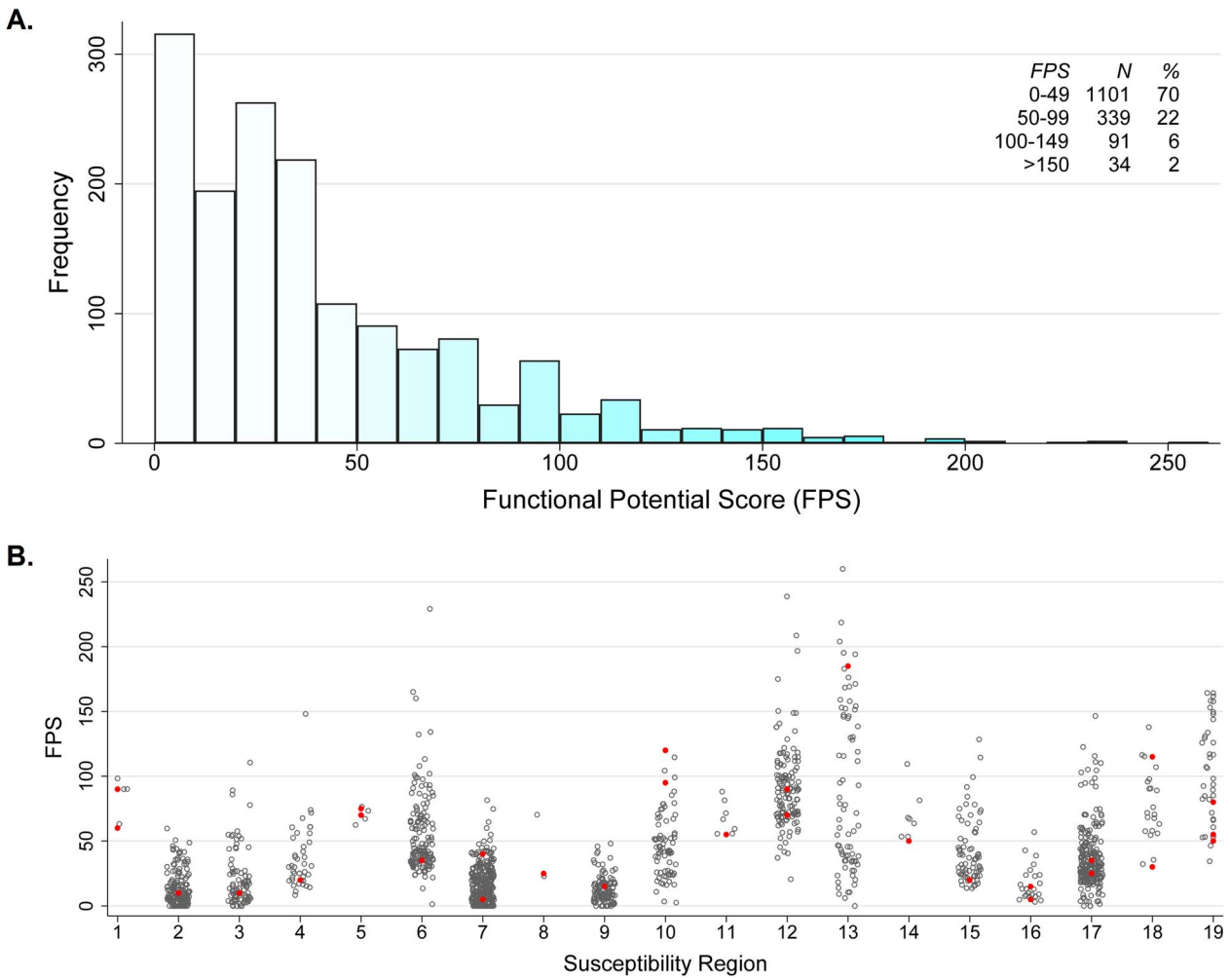

Figure 2. Functional potential scores (FPS). (**A**) Histogram distribution of FPS for 1565 candidate risk SNPs in 19 GWAS susceptibility regions. (**B**) Scatter plot of FPS for individual candidate risk variants in each of these 19 regions. Red, GWAS index SNP.

colocalization analysis to evaluate the posterior probability that a shared causal genetic signal underlies the observed GWAS association and a given eQTL association. Using a standard threshold of PP4 > 0.8 and PP3 < 0.2, our analysis indicated a high probability of colocalized genetic signals for BE/EAC risk and *CRTC1* expression in esophagus muscularis (PP4 = 0.99). Evidence for colocalization was also obtained for *TMEM59L* in the same tissue (PP4 = 0.91). GWAS and eQTL regional association plots were generated for visualization (Supplementary Material, Fig. S4).

Next, we consulted published 3D chromatin interaction datasets to assess the likelihood of physical contact between CER2 and regional gene promoters. Using promoter-capture Hi-C profiles generated in >25 human primary tissues (32), we extracted all chr19 fragments spanning variant rs10423674 and identified all interacting (promoter) fragments reported as statistically significant (FDR < 0.09). These data suggested the presence of 3D chromatin loops between the putative risk enhancer at 19p13.11 and the gene promoters for *COMP*, *KLHL26*, *UPF1* and *C19orf60*, in multiple tissues (Fig. 7, Supplementary Material, Table S4).

Discussion

This study represents the first reported experimental effort to bridge GWAS associations to biology in BE/EAC. Studies in the past decade have revealed an important role for inherited genetics in the etiology of BE/EAC ($h^2_g \sim 25$–35%), with nearly 20 genetic loci associated with susceptibility. GWAS associations in isolation can provide initial clues as to potential genes or processes implicated in risk, especially when coupled with informatics data mining, but experimental assays—often laborious and resource-intensive—remain essential for identifying causal variants, target genes and etiologic pathways. In this report, we developed a systematic informatics framework for prioritizing candidate risk variants with high estimated functional potential and applied this pipeline to guide our initial selection of GWAS loci for experimental interrogation. Using luciferase reporter assays in EAC cell lines, we identified rs10423674 as a functional risk variant that modulates the activity of a putative enhancer region at chr19p13.11. Using CRISPR genome editing and expression profiling, we identified two candidate gene targets of this risk enhancer, *CRTC1* and *COMP*, both previously implicated

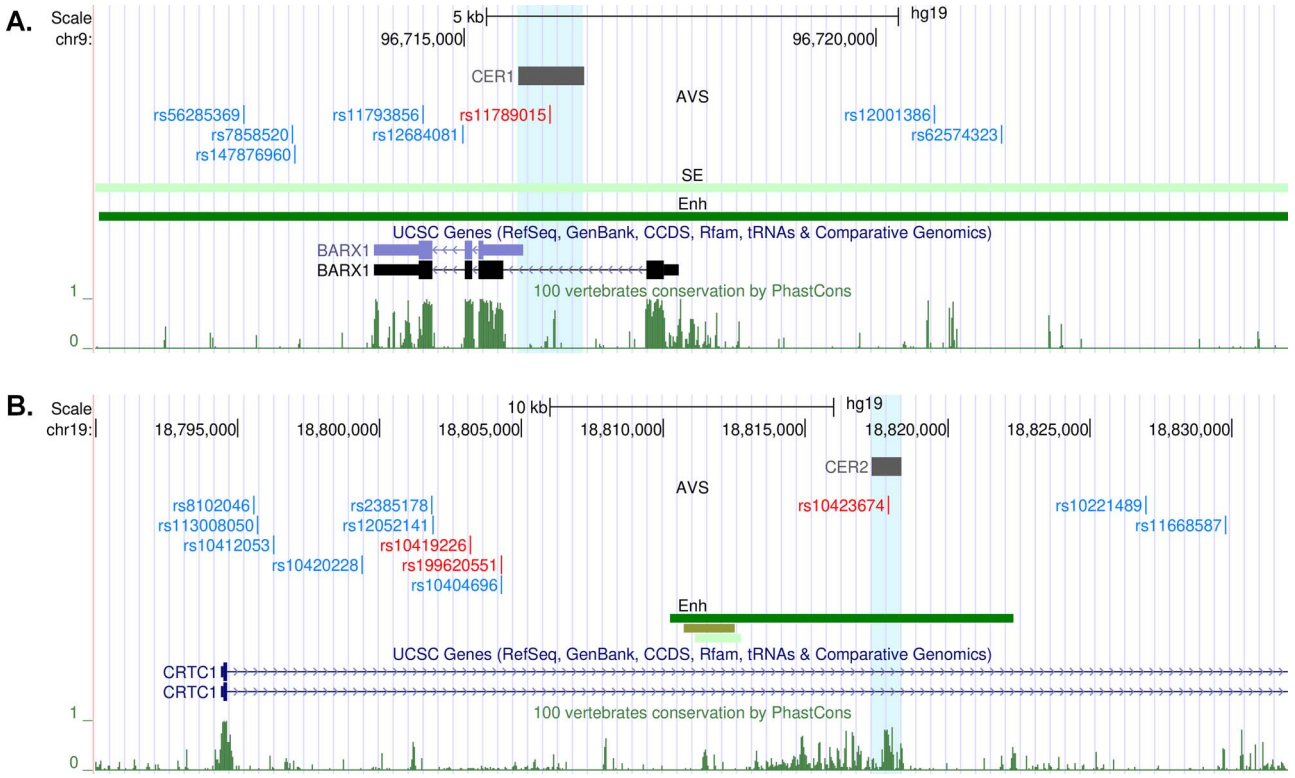

Figure 3. Genome Browser plots for selected GWAS loci. (**A**) Region 13 (9q22.32/*BARX1*), (**B**) Region 19 (19p13.11/*CRTC1*). red: index variants, blue $r^2 > 0.80$; SE/Enh, predicted super-enhancer or enhancer regions based on H3-K27ac ChIP-seq profiles in specific primary tissues: gastric (dark green), esophagus (brown) or stomach smooth muscle (light green) (Hsniz 2013). CER1: 9:96715651-96 716 450 (rs11789015/rs7872123); CER2: 19:18817324-18 818 396 (rs10423674) [http://genome.ucsc.edu].

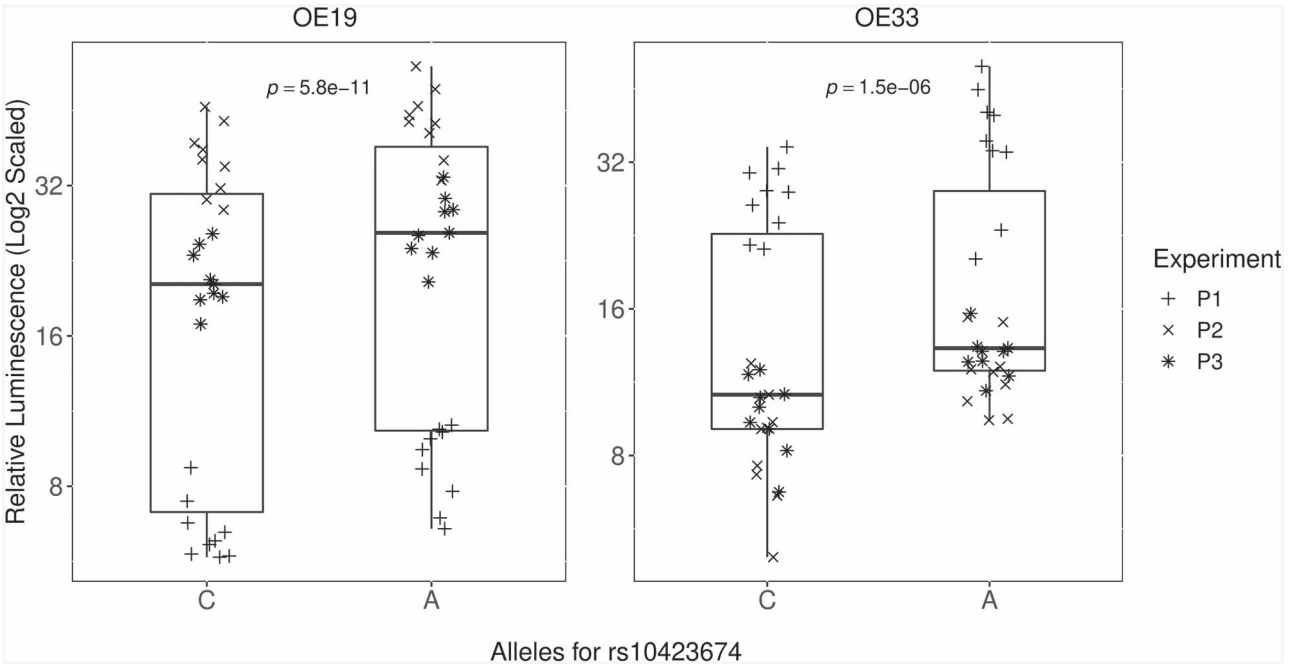

Figure 4. Luciferase reporter enhancer activity assays. The CER at 19p13.11 (CER2) was assessed for allele-specific activity, alongside positive and negative control fragments. Three independent experiments were performed in triplicate in two EAC cell lines (OE19 and OE33).

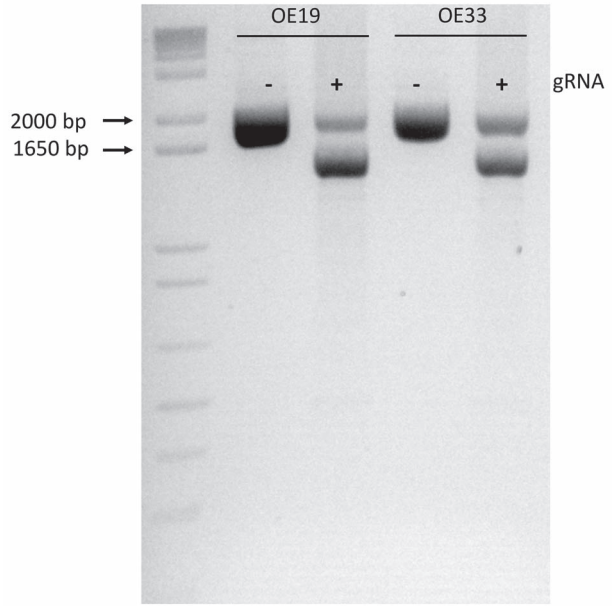

Figure 5. CRISPR-Cas9 genome editing of putative enhancer on 19p13.11. DNA gel electrophoresis showing genome editing of CER2, containing SNP rs10423674, in OE19 and OE33 lines. Region targeted by CRISPR gRNAs is ~0.45 kb. PCR amplified region is ~2 kb. gRNA (−): cells transfected with cas9 vector and gRNA empty vector (−); gRNA (+): cells transfected with cas9 vector and guide RNA target vectors.

in cancer development/progression. Our work provides new insight into the molecular basis of genetic risk for BE/EAC and establishes a generalizable framework to accelerate functional analyses across diverse disease states.

Despite gradual emergence of high-throughput experimental systems for assessing biological function of GWAS variants (33), a major bottleneck in the field remains the identification of functional/causal variants, genes and pathways which underlie genetic associations with disease susceptibility. In recent years, we and others have invested significant time and resources in experimental studies of a limited number of GWAS loci for colorectal cancer (CRC), glioma and other cancers (27–30). While successful, these efforts have often required experimental interrogation of multiple individual candidate variants and enhancer regions to identify the functional variant(s) revealing allele-specific enhancer activity. The general approach has been to use ChIP-seq profiles of enhancer histone marks (H3K27ac/H3K4me1) in disease-relevant cell types to identify putative enhancer regions spanning candidate variants. In this study, we built a more comprehensive and integrative framework for prioritizing variants by consolidating multiple classes of annotations into relative estimates of functional potential. Our 'FPS' flexibly accounts for enhancer histone marks and further incorporates features such as chromatin accessibility, eQTL associations and DNA sequence conservation. We used these scores to guide the selection and filtering process at two stages—first, to prioritize a small number of GWAS susceptibility regions for initial consideration; and second, at a given locus, to select a variant and CER for enhancer activity assays. While only two candidate loci were experimentally evaluated in this report, it is noteworthy that both regions were found to have enhancer activity in our assays, and one of the two exhibited allele-specific activity. A 'success rate' of ~50%, further supported by ongoing studies directed at testing additional prioritized loci, is considerably higher than that achieved in our prior work (27) and supports the utility of using FPS to inform candidate variant prioritization.

At the chr19p13.11 risk locus, rs10423674 was one of two variants (along with rs10419226) which satisfied the stringent Bonferroni significance threshold (5×10^{-8}) in the first published GWAS of BE/EAC (16). When the results from this prior study were combined with additional GWAS data in a subsequent meta-analysis, with imputation, a third variant was identified as the lead 'index' variant at this locus (rs199620551) (15). rs10423674 and rs199620551 exhibit strong LD as measured by D′ (0.96) but are only moderately correlated by r^2 (0.4), likely reflecting a sizable difference in minor allele frequency (34)

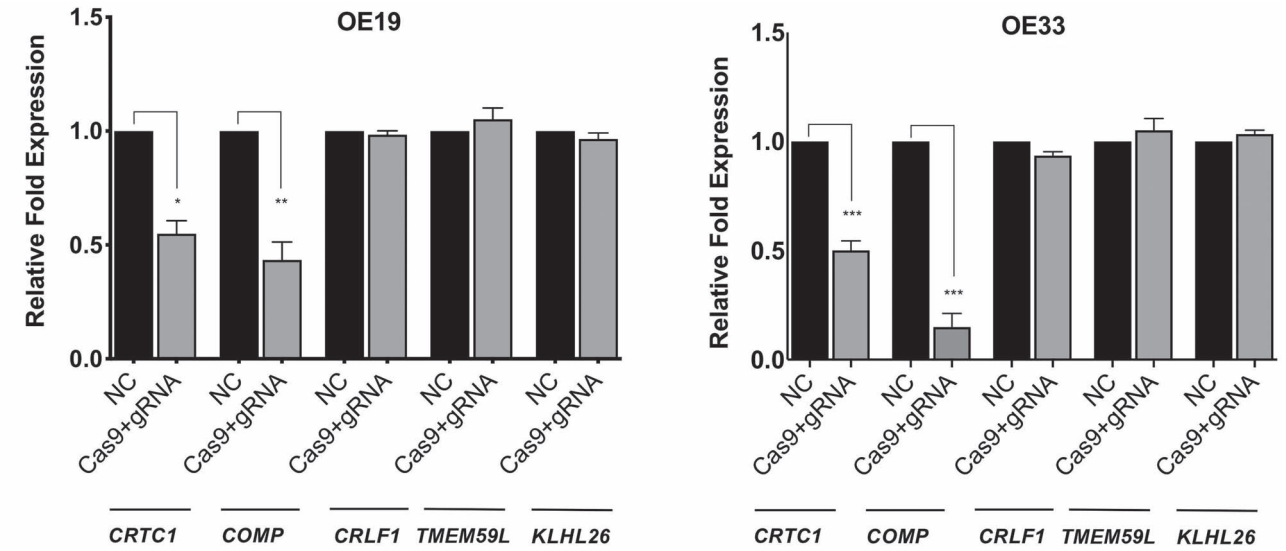

Figure 6. Gene expression changes following CRISPR-Cas9 deletion of putative enhancer on 19p13.11. CER2 was targeted for deletion in OE19 (left) and OE33 (right) cells using CRISPR-Cas9 technology. Pools of transfected cells were analyzed using TaqMan gene expression assays for *CRTC1*, *COMP*, *CRLF1*, *TMEM59L*, *KLHL26* in triplicate, in three independent experiments. NC: mock transfected parental cells; Cas9 + gRNA: cells transfected with Cas9 vector and guide RNA target vectors. $^{*}P < 0.05$; $^{**}P < 0.01$ and $^{***}P < 0.001$.

Table 2. Genotype-expression correlations for rs10423674 and candidate gene targets. Cis-eQTLs for rs10423674 and *CRTC1* or *COMP* in upper GI-tract tissues (GTEx v8: EUR). GEJ, gastroesophageal junction. rs10423674 reference allele (C), alternative allele (A)

	CRTC1			*COMP*		
Tissue	*P*	Slope	SE	*P*	Slope	SE
Esophagus_GEJ	0.752	0.016	0.051	**0.003**	**−0.229**	**0.076**
Esophagus_Mucosa	**0.025**	**−0.090**	**0.040**	0.581	−0.031	0.056
Esophagus_Muscularis	**0.098**	**0.064**	**0.038**	0.311	−0.060	0.059
Stomach	0.780	0.012	0.043	0.060	−0.122	0.064

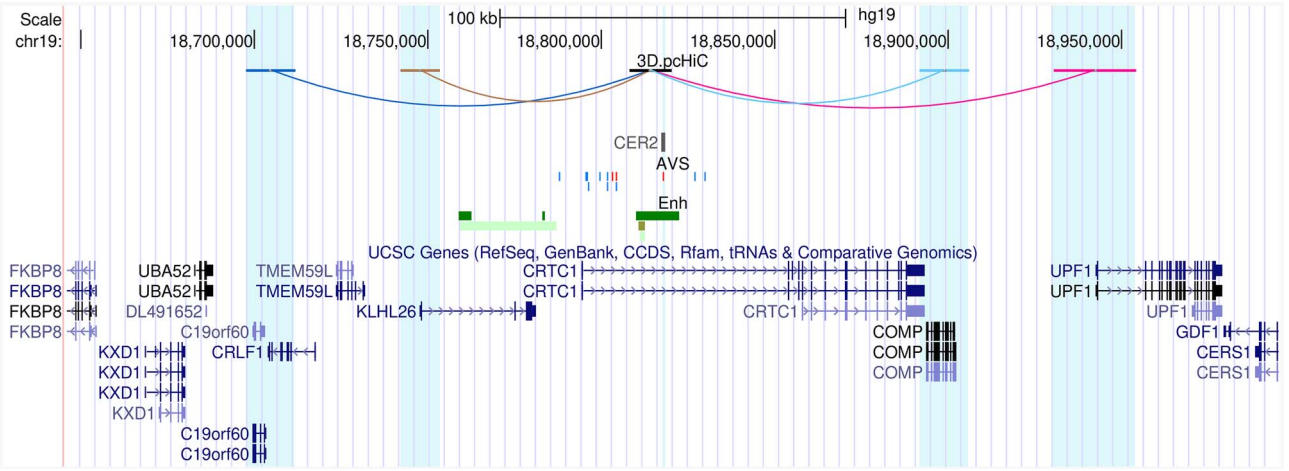

Figure 7. Regional 3D chromatin interactions with putative enhancer containing rs10423674. Promoter capture-Hi-C profiles from multiple primary tissues were queried for fragments spanning the candidate risk SNP. UCSC Genome Browser plot of significant interactions between such fragments and promoters of neighboring genes (±150 kb) [http://genome.ucsc.edu].

as reported in Levine *et al.* (16) (0.34 versus 0.46). For each of these variants, the GWAS risk estimate obtained for BE was very similar to the estimate for EAC, suggesting that this locus may primarily affect risk of BE development, rather than BE progression to cancer (15,16). Our selection of rs10423674 for functional testing was guided largely by FPS, which in turn was driven by two annotation features—DNA sequence conservation across vertebrates (35) and active histone marks (H3K27ac) predictive of an extended enhancer region in GI-tract tissue (25).

Two plausible risk genes at the chr19 GWAS locus emerged from our experimental studies, with further support obtained through informatic and statistical approaches, particularly for *CRTC1* (via colocalization). The protein encoded by *CRTC1*, CREB-regulated transcription coactivator 1, functions to stimulate gene expression programs governed by the transcription factor cyclic AMP response element-binding protein (CREB) (36). Interestingly, CRTC1 was first identified as a translocation fusion partner of the Notch coactivator MAML2 in mucoepidermoid carcinomas (37). Aberrant activation of CRTC1 has been implicated in the biology of several cancers, including lung (38,39), colorectal (40) and esophageal (41). CRTC1 dysregulation, via dephosphorylation and nuclear translocation, has been reported to occur downstream of elevated prostaglandin E2 signaling (40) and loss of the tumor suppressor kinase LKB1 (39,41). Putative pro-tumorigenic transcriptional targets linked to enhanced CRTC1 expression include interleukin-6 and cylcooxygenase-2 (40), two key genes previously implicated in BE/EAC pathogenesis (42,43). CRTC1 also plays important roles in neuronal signaling (44), and in animal models, it has been functionally linked to energy balance and feeding behaviors (45,46), circadian entrainment (47), memory formation (48) and systemic metabolism (49). Sex-biased associations with body mass index and body fat percentage have been reported for *CRTC1* genetic variants rs757318, rs3746266, rs6510997 in prior GWAS or candidate gene studies (50,51). Variant rs757318 is located 2.4 kb downstream of, and is moderately correlated with, rs10423674 ($r^2 = 0.57$).

The protein encoded by *COMP*, cartilage oligomeric matrix protein, is a large pentameric extracellular matrix (ECM) glycoprotein of the thrombospondin gene family (52). While originally isolated from cartilage, secreted COMP is found in a wide range of normal tissues and is thought to be important for ECM structural integrity and mechanical stress resistance. COMP expression has also been reported in multiple epithelial cancers (e.g. CRC, breast, prostate) and implicated in pro-tumorigenic activities (53–58). Experimental studies in CRC have linked *COMP* overexpression to enhanced epithelial-to-mesenchymal transition and *COMP* knockdown to reduced invasion/metastasis (55). Similar studies in breast cancer linked *COMP* overexpression to reductions in mitochondrial respiration, ER stress and apoptosis (54); higher COMP levels in human breast (54) and colon (55,59) tumors correlated with worse patient survival.

An important question when evaluating candidate functional variants in post-GWAS studies is the tissue and cell type in which the causal variant(s) exerts its functional/biological effects to alter disease risk. The etiology of many disease states, including BE/EAC, may be influenced by events in multiple tissues and

cell types, reflecting both direct and indirect effects. While BE is defined by the replacement of normal squamous epithelium in the distal esophagus by intestinal metaplasia, the precise cellular origins of this metaplastic tissue remain controversial (60,61). Currently, leading candidates for the BE cell-of-origin include various epithelial subpopulations, such as transitional basal epithelial cells (TBECs) (62), reserve embryonic cells (RECs) (63) and esophageal submucosal glands (64). Under conditions of chronic reflux into the lower esophagus, TBECs and/or RECs may become 'activated' and undergo a proliferative expansion. Under one scenario, a genetic risk variant for BE may lead to alterations in gene expression in these particular cells, directly modifying their propensity to give rise to BE. Alternatively, or in parallel, such a variant could modulate expression programs in different tissues, indirectly influencing susceptibility via effects on acid reflux or systemic inflammation, two established risk factors for BE/EAC. Recent genetic studies of gastro-esophageal reflux disease (GERD) suggest that a number of new loci identified for this trait are similarly associated with risk of BE/EAC (65).

Our current data provide support for a model whereby rs10423674 exerts regulatory function in EAC cells, resulting in altered expression of *CRTC1* and *COMP*, but it remains to be determined whether such regulation also occurs in the BE cell-of-origin and/or other cell types, and what the impact may be on BE/EAC etiology. Our colocalization analysis at the chr19p13.11 locus suggests that a shared genetic signal may underlie the GWAS risk association and the *CRTC1* eQTL in esophageal muscularis. This result further implicates *CRTC1* as a candidate gene target and points to smooth muscle—the primary constituent of the lower esophageal sphincter—as a potential 'causal tissue' for the chr19p13.11 risk locus. Notably, structural and functional abnormalities of the lower esophageal sphincter are known causes of GERD, the strongest risk factor identified for BE/EAC (66). Assessment of chromatin accessibility and enhancer histone marks across diverse tissues and cell types further suggests that the putative enhancer region containing rs10423674 may be active during fetal muscle development (Supplementary Material, Fig. S5). Intriguingly, an expanded colocalization scan we conducted across all 49 GTEx tissues implicated another gene as a potential target of the chr19 locus—cytokine receptor-like factor 1 (data not shown). *CRLF1* appears to be important for lower esophageal sphincter function via modulation of neurotrophic signaling in the nervous system. Loss-of-function mutations in *CRLF1* cause achalasia, an esophageal motility disorder (67).

There are some limitations to our study. We cannot discount the possibility that there are additional functional SNPs on 19p11.13 that may regulate gene expression, including via alternate mechanisms such as modulation of microRNA binding or alternative splicing. Our *in-vitro* assessment of enhancer activity was conducted in only two EAC cell lines, and it remains possible that a relevant enhancer(s) in this region is inactive in the cell lines tested but active in other cell lines. Another limitation is that in the CRISPR experiments, we have not determined the impact of SNP rs10423674 directly, but of the enhancer containing that SNP. Nevertheless, we believe that we provide strong supportive evidence for the identification of rs10423674 as a functional SNP relevant to BE/EAC development. We acknowledge that as chromatin profiles, 3D interaction maps and eQTL regulatory maps continue to accumulate, particularly at the single-cell level and under specific environmental conditions (68–71), our current FPS pipeline can be readily extended and refined via inclusion of features tailored to the disease of interest. Our understanding of the precise mechanisms whereby inherited genetic variation translates into altered disease risk will significantly expand as the resolution of our assays and experimental systems captures increasingly more of the dynamic biological complexity of gene regulatory networks across the spectrum of human development and disease pathogenesis (72).

In conclusion, this study represents the first reported experimental functional analysis of inherited genetic risk loci associated with BE/EAC. We find that rs10423674 (chr19p13.11) modulates the activity of a putative regulatory enhancer in EAC cell lines, which in turn correlates with altered expression of genes *CRTC1* and *COMP*. Future studies will determine whether expression levels of additional genes, including long-range targets and non-coding RNAs, are influenced by this enhancer; whether such regulation manifests in specific cell types critical to the etiology of BE/EAC and whether other functional variants at chr19p13.11 may in part drive the genetic association. Our work suggests that use of an integrative and customizable informatics pipeline and scoring system to prioritize candidate functional variants for experimental follow-up can substantially improve the efficiency of GWAS functional studies.

Materials and Methods

Selection of candidate risk variants

We identified 31 genome-wide-significant ($P < 5 \times 10^{-8}$) variants reported in published GWAS of BE/EAC, distributed across 19 susceptibility regions (Table 1) (15–18,20,21). Using the LDproxy Tool at NCBI (73), we selected all variants in moderate-to-strong LD ($r^2 > 0.6$) with the 31 'index variants' (CEU reference population), yielding a total of 1565 candidate functional/causal variants.

Variant annotations

The genomic and epigenomic context of each variant was investigated using data from public repositories and resources from published studies (Supplementary Material, Table S1). *Sequence-based features* were extracted from HaploReg v4.1 (35) and included dbSNP annotations (exon/intron/UTR/intergenic), GERP/SiPhy conservation calls and predicted DNA binding motifs altered by the variant. *Chromatin-based features* were extracted from the Roadmap Epigenome Project (74), ENCODE (75) and The Cancer Genome Atlas (76) and included physical accessibility profiles and inferred chromatin states derived from histone modification profiles. *Other features* such as eQTL and 3D genome interaction profiles (Hi-C) were extracted from the Genotype-Tissue Expression (GTEx) repository (31) and 4DGenome (77), respectively. The distributions of annotation categorizations across all 1565 variants are summarized in Supplementary Material, Fig. S1.

Assembly of FPS for prioritizing candidate risk variants and GWAS loci

To consolidate multiple annotation features ($n = 12$) into a single composite measure of estimated functional potential, we assigned numeric component scores to the specific categories/values of each feature. Given that chromatin-based features and eQTLs were derived in specific tissues, we first designated a limited set of 'key tissues' based on anticipated etiologic relevance to BE/EAC (Supplementary Material, Table S2). These included tissues of the upper GI tract (e.g. esophagus, stomach), as BE/EAC

arises in the lower esophagus near the gastroesophageal (GE) junction (1). A detailed description of the scoring scheme is provided in Supplementary material, Table S3. Annotations specific to key tissues were assigned the highest component scores, followed by those present in key and other tissues, or other tissues only. The FPS for each variant was calculated by summing across the individual feature component scores assigned to that variant. The overall rationale was to prioritize most highly variants with multiple indicators of functional potential, particularly those situated in likely enhancers in select tissues, accounting for DNA sequence, local chromatin and regulatory associations.

The FPS distribution obtained across all candidate risk variants was assessed, and scores of individual variants at each GWAS locus were visualized using dot plots. We used these scores to prioritize a subset of the 19 GWAS risk loci for subsequent experimental testing. After rank ordering the loci based on maximum FPS obtained for any of their candidate risk variants, we selected two top-scoring loci for analysis in this report. At each locus, final selection of a CER spanning one or more high-scoring candidate risk variants was guided by joint consideration of FPS with r^2 and genomic distance to the index variant(s), supplemented by visualization of individual annotation features via customized UCSC Genome Browser plots.

Cell culture

OE19 and OE33 EAC cell lines were obtained from ECACC (European Collection of Authenticated Cell Cultures) via MilliporeSigma (St. Louis, MO). Mycoplasma testing of the cell lines revealed no contamination. Stocks of the cell lines were frozen at low passage numbers. OE19 and OE33 cells were grown in DMEM (Thermo Fisher) supplemented with 5% Fetal Bovine Serum (Thermo Fisher) and 1% Penicillin/Streptomycin, and incubated at 37°C and 5% CO_2.

Plasmids and luciferase reporter assays

DNA fragments containing alternate alleles of each of the two candidate variants/haplotypes were PCR-amplified from normal human genomic DNA and subcloned into Sac I and Xho I restriction enzyme sites (in both orientations) upstream of a TK (thymidine kinase) minimal promoter-firefly-luciferase vector (30) using CloneAmp HiFi PCR Premix and the In-Fusion HD cloning kit (Takara). Plasmid clones were sequenced by Sanger sequencing (Genewiz) to confirm the presence of candidate variants and the absence of any PCR amplification-induced mutations.

A region from chr3p13 shown to have no activity in either of the cell lines served as the negative control, and a region of chr9q22.32 shown to have enhancer activity in both cell lines served as the positive control. For enhancer assays, OE19 and OE33 cells (1×10^5 cells/well) were seeded into 96-well plates. Cells were co-transfected with reporter plasmids and constitutively active pNL1.1.TK [Nluc/TK] Vector (Promega) using Lipofectamine 2000 Reagent (Thermo Fisher) according to the manufacturer's instructions. After 48 hours, cells were assayed for luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions and measured using a Synergy H1 Microplate Reader (BioTek). To quantify enhancer activity in cells transfected with candidate enhancer elements, luminescence resulting from the transcription and translation of firefly luciferase in the presence of luciferin was measured, and background luminescence in the absence of reagents was subtracted. To control for cell concentration, normalization with luminescence from constitutively active NanoLuc luciferase was performed (i.e. candidate enhancer/constitutive control). To control for assay artifact, normalization with luminescence from cells transfected with clones empirically found to have no enhancer activity was performed (i.e. candidate enhancer/negative control). Measurements for each enhancer were obtained in three wells (i.e. technical replicates) for three clones (i.e. biological replicates) for each of the alleles observed in human populations (i.e. experimental conditions) on three separate days (i.e. experimental validations) and in two independent cell lines (i.e. experimental validations). For statistical testing, measurements from the two cell lines were considered separately, and regression analysis was performed with generalized estimating equations in order to account for repeated measurements of clones (78,79).

CRISPR and gene expression assays

Upstream and downstream CRISPR gRNAs (guide RNAs) were designed flanking the CER using CRISPRscan (80) and synthesized by Synthego (Menlo Park, CA) (guide sequences; gRNA1 5' GTGAAAAGGCCCCATTCCCA 3′, gRNA2 5' GCCAAACCATTCAACGGGGA 3′). The chemical modifications 2'-O-Methyl at 3 first and last bases and 3′ phosphorothioate bonds between first 3 and last 2 bases were introduced into the gRNAs in order to provide superior editing in the cell lines used (Synthego) (guide sequences following chemical modifications; gRNA1 5' GUGAAAAGGCCCCAUUCCCA 3′, gRNA2 5' GCCAAACCAUUCAACGGGGA 3′). The Cas9 2NLS Nuclease was purchased from Synthego. OE19 and OE33 cells were electroporated by gRNA and Cas9 (1:3 ratio) using the Neon Transfection System (Thermo Fisher). Electroporated cells were allowed to grow for 48 hours prior to DNA and RNA harvesting. Genomic DNA was purified using the QIAamp DNA Mini Kit (Qiagen) and enhancer deletion was confirmed with PCR amplifications (forward primer 5' CTCACTGCAACCTCCTCCTC 3′, reverse primer 5' ATCACGTCCAAGCTCTCCAG 3′).

Quantitative real-time PCR

RNA was isolated using Trizol reagent (Thermo Fisher) and cDNA was synthesized from 2 μg of total RNA using the High Capacity Reverse Transcriptase cDNA kit (Thermo Fisher). Quantitative real-time polymerase chain reaction was performed using Superscript III kit for RT-PCR (Thermo Fisher) and amplified with TaqMan assays for selected genes mapping within 100 kb upstream and downstream of variant rs10423674—*CRTC1* (Hs00993064_m1), *COMP* (Hs00164359_m1), *CRLF1* (Hs00191064_m1), *TMEM95L* (Hs00201595_m1), *KLHL26* (Hs00217801_m1)—in three independent experiments and in triplicate for each RNA preparation on QuantStudio 5 (Thermo Fisher) and analyzed using GraphPad Prism (version 8.3.0, GraphPad Software, La Jolla, CA, USA). Reactions were normalized using the control gene *GUSB* (assay ID: Hs00939627_m1), and calculations were performed according to the $2^{\wedge ddCT}$ method. Fold change in expression was determined from three independent experimental repeats, each performed in duplicate, unless otherwise noted. Data were analyzed for statistical differences using an analysis of variance, with Bonferroni correction for multiple hypothesis testing. $^*P < 0.05$; $^{**}P < 0.01$ and $^{***}P < 0.001$ indicate levels of statistical significance.

eQTL and colocalization analysis

eQTL data for rs10423674 in GI-tract and other tissues were obtained from European-ancestry individuals included in the Genotype-Tissue Expression (GTEx) project v8 repository (31). Complete EUR summary statistics were downloaded from the Google cloud and converted from Parquet format to plain text files using Python. Colocalization analysis was performed using the COLOC software to evaluate the posterior probability ('PP4') of a shared causal signal underlying both the observed GWAS association (with BE/EAC risk) and the eQTL for a given gene in a selected tissue (81). The method also evaluates if the expression association and disease association are driven by two distinct causal variants (PP3). A high PP4 (>0.8) and a low PP3 (<0.2) indicate that a single variant is responsible for both the GWAS and eQTL signals. BEACON/Cambridge GWAS data (16) were used for colocalization.

Visualization of Hi-C chromosome conformation interactions

Hi-C and promoter-capture Hi-C profiles derived from the OE33 cell line (82) or multiple human primary tissues (32) were queried for fragments spanning the candidate risk variant rs10423674 (chr19p13.11). Interactions between such 'bait' fragments and promoter regions of neighboring genes (±150 kb) reported as statistically significant (FDR < 0.09) were depicted graphically using custom scripts generated for the UCSC Genome Browser (83).

Supplementary Material

Supplementary material is available at *HMG* online.

Authors' contributions

Conception and design: M.F.B., G.C. Data processing/extraction and bioinformatics: J.H.C., M.F.B. Laboratory assays: M.W.A. Analysis and/or interpretation of data: M.F.B, L.Y., J.H.C., M.W.A., G.C., S.G.D, J.Y.D. and T.L.V. Drafting of the manuscript: M.F.B., J.H.C., M.W.A. and G.C. Study supervision: M.F.B. and G.C. All authors approved of the final draft submitted.

Acknowledgments

We acknowledge investigators in the BEACON Consortium and collaborating consortia in the UK and Germany for their instrumental roles in prior GWAS of BE/EAC. In particular, we thank Drs David Whiteman and David Levine (BEACON GWAS); Drs Rebecca Fitzgerald, Paul D. Pharoah and Carlos Caldas from the Oesophageal Cancer Clinical and Molecular Stratification Consortium (Cambridge GWAS); Drs Janusz Jankowski, Ian Tomlinson and Claire Palles (Oxford GWAS); Dr Johannes Schumacher (Bonn GWAS) and Drs Stuart MacGregor and Puya Gharahkhani (meta-GWAS). Detailed information regarding the assembly of GWAS datasets and use of external genotyping data has been published previously (15–18).

Conflict of Interest statement. The authors declare no potential conflicts of interest.

Funding

National Cancer Institute (P30CA016056, Roswell Park Comprehensive Cancer Center Support Grant; R01CA204279 to G.C.; R01CA136725 to T.L.V. to support the BEACON GWAS).

References

1. Reid, B.J., Li, X., Galipeau, P.C. and Vaughan, T.L. (2010) Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. *Nat. Rev. Cancer*, **10**, 87–101.
2. Vaughan, T.L. and Fitzgerald, R.C. (2015) Precision prevention of oesophageal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.*, **12**, 243–248.
3. Coleman, H.G., Xie, S.-H. and Lagergren, J. (2017) The epidemiology of esophageal adenocarcinoma. *Gastroenterology*, **154**, 390–405.
4. Njei, B., McCarty, T.R. and Birk, J.W. (2016) Trends in esophageal cancer survival in United States adults from 1973 to 2009: a SEER database analysis. *J. Gastroenterol. Hepatol.*, **31**, 1141–1146.
5. Naef, A.P., Savary, M. and Ozzello, L. (1975) Columnar-lined lower esophagus: an acquired lesion with malignant predisposition. Report on 140 cases of Barrett's esophagus with 12 adenocarcinomas. *J. Thorac. Cardiovasc. Surg.*, **70**, 826–835.
6. Giroux, V. and Rustgi, A.K. (2017) Metaplasia: tissue injury adaptation and a precursor to the dysplasia-cancer sequence. *Nat. Rev. Cancer*, **17**, 594–604.
7. Cook, M.B., Corley, D.A., Murray, L.J., Liao, L.M., Kamangar, F., Ye, W., Gammon, M.D., Risch, H.A., Casson, A.G., Freedman, N.D. *et al.* (2014) Gastroesophageal reflux in relation to adenocarcinomas of the esophagus: a pooled analysis from the Barrett's and Esophageal adenocarcinoma Consortium (BEACON). *PLoS One*, **9**, e103508.
8. Cook, M.B., Kamangar, F., Whiteman, D.C., Freedman, N.D., Gammon, M.D., Bernstein, L., Brown, L.M., Risch, H.A., Ye, W., Sharp, L. *et al.* (2010) Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON Consortium. *J. Natl. Cancer Inst.*, **102**, 1344–1353.
9. Cook, M.B., Shaheen, N.J., Anderson, L.A., Giffen, C., Chow, W.H., Vaughan, T.L., Whiteman, D.C. and Corley, D.A. (2012) Cigarette smoking increases risk of Barrett's esophagus: an analysis of the barrett's and esophageal adenocarcinoma consortium. *Gastroenterology*, **142**, 744–753.
10. Hoyo, C., Cook, M.B., Kamangar, F., Freedman, N.D., Whiteman, D.C., Bernstein, L., Brown, L.M., Risch, H.A., Ye, W., Sharp, L. *et al.* (2012) Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the international BEACON consortium. *Int. J. Epidemiol.*, **41**, 1706–1718.
11. Lagergren, J., Bergström, R., Lindgren, A. and Nyrén, O. (1999) Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N. Engl. J. Med.*, **340**, 825–831.
12. Vaughan, T.L., Davis, S., Kristal, A. and Thomas, D.B. (1995) Obesity, alcohol, and tobacco as risk factors for cancers of the esophagus and gastric cardia: adenocarcinoma versus squamous cell carcinoma. *Cancer Epidemiol. Biomark. Prev.*, **4**, 85–92.
13. Smith, K.J., O'Brien, S.M., Smithers, B.M., Gotley, D.C., Webb, P.M., Green, A.C. and Whiteman, D.C. (2005) Interactions among smoking, obesity, and symptoms of acid reflux in Barrett's esophagus. *Cancer Epidemiol. Biomark Prev.*, **14**, 2481–2486.
14. Contino, G., Vaughan, T.L., Whiteman, D. and Fitzgerald, R.C. (2017) The evolving genomic landscape of Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology*, **153**, 657–673.e1.
15. Gharahkhani, P., Fitzgerald, R.C., Vaughan, T.L., Palles, C., Gockel, I., Tomlinson, I., Buas, M.F., May, A., Gerges, C.,

Anders, M. *et al.* (2016) Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. *Lancet Oncol.*, **17**, 1363–1373.
16. Levine, D.M., Ek, W.E., Zhang, R., Liu, X., Onstad, L., Sather, C., Lao-Sirieix, P., Gammon, M.D., Corley, D.A., Shaheen, N.J. *et al.* (2013) A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. *Nat. Genet.*, **45**, 1487–1493.
17. Palles, C., Chegwidden, L., Li, X., Findlay, J.M., Farnham, G., Giner, F.C., Peppelenbosch, M.P., Kovac, M., Adams, C.L., Prenen, H. *et al.* (2015) Polymorphisms near TBX5 and GDF7 are associated with increased risk for Barrett's esophagus. *Gastroenterology*, **148**, 367–378.
18. Su, Z., Gay, L.J., Strange, A., Palles, C., Band, G., Whiteman, D.C., Lescai, F., Langford, C., Nanji, M., Edkins, S. *et al.* (2012) Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nat. Genet.*, **44**, 1131–1136.
19. Ek, W.E., Levine, D.M., D'Amato, M., Pedersen, N.L., Magnusson, P.K.E., Bresso, F., Onstad, L.E., Schmidt, P.T., Törnblom, H., Nordenstedt, H. *et al.* (2013) Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. *J. Natl. Cancer Inst.*, **105**, 1711–1718.
20. Dai, J.Y., de Dieu Tapsoba, J., Buas, M.F., Onstad, L.E., Levine, D.M., Risch, H.A., Chow, W.-H., Bernstein, L., Ye, W., Lagergren, J. *et al.* (2015) A newly identified susceptibility locus near FOXP1 modifies the association of gastroesophageal reflux with Barrett's esophagus. *Cancer Epidemiol. Biomark. Prev.*, **24**, 1739–1747.
21. Dai, J.Y., Tapsoba, J.d.D., Buas, M.F., BEACON Consortium, Risch, H.A. and Vaughan, T.L. (2016) Constrained score statistics identify genetic variants interacting with multiple risk factors in Barrett's Esophagus. *Am. J. Hum. Genet.*, **99**, 352–365.
22. Buas, M.F., He, Q., Johnson, L.G., Onstad, L., Levine, D.M., Thrift, A.P., Gharahkhani, P., Palles, C., Lagergren, J., Fitzgerald, R.C. *et al.* (2017) Germline variation in inflammation-related pathways and risk of Barrett's oesophagus and oesophageal adenocarcinoma. *Gut*, **66**, 1739–1747.
23. Buas, M.F., Levine, D.M., Makar, K.W., Utsugi, H., Onstad, L., Li, X., Galipeau, P.C., Shaheen, N.J., Hardie, L.J., Romero, Y. *et al.* (2014) Integrative post-genome-wide association analysis of CDKN2A and TP53 SNPs and risk of esophageal adenocarcinoma. *Carcinogenesis*, **35**, 2740–2747.
24. Dighe, S.G., Chen, J., Yan, L., He, Q., Gharahkhani, P., Onstad, L., Levine, D.M., Palles, C., Ye, W., Gammon, M.D. *et al.* (2020) Germline variation in the insulin-like growth factor pathway and risk of Barrett's esophagus and esophageal adenocarcinoma. *Carcinogenesis*, **42**, 369–377.
25. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
26. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190–1195.
27. Biancolella, M., Fortini, B.K., Tring, S., Plummer, S.J., Mendoza-Fandino, G.A., Hartiala, J., Hitchler, M.J., Yan, C., Schumacher, F.R., Conti, D.V. *et al.* (2014) Identification and characterization of functional risk variants for colorectal cancer mapping to chromosome 11q23.1. *Hum. Mol. Genet.*, **23**, 2198–2209.
28. Fortini, B.K., Tring, S., Plummer, S.J., Edlund, C.K., Moreno, V., Bresalier, R.S., Barry, E.L., Church, T.R., Figueiredo, J.C. and Casey, G. (2014) Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *PLoS One*, **9**, e111914.
29. Fortini, B.K., Tring, S., Devall, M.A., Ali, M.W., Plummer, S.J. and Casey, G. (2021) SNPs associated with colorectal cancer at 15q13.3 affect risk enhancers that modulate GREM1 gene expression. *Hum. Mutat.*, **42**, 237–245.
30. Ali, M.W., Patro, C.P.K., Zhu, J.J., Dampier, C.H., Plummer, S.J., Kuscu, C., Adli, M., Lau, C., Lai, R.K. and Casey, G. (2020) A functional variant on 20q13.33 related to glioma risk alters enhancer activity and modulates expression of multiple genes. *Hum. Mutat.*, **42**, 77–88.
31. Consortium, G. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, N.Y.)*, **369**, 1318–1330.
32. Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S. *et al.* (2019) A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.*, **51**, 1442–1449.
33. Montalbano, A., Canver, M.C. and Sanjana, N.E. (2017) High-throughput approaches to pinpoint function within the non-coding genome. *Mol. Cell*, **68**, 44–59.
34. Wray, N.R. (2005) Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res. Hum. Genet.*, **8**, 87–94.
35. Ward, L.D. and Kellis, M. (2016) HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.*, **44**, D877–D881.
36. Iourgenko, V., Zhang, W., Mickanin, C., Daly, I., Jiang, C., Hexham, J.M., Orth, A.P., Miraglia, L., Meltzer, J., Garza, D. *et al.* (2003) Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 12147–12152.
37. Tonon, G., Modi, S., Wu, L., Kubo, A., Coxon, A.B., Komiya, T., O'Neil, K., Stover, K., El-Naggar, A., Griffin, J.D. *et al.* (2003) T(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a notch signaling pathway. *Nat. Genet.*, **33**, 208–213.
38. Feng, Y., Wang, Y., Wang, Z., Fang, Z., Li, F., Gao, Y., Liu, H., Xiao, T., Li, F., Zhou, Y. *et al.* (2012) The CRTC1-NEDD9 signaling axis mediates lung cancer progression caused by LKB1 loss. *Cancer Res.*, **72**, 6502–6511.
39. Cao, C., Gao, R., Zhang, M., Amelio, A.L., Fallahi, M., Chen, Z., Gu, Y., Hu, C., Welsh, E.A., Engel, B.E. *et al.* (2015) Role of LKB1-CRTC1 on glycosylated COX-2 and response to COX-2 inhibition in lung cancer. *J. Natl. Cancer Inst.*, **107**, 358.
40. Schumacher, Y., Aparicio, T., Ourabah, S., Baraille, F., Martin, A., Wind, P., Dentin, R., Postic, C. and Guilmeau, S. (2016) Dysregulated CRTC1 activity is a novel component of PGE2 signaling that contributes to colon cancer growth. *Oncogene*, **35**, 2602–2614.
41. Gu, Y., Lin, S., Li, J.-L., Nakagawa, H., Chen, Z., Jin, B., Tian, L., Ucar, D.A., Shen, H., Lu, J. *et al.* (2012) Altered LKB1/CREB-regulated transcription co-activator (CRTC) signaling axis promotes esophageal cancer cell migration and invasion. *Oncogene*, **31**, 469–479.
42. Dvorak, K. and Dvorak, B. (2013) Role of interleukin-6 in Barrett's esophagus pathogenesis. *World J. Gastroenterol.*, **19**, 2307–2312.

43. Shimizu, D., Vallböhmer, D., Kuramochi, H., Uchida, K., Schneider, S., Chandrasoma, P.T., Shimada, H., DeMeester, T.R., Danenberg, K.D., Peters, J.H. *et al.* (2006) Increasing cyclooxygenase-2 (cox-2) gene expression in the progression of Barrett's esophagus to adenocarcinoma correlates with that of Bcl-2. *Int. J. Cancer*, **119**, 765–770.
44. Ch'ng, T.H., Uzgil, B., Lin, P., Avliyakulov, N.K., O'Dell, T.J. and Martin, K.C. (2012) Activity-dependent transport of the transcriptional coactivator CRTC1 from synapse to nucleus. *Cell*, **150**, 207–221.
45. Rossetti, C., Sciarra, D., Petit, J.-M., Eap, C.B., Halfon, O., Magistretti, P.J., Boutrel, B. and Cardinaux, J.-R. (2017) Gender-specific alteration of energy balance and circadian locomotor activity in the Crtc1 knockout mouse model of depression. *Transl. Psychiatry*, **7**, 1269.
46. Altarejos, J.Y., Goebel, N., Conkright, M.D., Inoue, H., Xie, J., Arias, C.M., Sawchenko, P.E. and Montminy, M. (2008) The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nat. Med.*, **14**, 1112–1117.
47. Jagannath, A., Butler, R., Godinho, S.I.H., Couch, Y., Brown, L.A., Vasudevan, S.R., Flanagan, K.C., Anthony, D., Churchill, G.C., Wood, M.J.A. *et al.* (2013) The CRTC1-SIK1 pathway regulates entrainment of the circadian clock. *Cell*, **154**, 1100–1111.
48. Nonaka, M., Kim, R., Fukushima, H., Sasaki, K., Suzuki, K., Okamura, M., Ishii, Y., Kawashima, T., Kamijo, S., Takemoto-Kimura, S. *et al.* (2014) Region-specific activation of CRTC1-CREB signaling mediates long-term fear memory. *Neuron*, **84**, 92–106.
49. Burkewitz, K., Morantte, I., Weir, H.J.M., Yeo, R., Zhang, Y., Huynh, F.K., Ilkayeva, O.R., Hirschey, M.D., Grant, A.R. and Mair, W.B. (2015) Neuronal CRTC-1 governs systemic mitochondrial metabolism and lifespan via a catecholamine signal. *Cell*, **160**, 842–855.
50. Lu, Y., Day, F.R., Gustafsson, S., Buchkovich, M.L., Na, J., Bataille, V., Cousminer, D.L., Dastani, Z., Drong, A.W., Esko, T. *et al.* (2016) New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.*, **7**, 10495.
51. Choong, E., Quteineh, L., Cardinaux, J.-R., Gholam-Rezaee, M., Vandenberghe, F., Dobrinas, M., Bondolfi, G., Etter, M., Holzer, L., Magistretti, P. *et al.* (2013) Influence of CRTC1 polymorphisms on body mass index and fat mass in psychiatric patients and the general adult population. *JAMA Psychiat.*, **70**, 1011–1019.
52. Posey, K.L., Coustry, F. and Hecht, J.T. (2018) Cartilage oligomeric matrix protein: COMPopathies and beyond. *Matrix Biol.*, **71–72**, 161–173.
53. Papadakos, K.S., Bartoschek, M., Rodriguez, C., Gialeli, C., Jin, S.-B., Lendahl, U., Pietras, K. and Blom, A.M. (2019) Cartilage oligomeric matrix protein initiates cancer stem cells through activation of Jagged1-Notch3 signaling. *Matrix Biol.*, **81**, 107–121.
54. Englund, E., Bartoschek, M., Reitsma, B., Jacobsson, L., Escudero-Esparza, A., Orimo, A., Leandersson, K., Hagerling, C., Aspberg, A., Storm, P. *et al.* (2016) Cartilage oligomeric matrix protein contributes to the development and metastasis of breast cancer. *Oncogene*, **35**, 5585–5596.
55. Zhong, W., Hou, H., Liu, T., Su, S., Xi, X., Liao, Y., Xie, R., Jin, G., Liu, X., Zhu, L. *et al.* (2020) Cartilage oligomeric matrix protein promotes epithelial-mesenchymal transition by interacting with transgelin in colorectal cancer. *Theranostics*, **10**, 8790–8806.
56. Liao, Q., Kleeff, J., Xiao, Y., Di Cesare, P.E., Korc, M., Zimmermann, A., Büchler, M.W. and Friess, H. (2003) COMP is selectively up-regulated in degenerating acinar cells in chronic pancreatitis and in chronic-pancreatitis-like lesions in pancreatic cancer. *Scand. J. Gastroenterol.*, **38**, 207–215.
57. Nfonsam, V.N., Jecius, H.C., Janda, J., Omesiete, P.N., Elquza, E., Scott, A.J., Nfonsam, L.E. and Jandova, J. (2020) Cartilage oligomeric matrix protein (COMP) promotes cell proliferation in early-onset colon cancer tumorigenesis. *Surg. Endosc.*, **34**, 3992–3998.
58. Englund, E., Canesin, G., Papadakos, K.S., Vishnu, N., Persson, E., Reitsma, B., Anand, A., Jacobsson, L., Helczynski, L., Mulder, H. *et al.* (2017) Cartilage oligomeric matrix protein promotes prostate cancer progression by enhancing invasion and disrupting intracellular calcium homeostasis. *Oncotarget*, **8**, 98298–98311.
59. Nfonsam, V.N., Nfonsam, L.E., Chen, D., Omesiete, P.N., Cruz, A., Runyan, R.B. and Jandova, J. (2019) COMP gene coexpresses with EMT genes and is associated with poor survival in colon cancer patients. *J. Surg. Res.*, **233**, 297–303.
60. McDonald, S.A.C., Lavery, D., Wright, N.A. and Jansen, M. (2015) Barrett oesophagus: lessons on its origins from the lesion itself. *Nat. Rev. Gastroenterol. Hepatol.*, **12**, 50–60.
61. Que, J., Garman, K.S., Souza, R.F. and Spechler, S.J. (2019) Pathogenesis and cells of origin of Barrett's Esophagus. *Gastroenterology*, **157**, 349–364.e1.
62. Jiang, M., Li, H., Zhang, Y., Yang, Y., Lu, R., Liu, K., Lin, S., Lan, X., Wang, H., Wu, H. *et al.* (2017) Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. *Nature*, **550**, 529–533.
63. Wang, X., Ouyang, H., Yamamoto, Y., Kumar, P.A., Wei, T.S., Dagher, R., Vincent, M., Lu, X., Bellizzi, A.M., Ho, K.Y. *et al.* (2011) Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell*, **145**, 1023–1035.
64. Owen, R.P., White, M.J., Severson, D.T., Braden, B., Bailey, A., Goldin, R., Wang, L.M., Ruiz-Puig, C., Maynard, N.D., Green, A. *et al.* (2018) Single cell RNA-seq reveals profound transcriptional similarity between Barrett's oesophagus and oesophageal submucosal glands. *Nat. Commun.*, **9**, 4261.
65. An, J., Gharahkhani, P., Law, M.H., Ong, J.-S., Han, X., Olsen, C.M., Neale, R.E., Lai, J., Vaughan, T.L., Gockel, I. *et al.* (2019) Gastroesophageal reflux GWAS identifies risk loci that also associate with subsequent severe esophageal diseases. *Nat. Commun.*, **10**, 4219.
66. Marseglia, L., Manti, S., D'Angelo, G., Gitto, E., Salpietro, C., Centorrino, A., Scalfari, G., Santoro, G., Impellizzeri, P. and Romeo, C. (2015) Gastroesophageal reflux and congenital gastrointestinal malformations. *World J. Gastroenterol.*, **21**, 8508–8515.
67. Busch, A., Žarković, M., Lowe, C., Jankofsky, M., Ganschow, R., Buers, I., Kurth, I., Reutter, H., Rutsch, F. and Hübner, C.A. (2017) Mutations in CRLF1 cause familial achalasia. *Clin. Genet.*, **92**, 104–108.
68. Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H. *et al.* (2020) A human cell atlas of fetal chromatin accessibility. *Science (New York, N.Y.)*, **370**, eaba7612.
69. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E. *et al.* (2020) Global reference mapping of human transcription factor footprints. *Nature*, **583**, 729–736.
70. van der Wijst, M., de Vries, D.H., Groot, H.E., Trynka, G., Hon, C.C., Bonder, M.J., Stegle, O., Nawijn, M.C., Idaghdour, Y., van der Harst, P. *et al.* (2020) The single-cell eQTLGen consortium. *elife*, **9**, e52155.

71. Kempfer, R. and Pombo, A. (2020) Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.*, **21**, 207–226.
72. Cha, J. and Lee, I. (2020) Single-cell network biology for resolving cellular heterogeneity in human diseases. *Exp. Mol. Med.*, **52**, 1798–1808.
73. Machiela, M.J. and Chanock, S.J. (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics (Oxford, England)*, **31**, 3555–3557.
74. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
75. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.
76. Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W. *et al.* (2018) The chromatin accessibility landscape of primary human cancers. *Science (New York, N.Y.)*, **362**, eaav1898.
77. Teng, L., He, B., Wang, J. and Tan, K. (2015) 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics (Oxford, England)*, **31**, 2560–2564.
78. Zeger, S.L. and Liang, K.Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.
79. Goldhoff, P., Warrington, N.M., Limbrick, D.D., Hope, A., Woerner, B.M., Jackson, E., Perry, A., Piwnica-Worms, D. and Rubin, J.B. (2008) Targeted inhibition of cyclic AMP phosphodiesterase-4 promotes brain tumor regression. *Clin. Cancer Res.*, **14**, 7717–7725.
80. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K. and Giraldez, A.J. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods*, **12**, 982–988.
81. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
82. Akdemir, K.C., Le, V.T., Chandran, S., Li, Y., Verhaak, R.G., Beroukhim, R., Campbell, P.J., Chin, L., Dixon, J.R., Futreal, P.A. *et al.* (2020) Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat. Genet.*, **52**, 294–305.
83. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.