OXFORD

Human Molecular Genetics, 2016, 1–8

doi: 10.1093/hmg/ddv512
Advance Access Publication Date: 23 December 2015
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Chronic gastroesophageal reflux disease shares genetic background with esophageal adenocarcinoma and Barrett's esophagus

Puya Gharahkhani[1,*], Joyce Tung[3], David Hinds[3], Aniket Mishra[1,4], Barrett's and Esophageal Adenocarcinoma Consortium (BEACON), Thomas L. Vaughan[5], David C. Whiteman[2] and Stuart MacGregor[1], on behalf of the BEACON study investigators

[1]Statistical Genetics Laboratory, [2]Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia, [3]23andMe, Mountain View, CA 94035, USA, [4]Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, The Netherlands and [5]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, USA

*To whom correspondence should be addressed. Tel: +61 733620219; Email: puya.gharahkhani@qimrberghofer.edu.au

Abstract

Esophageal adenocarcinoma (EA) is a rapidly fatal cancer with rising incidence in the developed world. Most EAs arise in a metaplastic epithelium, Barrett's esophagus (BE), which is associated with greatly increased risk of EA. One of the key risk factors for both BE and EA is chronic gastroesophageal reflux disease (GERD). This study used the linkage disequilibrium (LD) score regression and genomic profile risk scoring approaches to investigate the contribution of multiple common single-nucleotide polymorphisms (SNPs) to the risk of GERD, and the extent of genetic overlap between GERD and BE or EA. Using LD score regression, we estimated an overall phenotypic variance of 7% (95% CI 3–11%) for GERD explained by all the genotyped SNPs. A genetic correlation of 77% (s.e. = 24%, $P = 0.0012$) between GERD and BE and 88% between GERD and EA (s.e. = 25%, $P = 0.0004$) was estimated using the LD score regression approach. Results from the genomic profile risk scoring approach, as a robustness check, were broadly similar to those from the LD score regression. This study provides the first evidence for a polygenic basis for GERD and supports for a polygenic overlap between GERD and BE, and GERD and EA.

Introduction

Esophageal adenocarcinoma (EA) (OMIM: 614266) is a rapidly fatal cancer with rising incidence in the developed world. EA has a high mortality rate, with fewer than 20% of patients surviving 5 years (1). Barrett's esophagus (BE) (OMIM: 614266) is a precancerous metaplastic change of the normal stratified squamous epithelium of the esophagus to columnar epithelium containing goblet cells (2). Every year 0.3% of patients with BE develop EA (3). BE has prevalence of 1.6% while EA has a lifetime risk of 0.25% (http://seer.cancer.gov/archive/csr/1975_2009_pops09/, 1 March 2015, date last accessed) (4,5). Frequent and chronic gastroesophageal reflux (hereafter referred to as gastroesophageal reflux

Received: September 2, 2015. **Revised:** November 19, 2015. **Accepted:** December 10, 2015

disease or 'GERD') (OMIM: 109350) is the strongest known risk factor for both BE and EA (6,7). GERD has a prevalence of 18% in Western countries (8).

Genetic predisposition to GERD, BE and EA is incompletely understood. Previous twin studies estimated a heritability of 30–40% for GERD (9). Some other studies had suggested involvement of genetic factors in development of GERD, BE and EA (10–13). GWAS studies have identified four loci associated with development of BE and four additional loci associated with development of both BE and EA (14–16). GWAS studies have not identified genome-wide significant hits for GERD to date.

Some investigators have inferred a shared genetic basis between GERD, BE and EA because the risk for these diseases is increased when a relative is affected with any of these three diseases (17–20). Our recent study using genome-wide single-nucleotide polymorphisms (SNPs) found a significant genetic overlap between BE and EA, but not between GERD and BE or EA (21). Furthermore, while that study reported a significant contribution of common variants to BE and EA (array heritability 35 and 25%, respectively), the contribution of common genetic variants to risk of GERD was not significantly different from zero (21). However, that study was limited by small numbers of patients with GERD, which may have resulted in false-negative findings with regards to the genetic contribution to GERD and to the overlap between GERD and BE/EA. Thus, whether common genetic variants make any contribution to GERD predisposition and the extent of any genetic overlap between GERD and BE/EA remains open to question.

Accordingly, we aimed to investigate the contribution of common genetic variants to risk of GERD. Further aims were to estimate polygenic overlap and genetic correlation between GERD and BE or EA. Moreover, we aimed to examine for overlaps between the top loci associated with BE/EA and GERD and to investigate novel shared genetic loci between these diseases using meta-analysis of the GWAS results in the cohorts available for this study. In addition to the meta-analysis of the three diseases together, we also aimed to investigate whether meta-analysis of the GWAS results from the large GERD cohorts available for this study can help us identify novel genetic loci associated with GERD alone. Finally, we investigated whether there are any pathways enriched for GERD that are also enriched for BE or EA.

Results

Common SNPs contribute to the genetic basis of GERD

We used the linkage disequilibrium (LD) score regression approach (22,23) to estimate the overall phenotypic variance for GERD explained by cumulative effects of all the SNPs used in this study (SNP heritability) using 8 743 GERD cases and 43 932 controls from 23andMe. We estimated an SNP heritability of 7% (95% CI 3–11%) for GERD (on liability scale assuming prevalence of 18% for GERD).

For profile risk scoring, we used the 23andMe GERD data as the discovery set and 880 GERD cases and 1210 GERD free controls from Barrett's and Esophageal Adenocarcinoma Consortium (BEACON) who were not affected by BE or EA as the target set. Profile scores calculated using the top 1% of SNPs from the GWAS of GERD in the 23andMe discovery cohort significantly ($P = 2.2 \times 10^{-4}$) predicted risk of GERD in BEACON, the target cohort. Nagelkerke's pseudo R^2 as measure of goodness of fit was ~0.9% (Fig. 1A), corresponding to 0.9% of phenotypic variance on the liability scale explained by the risk scores (assuming prevalence of 18% for GERD). Risk scoring analysis produced consistent results before and after LD pruning (Fig. 1 and Supplementary Material, Fig. S1).

This is the first evidence to date for a polygenic role in the development of GERD. Increasing the proportions of SNPs from the top 1% to all the SNPs in the analysis reduced the Nagelkerke's pseudo R^2 in the BEACON cohort (Fig. 1A). This suggests that while there may be true positive SNPs ranked further down the list, the true positives are mixed in with a large number of null effect SNPs, such that the overall prediction decreases in accuracy.

Genetic overlap between GERD and BE or EA

The LD score regression approach gave heritability estimates (on liability scale) of 25% (s.e. = 5%) for BE and 16% (s.e. = 5%) for EA using 2410 BE cases, 1510 EA cases and 3203 controls from BEACON. Genetic correlation estimates were 77% (s.e. = 24%, $P = 0.0012$) between GERD (23andMe GERD data) and BE (BEACON data), and 88% (s.e. = 25%, $P = 0.0004$) between GERD (23andMe GERD data) and EA (BEACON data). The intercept of the cross-trait LD score regression was 0.0012 (s.e. = 0.0054) for GERD and BE bivariate analysis and −0.0092 (s.e. = 0.0055) for GERD and EA bivariate analysis, indicating that there was no sample overlap between 23andMe and BEACON cohorts.

For profile risk scoring, we used the 23andMe GERD data as the discovery set and BE or EA data from BEACON as the target set. Profile scores including all the SNPs (729 324 SNPs) from 23andMe significantly ($P = 3.6 \times 10^{-7}$) predicted risk of BE in the BEACON cohort with Nagelkerke's pseudo R^2 of 0.6% (Fig. 1B). The BE phenotypic variance on the liability scale explained by the risk scores was 0.3% based on the measure of the Nagelkerke's pseudo R^2 and assuming a prevalence of 1.6% for BE. The significance level and the measure of goodness of fit improved by increasing the proportion of SNPs in the analysis (Fig. 1B), consistent with there being a polygenic basis for BE. Risk scoring analysis produced consistent results before and after LD pruning (Fig. 1 and Supplementary Material, Fig. S1).

Similarly, we found the top 75% of GERD-derived SNPs significantly ($P = 0.009$) predicted risk of EA in BEACON cohort with Nagelkerke's pseudo R^2 of 0.2%. The EA phenotypic variance on the liability scale explained by the risk scores was 0.07% based on the measure of the Nagelkerke's pseudo R^2 and assuming the lifetime risk of 0.25% for EA. An increasing trend was observed in Nagelkerke's pseudo R^2 by increasing the proportion of top SNPs in the analysis (Fig. 1C). Risk scoring analysis after LD pruning produced results consistent with those of risk scoring analysis before LD pruning (Fig. 1 and Supplementary Material, Fig. S1); however, the results were not statistically significant after LD pruning (Supplementary Material, Fig. S1).

To ensure that the above results for BE and EA were not influenced by the individuals affected by GERD, we also performed the risk scoring analyses for BE and EA once after removing people affected by GERD and once after removing people unaffected by GERD from the analysis. 1173 BE cases and 2323 controls for the analysis without GERD (including 1473 with unknown GERD status), and 1237 BE cases and 880 controls for the analysis with GERD were used. Similarly, 910 EA cases and 2323 controls for the analysis without GERD (including 1384 with unknown GERD status), and 600 EA cases and 880 controls for the analysis with GERD were used. The results of these analyses (data not shown) were consistent with those obtained for BE and EA regardless of the GERD status. However, as expected, the results were less statistically significant owing to smaller sample sizes left for these analyses.

Hence, the results of this study show for the first time that there is a shared genetic basis (polygenic overlap) between GERD and BE and between GERD and EA.

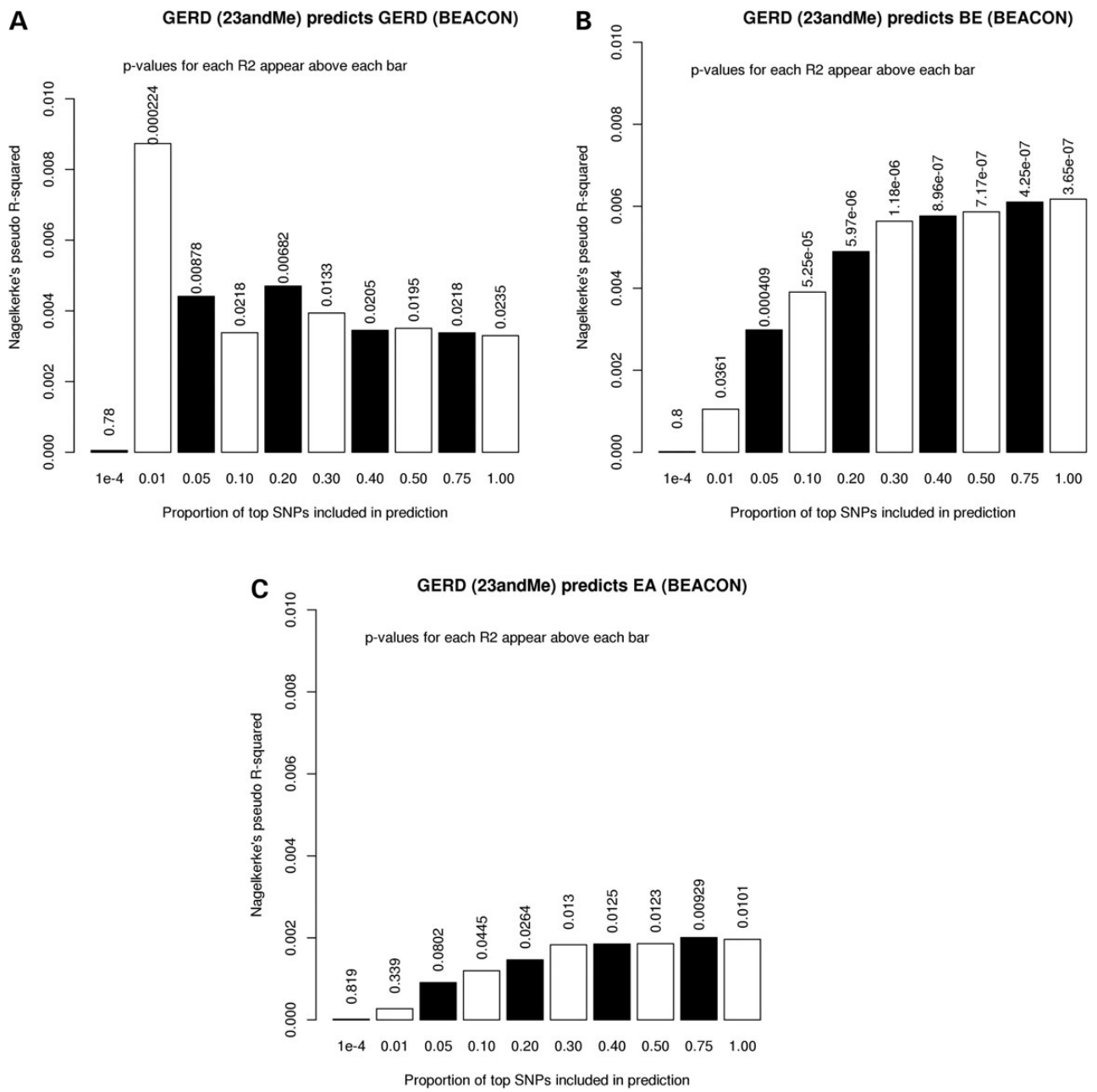

Figure 1. Prediction of risk for GERD, BE and EA in the BEACON cohort based on aggregation of SNP effects for GERD in 23andMe cohort. This figure summarizes the logistic regression results (with sex, age and the first four principal components fitted as covariates) for association of risk scores (calculated based on SNP effects obtained from GWAS of GERD in 23andMe cohort; more details are in the Materials and Methods section) with GERD, BE and EA status in BEACON cohort. The Y-axis shows the Nagelkerke's pseudo R^2 from the logistic regression as measure of goodness of fit. The X-axis shows the proportion of top GERD-associated SNPs used for calculation of the risk scores in BEACON cohort. The P-values for the Nagelkerke's pseudo R^2 are present above each bar. (**A**) GERD-associated SNPs in 23andMe predict risk for GERD in BEACON cohort (people who were not affected by BE or EA were included in this analysis). (**B**) GERD-associated SNPs in 23andMe predict risk for BE in BEACON (**C**) GERD-associated SNPs in 23andMe predict risk for EA in BEACON.

Does the genetic overlap extend to the known loci?

We also investigated whether the genetic overlap between GERD and BE/EA extends to the previously identified genome-wide significant SNPs for BE/EA (14–16). For this purpose, we investigated association of the index SNPs in the eight previously reported BE/EA loci with GERD in 23andMe cohort (Table 1). The index SNPs on chromosomes 19 and 3 (rs10419226 and rs2687201, respectively) were nominally associated with GERD ($P < 0.05$) (Table 1). Although the other six odds ratios (ORs) were weak, seven of the eight known BE/EA loci show the same direction of effect on GERD ($P = 0.04$ for exact binomial test for seven out of eight having the same direction of effect). However, the risk estimates were generally small for GERD (OR close to 1) and the 95% confidence intervals for ORs do not overlap with the estimates of the effects on EA/BE (Table 1).

Shared novel loci for BE/EA and GERD?

To investigate whether the genetic overlap between GERD and BE/EA could help us identify shared genetic loci associated with

Table 1. Association of the index SNPs in the previously identified BE/EA loci with the GERD in 23andMe cohort

					BE/EA in previous studies					GERD in 23andMe		
Chr	SNP	Position[a]	A1[b]	A2	Trait	Pubmed ID	*P*	OR	95% CI	*P*	OR	95% CI
19	rs10419226	18 803 172	T	G	BE/EA combined	24 121 790	3.6×10^{-10}	1.18	1.12–1.24	0.03832	1.04	1.00–1.07
9	rs11789015	96 716 028	G	A	BE/EA combined	24 121 790	1.0×10^{-9}	0.83	0.79–0.88	0.9152	0.998	0.96–1.04
3	rs2687201	70 928 930	A	C	BE/EA combined	24 121 790	5.5×10^{-9}	1.18	1.12–1.25	0.0252	1.04	1.01–1.09
16	rs9936833	86 403 118	G	A	BE	22 961 001	2.7×10^{-10}	1.14	1.10–1.19	0.6148	1.01	0.97–1.04
6	rs9257809	29 356 331	A	G	BE	22 961 001	4.1×10^{-9}	1.21	1.13–1.28	0.09692	1.05	0.99–1.12
2	rs3072	20 878 406	G	A	BE	25 447 851	1.8×10^{-11}	1.14	1.09–1.18	0.96065	0.999	0.96–1.03
12	rs2701108	11 467 4261	G	A	BE	25 447 851	7.5×10^{-9}	0.90	0.86–0.93	0.11538	0.972	0.94–1.01
15	rs3784262	58 253 106	G	A	BE/EA combined	25 447 851	3.7×10^{-9}	0.90	0.87–0.93	0.80412	0.996	0.96–1.03

Chr, chromosome; CI, confidence interval.
[a]Position of SNPs in build 37.
[b]Effect allele.

these diseases, we performed a fixed-effects meta-analysis of GWAS results of BE/EA/GERD combined. Here, cases were defined as being either BE/EA in BEACON (3920 BE/EA cases) or GERD in 23andMe (8743 GERD cases). Controls were no BE/EA in BEACON (3203 controls) or no GERD in 23andMe (43 932 controls). Association results for the top SNPs in BEACON cohort are shown in Supplementary Material, Table S1. None of the SNPs reached the genome-wide significance level ($P < 5 \times 10^{-8}$) in this meta-analysis (Supplementary Material, Table S2).

We also performed a multivariate GWAS using GWAS summary statistics from GERD in 23andMe and BE/EA in BEACON. Only rs10419226 on chromosome 19 was genome-wide significant ($P = 1.3 \times 10^{-8}$) in the multivariate GWAS. In the single trait GWAS, this SNP was also associated with BE ($P = 2.6 \times 10^{-8}$) and EA ($P = 1.83 \times 10^{-5}$) in the BEACON cohort, and nominally ($P = 0.038$) associated with GERD in 23andMe cohort.

We also performed a fixed-effects meta-analysis between the GWAS results for GERD in BEACON (2718 GERD cases and 2662 controls) and 23andMe (8743 GERD cases and 43 932 controls) to identify any loci associated with GERD at a genome-wide significance level ($P < 5 \times 10^{-8}$). None of the SNPs reached the genome-wide significance level ($P < 5 \times 10^{-8}$) in this meta-analysis (Supplementary Material, Table S3). The Q–Q and Manhattan plots for this meta-analysis are in Supplementary Material, Figures S2 and S3.

Shared pathways between GERD and BE, and GERD and EA

To investigate whether there is any pathway enriched in GERD that is also enriched in BE or EA, we performed pathway analysis using DEPICT approach (24). We performed separate analysis for each phenotype and then combined the pathway association test statistics for GERD and BE, and GERD and EA using Fisher's combined test. Supplementary Material, Tables S4 shows the top five pathways for each phenotype separately, and Supplementary Material, Table S5 shows the top five pathways for the combined analyses. None of the pathways were significant ($P < 10^{-6}$ Bonferroni-corrected significance threshold) in the pathway analysis for each phenotype separately as well as in the combined analyses. However, the 'transforming growth factor beta (TGF-β)-activated receptor activity' pathway (GO:0005024) came close to being significant ($P = 9.5 \times 10^{-6}$) in the combined GERD and BE analysis, followed by the 'TGF-β receptor binding' pathway (GO:0005160) ($P = 3.5 \times 10^{-5}$). However, these pathways were not significant for EA ($P = 0.60$ and $P = 0.07$ for GO:0005024 and GO:0005160, respectively).

Discussion

This study has two significant findings on genetic etiology of GERD. Firstly, for the first time, we showed that GERD has a polygenic background. The top 1% of SNPs associated with GERD in a discovery cohort significantly predicted the risk of GERD in an independent target cohort. Previous studies have not been able to detect any genome-wide significant ($P < 5 \times 10^{-8}$) hits for GERD. However, despite us demonstrating a significant polygenic basis to GERD, our analysis of the BEACON and 23andMe data (either taken individually or combined) failed to clearly identify (at genome-wide significant levels) any specific SNPs conferring GERD susceptibility. However, we estimated an overall phenotypic variance of 7% in GERD explained by the SNPs used in this study using the LD score regression approach. This explains a proportion of the heritability estimated from twin studies of GERD (9). As with other complex traits for which a polygenic basis has been established, ever larger sample sizes will enable identification of increasing numbers of genome-wide significant SNPs (25,26).

The second significant finding of this study is that GERD appears to have a significant genetic overlap with BE and EA. Using the LD score regression approach, we estimated a considerable genetic correlation between GERD and BE/EA, consistent with a significant polygenic overlap between GERD and BE/EA using the polygenic risk scoring approach. The overlap seems to involve large numbers of loci with small effects rather than small numbers of loci with large effects.

We used the profile risk scoring approach in this study as a robustness check to ensure that the findings are consistent with those obtained from the LD score regression approach. This is because the profile risk scoring approach uses the raw data for the target sets when compared with the LD score regression approach that uses only summary statistics. In addition, profile risk scoring is useful to infer what proportions of SNPs better explain the genetic overlap between the discovery set and the target set. The results obtained from both approaches were consistent with a polygenic background for GERD and a significant genetic overlap between GERD and BE and GERD and EA.

We previously showed that combined GWAS analysis for BE and EA [which have a high genetic correlation (rg = 1)] increased the statistical power to detect novel loci for those diseases (14). As we identified a considerable polygenic overlap and genetic correlation between GERD and BE or EA in this study, combined GWAS analysis or meta-analysis for these diseases may increase the statistical power to detect loci with pleiotropic effects. Despite this, combined GWAS analysis may not always result in

increased statistical power, especially when the genetic correlation between the traits is not high, and for loci without pleiotropic effects. We did not detect any shared loci at genome-wide significance level in a meta-analysis between the GWAS results for BE/EA combined, and GERD. This might be explained, at least in part, by the polygenic basis of these three diseases where large numbers of loci contribute to their development, each with a small effect size. More powerful meta-analyses with larger sample sizes will be required to unambiguously identify specific loci at genome-wide significance level for these diseases.

Although the polygenic nature of these traits has made it difficult to identify specific loci, rs10419226 on chromosome 19 was genome-wide significant in the multivariate GWAS for correlated traits using GWAS summary statistics (27). This locus was previously identified as a risk locus for BE and EA (14). Although not associated with GERD at genome-wide significant level in the single trait GWAS, multivariate GWAS in this study suggests that this locus may also be a risk factor for GERD.

While we used estimates of the prevalence of BE and GERD in general population to estimate SNP heritability of these diseases on liability scale, the lifetime risk was used for EA. This is because prevalence is a suitable measure of case burden in the general population for conditions that are not fatal. As EA is a fatal cancer with a low survival rate, the prevalence of the disease in general population is low; hence, we used estimates of lifetime risk for EA in this study.

The results of this study suggest that our previous study, in which we did not detect a significant genetic overlap between GERD and BE or EA (21), was adversely affected by the small sample size available for GERD in that study.

Shared genetic background between GERD, BE and EA has also clinical implications. For example, it shows that people suffering from GERD may have a genetic architecture that puts them in a higher risk of developing BE or EA. A recent study showed that using multiple correlated traits could significantly increase risk prediction accuracy (28). Dissecting the genetic overlap between these diseases can also be helpful in increasing our knowledge of the etiology of these diseases and in future treatments targeting shared molecular pathways involved in pathogenesis of these diseases.

Epidemiological studies show that frequent gastroesophageal reflux is associated with BE and EA (29,30). Mendelian randomization approaches may be useful to help determine whether the association is causal. Our data showing genetic overlap between GERD, BE and EA are consistent with causality; however, our sample size here is not large enough to enable robust conclusions to be drawn using Mendelian randomization.

The combined pathways analysis in this study suggested that 'TGF-β-activated receptor activity' might be a common pathway involved in the development of GERD and BE. This pathway involves combining with and transmitting signal from TGF-β for subsequent catalytic activities within cells (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005024, 1 May 2015, date last accessed). Several studies have shown that changes in the expression of TGF-β and its signaling mediators are involved in the BE carcinogenesis and EA progression (31–33). On the other hand, it has been suggested that esophageal exposure to gastric acid in GERD promotes tissue remodeling through TGF-β1 stimulation of the differentiation of fibroblasts into myofibroblasts (34). Increased expression of TGF-α and epidermal growth factor receptors have also been reported in rat chronic reflux esophagitis (35). These findings suggest that changes in the expression of TGF-β and its signaling mediators could play an important role in the development of GERD and BE, and progression to EA.

In summary, we showed that GERD has a polygenic basis with a significant genetic overlap with BE and EA. These results suggest that combining data from GERD, BE and EA will help in future identification of shared genetic loci between these diseases. Our work has increased our knowledge of the etiology of these diseases and in the future this may inform development of treatments targeting the shared molecular pathways.

Materials and Methods

23andMe cohort

The 23andMe discovery cohort included summary statistics from a GWAS of 8743 GERD cases versus 43 932 controls of European ancestry. GERD cases were defined as people who had been diagnosed with GERD by medical doctors, and controls were people who had never been diagnosed with heartburn, acid reflux or GERD (Supplementary Material). Details of genotyping methods, quality control, imputation and association analysis are in Supplementary Material.

Barrett's and Esophageal Adenocarcinoma Consortium (BEACON) cohort

Following data cleaning 2410 BE cases, 1510 EA cases and 3203 controls were available from BEACON cohort for this study. Approximately 1020 individuals from the controls were 'MD Anderson controls', the cancer-free individuals of European ancestry who were screened for melanoma at the MD Anderson Cancer Center in Houston, Texas. For GERD, 2718 cases and 2662 controls from BEACON were used in this study. Histological confirmation of EA was carried out for all the participating studies in the BEACON cohort. BE was defined as identification of goblet cells in intestinal metaplastic columnar epithelium in biopsy taken from tubular esophagus. GERD data were collected via standardized questionnaires, usually through personal interviews. The BEACON cohort and the participating studies have been previously used for studies on BE, EA and GERD (6,14,21,36,37). Details of genotyping methods and the quality control are in Supplementary Material.

LD score regression

We used the LD score regression approach (22,23) to estimate SNP heritability for GERD, BE and EA as well as to estimate genetic correlation between GERD and BE or EA. The intercept of the cross-trait LD score regression was used as an estimate to ensure that there was no significant sample overlap between the individuals participated in 23andMe study and those participated in the BEACON [an intercept close to zero indicates no sample overlap (22)].

Risk scores

We used the genomic profile risk scoring method as a robustness check to ensure that the findings are consistent between this method and those form the LD score regression approach. We investigated genetic overlap between GERD and BE or EA by using a large number of autosomal SNPs demonstrated to be associated with GERD in 23andMe cohort to predict the risk of GERD, BE and EA in the BEACON cohort. This method (referred to as 'profile risk scoring') has been used previously in studies exploring the genetic architecture of the other complex traits such as schizophrenia and bipolar disorder (38) and endometriosis (39). Prespecified subsets of the top SNPs associated with GERD (top 10% *P*-values, top 20% *P*-values . . .) from the 23andMe discovery

cohort were used to score the individuals in the target cohort (BEACON). The score for each individual in BEACON was calculated by summing the number of risk alleles weighted by their effect sizes obtained from the GWAS of GERD in the 23andMe discovery cohort. Profile risk scoring analysis was performed with and without LD pruning. LD pruning was done in PLINK (40) by removing SNPs in high LD ($r^2 > 0.2$) with SNPs having the best *P*-values across 500 kb regions. After LD pruning, 155 168 independent SNPs were taken forward for analysis.

Statistical analysis

To estimate the association of the profile scores with the GERD, BE and EA status in BEACON target cohort, we performed a logistic regression with sex, age and the first four principal components used as covariates. To obtain a better measure of goodness of fit for the genetic risk scoring analysis, Nagelkerke's pseudo R^2 as a measure of model fit from the logistic regression was converted to R^2 measure on the liability scale using the approach described previously (41). Profile risk scoring for GERD in BEACON was performed for the individuals who were not affected by BE or EA (BEACON population controls who were free of both BE and EA, including 880 GERD cases and 1210 GERD free controls).

Genome-wide association analysis for GERD and combined BE/EA status in BEACON cohort was performed in PLINK (40) using an additive model with sex, age and the first four principal components fitted as covariates (the GWAS summary statistics are publically available at http://www.qimrberghofer.edu.au/chronic-gastroesophageal-reflux-disease/, 8 December 2015, date last accessed). Summary results from these GWAS analyses were then used to perform meta-analysis between GERD in 23andMe and GERD or BE/EA in BEACON. We also performed a multivariate GWAS for correlated traits using the single trait GWAS summary statistics for GERD, BE and EA, as described previously (27).

Meta-analysis

Fixed-effects meta-analysis was performed in METAL (42) between the GWAS summary results of GERD in 23andMe cohort and GERD or combined BE/EA in the BEACON cohort. The presence of heterogeneity between the cohorts was investigated using the I2 statistic, as implemented in METAL.

Pathway analysis

We used DEPICT (24) to perform pathway analysis for GERD, BE and EA. Further, we meta-analyzed the pathway association test statistics for GERD and BE, and GERD and EA using Fisher's combined test.

Consortia

Members of BEACON consortium

Marilie D. Gammon, Douglas A. Corley, Nicholas J. Shaheen, Nigel C. Bird, Laura J. Hardie, Liam J. Murray, Brian J. Reid, Wong-Ho Chow, Harvey A. Risch, Weimin Ye, Geoffrey Liu, Yvonne Romero, Leslie Bernstein, Anna H. Wu, David C. Whiteman and Thomas L. Vaughan.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We thank 23andMe for providing us with GERD GWAS summary data. We also thank the customers and employees of 23andMe for making this work possible. We acknowledge the database of Genotypes and Phenotypes (dbGaP) for providing access approval to the MD Anderson controls data.

Conflict of Interest statement. None declared.

Funding

This work was supported by funding from the US National Cancer Institute at the National Institutes of Health (grant number R01CA136725 awarded to T.L.V. and D.C.W.). S.M. is supported by Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC) Fellowships. D.C.W. is supported by Australian NHMRC Fellowship. The Swedish Esophageal Cancer Study was funded by grants (R01 CA57947-03) from the National Cancer Institute, and the Swedish Cancer Society (4559-B01-01XAA and 4758-B02-01XAB). The Kaiser Permanente Study was supported by US NIH grants R01DK63616 and R01CA59636) and from the California Tobacco Related Research Program (3RT-0122 and 10RT-0251). The MD Anderson controls were drawn from dbGaP (study accession: phs000187.v1.p1) following approval. Genotyping of these controls (C. Amos, PI) were performed through the University of Texas MD Anderson Cancer Center (UTMDACC) and the Center for Inherited Disease Research (CIDR), supported in part by NIH grants R01CA100264, P30CA016672 and R01CA133996, the UTMDACC NIH SPORE in Melanoma 2P50CA093459, as well as by the Marit Peterson Fund for Melanoma Research. CIDR is supported by contract HHSN268200782096C. Funding to pay the Open Access publication charges for this article was provided by the Australian Research Council.

References

1. Holmes, R.S. and Vaughan, T.L. (2007) Epidemiology and pathogenesis of esophageal cancer. *Semin. Radiat. Oncol.*, **17**, 2–9.
2. Wong, D.J., Paulson, T.G., Prevo, L.J., Galipeau, P.C., Longton, G., Blount, P.L. and Reid, B.J. (2001) p16(INK4a) lesions are common, early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. *Cancer Res.*, **61**, 8284–8289.
3. Desai, T.K., Krishnan, K., Samala, N., Singh, J., Cluley, J., Perla, S. and Howden, C.W. (2012) The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut*, **61**, 970–976.
4. Ronkainen, J., Aro, P., Storskrubb, T., Johansson, S.E., Lind, T., Bolling-Sternevald, E., Vieth, M., Stolte, M., Talley, N.J. and Agreus, L. (2005) Prevalence of Barrett's esophagus in the general population: an endoscopic study. *Gastroenterology*, **129**, 1825–1831.
5. Powell, J., McConkey, C.C., Gillison, E.W. and Spychal, R.T. (2002) Continuing rising trend in oesophageal adenocarcinoma. *Int. J. Cancer*, **102**, 422–427.
6. 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.
7. 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.

8. Mohammed, I., Cherkas, L.F., Riley, S.A., Spector, T.D. and Trudgill, N.J. (2003) Genetic influences in gastro-oesophageal reflux disease: a twin study. *Gut*, **52**, 1085–1089.
9. Cameron, A.J., Lagergren, J., Henriksson, C., Nyren, O., Locke, G.R. III and Pedersen, N.L. (2002) Gastroesophageal reflux disease in monozygotic and dizygotic twins. *Gastroenterology*, **122**, 55–59.
10. Poynton, A.R., Walsh, T.N., O'Sullivan, G. and Hennessy, T.P. (1996) Carcinoma arising in familial Barrett's esophagus. *Am. J. Gastroenterol.*, **91**, 1855–1856.
11. Romero, Y., Cameron, A.J., Locke, G.R., Schaid, D.J., Slezak, J.M., Branch, C.D. and Melton, L.J. (1997) Familial aggregation of gastroesophageal reflux in patients with Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology*, **113**, 1449–1456.
12. Chak, A., Faulx, A., Kinnard, M., Brock, W., Willis, J., Wiesner, G.L., Parrado, A.R. and Goddard, K.A. (2004) Identification of Barrett's esophagus in relatives by endoscopic screening. *Am. J. Gastroenterol.*, **99**, 2107–2114.
13. Fitzgerald, R.C. (2005) Complex diseases in gastroenterology and hepatology: GERD, Barrett's, and esophageal adenocarcinoma. *Clin. Gastroenterol. Hepatol.*, **3**, 529–537.
14. 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.
15. 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.
16. Palles, C., Chegwidden, L., Li, X., Findlay, J.M., Farnham, G., Castro Giner, F., 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.
17. Chak, A., Lee, T., Kinnard, M.F., Brock, W., Faulx, A., Willis, J., Cooper, G.S., Sivak, M.V. Jr. and Goddard, K.A. (2002) Familial aggregation of Barrett's oesophagus, oesophageal adenocarcinoma, and oesophagogastric junctional adenocarcinoma in Caucasian adults. *Gut*, **51**, 323–328.
18. Munitiz, V., Parrilla, P., Ortiz, A., Martinez-de-Haro, L.F., Yelamos, J. and Molina, J. (2008) High risk of malignancy in familial Barrett's esophagus: presentation of one family. *J. Clin. Gastroenterol.*, **42**, 806–809.
19. Groves, C., Jankowski, J., Barker, F. and Holdstock, G. (2005) A family history of Barrett's oesophagus: another risk factor? *Scand. J. Gastroenterol.*, **40**, 1127–1128.
20. Sun, X., Chandar, A.K., Elston, R. and Chak, A. (2014) What we know and what we need to know about familial gastroesophageal reflux disease and Barrett's esophagus. *Clin. Gastroenterol. Hepatol.*, **12**, 1664–1666.
21. Ek, W.E., Levine, D.M., D'Amato, M., Pedersen, N.L., Magnusson, P.K., Bresso, F., Onstad, L.E., Schmidt, P.T., Tornblom, 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.
22. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., ReproGen_Consortium, Psychiatric_Genomics_Consortium, Genetic_Consortium_for_Anorexia_Nervosa_of_the_Wellcome_Trust_Case_Control_Consortium_3, , Duncan, L. *et al.* (2015) An atlas of genetic correlations across human diseases and traits. *Nat. Genet.*, **47**, 1236–1241.
23. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S. and Yang, J., Schizophrenia_Working_Group_of_the_Psychiatric_Genomics_Consortium, Patterson, N., Daly, M.J. and Neale, A.L.P.B.M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.
24. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T. *et al.* (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.*, **6**, 5890.
25. Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.
26. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z. *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.*, **46**, 1173–1186.
27. Bolormaa, S., Pryce, J.E., Reverter, A., Zhang, Y., Barendse, W., Kemper, K., Tier, B., Savin, K., Hayes, B.J. and Goddard, M.E. (2014) A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.*, **10**, e1004198.
28. Maier, R., Moser, G., Chen, G.B., Ripke, S., Coryell, W., Potash, J.B., Scheftner, W.A., Shi, J., Weissman, M.M., Hultman, C.M. *et al.* (2015) Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am. J. Hum. Genet.*, **96**, 283–294.
29. Kamangar, F., Chow, W.H., Abnet, C.C. and Dawsey, S.M. (2009) Environmental causes of esophageal cancer. *Gastroenterol. Clin. North Am.*, **38**, 27–57, vii.
30. 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. Biomarkers Prev.*, **14**, 2481–2486.
31. von Rahden, B.H., Stein, H.J., Feith, M., Puhringer, F., Theisen, J., Siewert, J.R. and Sarbia, M. (2006) Overexpression of TGF-beta1 in esophageal (Barrett's) adenocarcinoma is associated with advanced stage of disease and poor prognosis. *Mol. Carcinog.*, **45**, 786–794.
32. Onwuegbusi, B.A., Aitchison, A., Chin, S.F., Kranjac, T., Mills, I., Huang, Y., Lao-Sirieix, P., Caldas, C. and Fitzgerald, R.C. (2006) Impaired transforming growth factor beta signalling in Barrett's carcinogenesis due to frequent SMAD4 inactivation. *Gut*, **55**, 764–774.
33. Hong, S., Lee, H.J., Kim, S.J. and Hahm, K.B. (2010) Connection between inflammation and carcinogenesis in gastrointestinal tract: focus on TGF-beta signaling. *World J. Gastroenterol.*, **16**, 2080–2093.
34. Jarmuz, T., Roser, S., Rivera, H., Gal, A. and Roman, J. (2004) Transforming growth factor-beta1, myofibroblasts, and tissue remodeling in the pathogenesis of tracheal injury: potential role of gastroesophageal reflux. *Ann. Otol. Rhinol. Laryngol.*, **113**, 488–497.
35. Fujiwara, Y., Higuchi, K., Hamaguchi, M., Takashima, T., Watanabe, T., Tominaga, K., Oshitani, N., Matsumoto, T. and Arakawa, T. (2004) Increased expression of transforming growth factor-alpha and epidermal growth factor receptors in rat chronic reflux esophagitis. *J. Gastroenterol. Hepatol.*, **19**, 521–527.
36. Kubo, A., Cook, M.B., Shaheen, N.J., Vaughan, T.L., Whiteman, D.C., Murray, L. and Corley, D.A. (2013) 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*, **62**, 1684–1691.
37. 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.
38. Purcell, S., Wray, N., Stone, J., Visscher, P., O'Donovan, M., Sullivan, P. and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748–752.
39. Painter, J., Anderson, C., Nyholt, D., Macgregor, S., Lin, J., Lee, S., Lambert, A., Zhao, Z., Roseman, F., Guo, Q. *et al.* (2011) Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat. Genet.*, **43**, 51–54.
40. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
41. Lee, S.H., Goddard, M.E., Wray, N.R. and Visscher, P.M. (2012) A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.*, **36**, 214–224.
42. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.