First genome-wide association study of esophageal atresia identifies three genetic risk loci at CTNNA3, FOXF1/FOXC2/FOXL1, and HNF1B

Jan Gehlen, 1,60 Ann-Sophie Giel, 1,60 Ricarda Köllges, 2,3,60 Stephan L. Haas, 3 Rong Zhang, 2,3,3 Jiri Trcka, 4 Ayse Ö. Sungur, 5 Florian Renziehausen, 5,3 Dorothea Bornholdt, 1 Daphne Jung, 1,6 Paul D. Hoyer, 6 Agneta Nordenskjöld, 7,8 Dick Tibboel, 9 John Vlot, 9 Manon C.W. Spaander, 10 Robert Smigiel, 11 Dariusz Patkowski, 12 Nel Rheevel, 13 Iris AML. van Rooij, 13 Ivo de Blauw, 14 Alice Hölscher, 15 Marcus Pauly, 16 Andreas Leutner, 17 Joerg Fuchs, 18 Joel Niethammer, 18 Maria-Theodora Melissari, 19 Ekkehart Jenetzky, 20,21 Nadine Zwink, 21 Holger Thiele, 22 Alina Christine Hilger, 2,23 Timo Hess, 1 Jessica Trautmann, 2,3,3 Matthias Marks, 24,25 Martin Baumgarten, 1 Gaby Bläss, 26 Mikael Landén, 27,28

Summary

Esophageal atresia with or without tracheoesophageal fistula (EA/TEF) is the most common congenital malformation of the upper digestive tract. This study represents the first genome-wide association study (GWAS) to identify risk loci for EA/TEF. We used a European case-control sample comprising 764 EA/TEF patients and 5,778 controls and observed genome-wide significant associations at three loci. On chromosome 10q21 within the gene CTNNA3 (p = 2.11 × 10⁻⁷; odds ratio [OR] = 3.94; 95% confidence interval [CI], 3.10–5.00), on chromosome 16q24 next to the FOX gene cluster (p = 2.25 × 10⁻¹⁰; OR = 1.47; 95% CI, 1.38–1.55) and on chromosome 17q12 next to the gene HNF1B (p = 3.35 × 10⁻¹⁰; OR = 1.75; 95% CI, 1.64–1.87). We next carried out an esophageal/tracheal transcriptome profiling in rat embryos at four selected embryonic time points. Based on these data and on already published data, the implicated genes at all three GWAS loci are promising candidates for EA/TEF development. We also analyzed the genetic EA/TEF architecture beyond the single marker level, which revealed an estimated single-nucleotide polymorphism (SNP)-based heritability of around 37% ± 14% standard deviation. In addition, we examined the polygenicity of EA/TEF and found that EA/TEF is less polygenic than other complex genetic diseases. In conclusion, the results of our study contribute to a better understanding on the underlying genetic architecture of ET/TEF with the identification of three risk loci and candidate genes.

Esophageal atresia with or without tracheoesophageal fistula (EA/TEF) (MIM: 189960) is the most common developmental malformation of the upper digestive tract. EA/TEF comprises five anatomical subtypes that are classified according to location and tracheoesophageal connection. The most common subtype is Gross type C. Here, the proximal esophagus ends blindly in the upper mediastinum, whereas the distal esophagus forms a fistula with

(Author list continued on next page)

(Affiliations continued on next page)
the trachea. Around 50% of EA/TEF patients present with additional malformations,3,4 the majority of these are defects from the VATER/VACTERL association spectrum (MIM: 192350).5,6

The birth prevalence of EA/TEF is around 1 in 4,1007 in the European population and the recurrence risk in children of affected individuals is approximately 1%–2.4%.8–12 Thus, children of EA/TEF patients display an 80-fold increase in recurrence risk compared with the general population. Besides this formal evidence for the involvement of genetic factors in disease etiology, research has shown that 6%–10% of EA/TEF patients carry chromosomal aberrations.3,4,6,8,13 The most frequently observed aberrations are trisomy 13, 18, and 21 as well as 13q-, 17q-, and 22q11-deletion syndromes.13 Further evidence for the involvement of genetic factors in EA/TEF etiology is provided by the existence of more than 50 distinct EA/TEF-related genetic syndromes, associations, and sequences.6,13,14 For several of these syndromes the causal genes are known; for example, MYCN in Feingold syndrome 1 (MIM: 164280),5 GLI2 in Pallister-Hall syndrome (MIM: 146510),16 and CHD7 in CHARGE syndrome (MIM: 214800).17 While these studies have generated insights into the etiology of syndromic patients, the genetic causes of non-syndromic EA/TEF remain elusive. The vast majority of these cases most likely develop EA/TEF from a complex genetic background.

This study represents the first genome-wide association study (GWAS) to identify genetic risk factors for non-syndromic EA/TEF. In total we analyzed 764 patients and 5,778 ethnically matched controls who were part of three different case-control cohorts. From Germany the Netherlands 450 cases and 4,420 controls were included, from Sweden 167 cases and 971 controls, and from Poland/Czech Republic 147 cases and 387 controls. A detailed description of all cohorts is provided as supplemental information. All cohorts were genotyped using high-density single-nucleotide polymorphism (SNP) arrays. After quality control (QC) all three case-control cohorts were subjected to imputation using the TOPMed Imputation Server and TOPMed Reference panel.18–20 In total, 9,140,693 imputed SNPs with R² > 0.3 and a minor allele frequency (MAF) of >1% in cases and a minor allele count of >20 in cases and controls combined were then tested separately in each cohort using PLINK2 logistic regression with Firth correction.21 For this analysis, the first five principal components were used as covariates. After computing the associations at single cohort level we performed a meta-analysis considering a fixed-effects inverse variance-weighting approach using METAL.22 A detailed description of the genome-wide genotype data, QC, imputation, and association analyses is provided as supplemental information. The genome inflation factor lambda of the resulting GWAS meta-analysis was 1.08.
The Q-Q and Manhattan plots are shown in Figures S1 and S2. In total, three genomic regions—on chromosomes 10q21, 16q24, and 17q12—showed genome-wide significant association with EA/TEF (Table 1).

On chromosome 10, the variant with the most significant association to EA/TEF was rs187017665 (p = 2.11 × 10^{-8}; odds ratio [OR] = 3.94; 95% confidence interval [CI], 3.10–5.00) (Table 1; Figure 1). The SNP is located within intron 15 of the gene CTNNA3 (catenin alpha 3) (MIM: 607667), which encodes an actin-filament binding protein that is a critical subunit of the cadherin-catenin cell-cell adhesion complex.23 Although CTNNA3 has not been described in the context of EA/TEF so far, it represents an interesting candidate gene. At the protein level, CTNNA3 is a direct binding partner of CTNNB1 (an interesting candidate gene. At the protein level, CTNNA3 is a direct binding partner of CTNNB1, which plays a key role in the proliferation and cellular differentiation of the intestinal epithelium during embryogenesis.24 Accordingly, inactivation of CTNNB1 in mice leads to a TEF phenotype, which is one of the most established models to study the biology of EA/TEF development.25 The risk allele of rs187017665 is rare in the European population (MAF < 0.005% according to gnomAD).26 Accordingly, public genomic data are sparse for excluding that a deleterious coding variant in CTNNA3 is in linkage disequilibrium (LD) with the risk variant at this locus. For that reason, we selected five patients carrying the EA/TEF-contributing risk allele at rs187017665 and sequenced all 17 coding exons of CTNNA3 (supplemental information). However, we could not identify any functional-relevant variant within CTNNA3 (missense, nonsense, or location in the splice-donor or effector site) (data not shown).

On chromosome 16, the SNP with the most significant EA/TEF association was rs8046904 (p = 2.25 × 10^{-10}; OR = 1.47; 95% CI, 1.38–1.55) (Table 1; Figure 1). The risk locus harbors the FOX gene cluster, which comprises FOXF1 (forkhead box F1) (MIM: 601089), FOXC2 (forkhead box C2) (MIM: 602402), and FOXL1 (forkhead box L1) (MIM: 603252). All three genes play an essential role in foregut development.25 Furthermore, Stankiewicz et al. identified risk conferring de novo duplications comprising HNF1B in patients with syndromic EA/TEF.33 Furthermore, TEF and other atresia phenotypes, such as biliary atresia and duodenal atresia, have been reported in patients with de novo HNF1B deletions or point mutations.34–36 Of the GWAS-associated SNPs at this locus, rs3094503 is the most probable causative EA/TEF variant since it is not in high LD with any other SNPs (r^2 < 0.5). In addition, rs3094503 has a CADD score37 of 21 suggesting a possible deleterious effect of this variant. Furthermore, according to HaploReg v.4.1,38 rs3094503 is located in an enhancer that is active in numerous embryonic and gastrointestinal tissues and leads to the alteration of multiple transcription factor binding sites. However, no study to date has reported any direct influence of this variant on HNF1B expression or function, maybe due to its temporo-spatial specific expression.

We next used esophageal/tracheal transcriptome data from rat embryos to get first insights into the gene expression profile at the identified EA/TEF GWAS loci. A detailed description of this analysis is provided as supplemental information. In brief, transcriptome-wide expression analyses were performed using esophageal/tracheal tissue from four selected embryonic time points (embryonic day 11 [E11], E12, E13, and E14). Research indicates that these time points are crucial in terms of EA/TEF development.25 In addition, esophageal tissue from E21 was used to represent a time point of no relevance to EA/TEF development. For each time point esophageal/tracheal RNA from five embryos were used and expression profiling was performed using QuantSeq (Lexogen, Vienna, Austria) and a HiSeq 2500 platform (Illumina, San Diego, USA). QuantSeq is an RNA quantification method, in which next-generation sequencing libraries are sequenced at the end of the 3’ poly(A) tail and then quantified in a subsequent analysis.39 Of all expressed rat embryonic transcripts, 14,075 could be annotated in the human genome via Ensembl biomaRt homology mapping.39 In each of the three implicated GWAS regions we focused on the expression of all genes localized within a 500 kb window surrounding the most significant associated EA/TEF variant. The corresponding genes at each locus are shown in Figure 1. On human chromosome 10q21 we observed a significant differential expression of Ctnna3 in embryonic tissues (mean expression E11–E14) compared with E21.
Table 1. SNPs with genome-wide significant association in the present EA/TEF GWAS

| SNP     | Chr | Pos in bp | EA   | OA       | p         | OR   | 95% CI   | Meta-analysis | Germany/the Netherlands | Poland/Czech Republic | Sweden |
|---------|-----|-----------|------|----------|-----------|------|----------|--------------|------------------------|----------------------|--------|
|         |     |           |      |          |           |      |          | Ca | Co | OR         | 95% CI  | Ca | Co | OR         | 95% CI  | Ca | Co | OR         | 95% CI  | Ca | Co | OR         | 95% CI  |
| rs187017665 | 10  | 66,039,525 | A    | G        | 2.11 × 10^-8 | 3.94 | 3.10-5.00 | 0.021 | 0.006 | 1.23 × 10^-7 | 4.42 | 3.36-5.81 | 0.003 | 0.001 | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 0.018 | 0.007 | 0.043 |
| rs183405336 | 10  | 66,076,833 | C    | T        | 3.03 × 10^-8 | 3.87 | 3.05-4.91 | 0.021 | 0.006 | 1.43 × 10^-7 | 4.39 | 3.34-5.77 | 0.003 | 0.001 | – | – | – | – | – | – | – | – | – | – | – | – | – | 0.018 | 0.008 | 0.051 |
| rs552788622 | 10  | 66,236,447 | C    | CTG      | 4.78 × 10^-8 | 3.78 | 2.98-4.79 | 0.021 | 0.006 | 2.05 × 10^-7 | 4.29 | 3.27-5.64 | 0.003 | 0.004 | – | – | – | – | – | – | – | – | – | – | – | – | 0.018 | 0.008 | 0.055 |
| rs931458   | 16  | 86,372,699 | A    | C        | 1.43 × 10^-3 | 1.43 | 1.35-1.51 | 0.637 | 0.556 | 2.40 × 10^-7 | 1.46 | 1.36-1.57 | 0.633 | 0.531 | 0.007 | 1.56 | 1.33-1.83 | 0.623 | 0.563 | 0.047 | 1.27 | 1.13-1.43 |
| rs8046904  | 16  | 86,373,131 | G    | C        | 2.25 × 10^-10 | 1.47 | 1.38-1.55 | 0.666 | 0.585 | 2.73 × 10^-7 | 1.48 | 1.37-1.59 | 0.637 | 0.541 | 0.019 | 1.48 | 1.26-1.74 | 0.677 | 0.592 | 0.004 | 1.43 | 1.27-1.62 |
| rs8069412  | 17  | 37,667,572 | C    | CT       | 5.90 × 10^-10 | 1.55 | 1.45-1.67 | 0.208 | 0.153 | 2.23 × 10^-3 | 1.47 | 1.35-1.61 | 0.172 | 0.163 | 0.747 | 1.07 | 0.87-1.32 | 0.298 | 0.171 | 1.42 | 2.04 | 1.79-2.33 |
| rs8069412  | 17  | 37,670,030 | G    | A        | 2.59 × 10^-10 | 1.57 | 1.46-1.68 | 0.209 | 0.153 | 1.44 × 10^-3 | 1.48 | 1.36-1.62 | 0.173 | 0.166 | 0.812 | 1.05 | 0.85-1.30 | 0.300 | 0.171 | 6.11 × 10^-8 | 2.09 | 1.83-2.38 |
| rs3094503  | 17  | 37,670,407 | C    | A        | 3.35 × 10^-16 | 1.75 | 1.64-1.87 | 0.268 | 0.179 | 3.35 × 10^-11 | 1.76 | 1.62-1.92 | 0.198 | 0.180 | 0.699 | 1.08 | 0.88-1.33 | 0.310 | 0.181 | 5.80 × 10^-8 | 2.10 | 1.84-2.40 |

For each variant the chromosomal position (Chr, Pos) is given in bp according to hg38. The allele frequency for each SNP in cases (Ca) and controls (Co) refers to the effect allele (EA). The allele frequency of the other allele (OA) is not shown. p values (p), odds ratio (OR), and the corresponding 95% confidence interval (CI) are shown. The frequency of risk alleles for SNPs on chromosome 10q21 were too small (<0.005) in Polish/Czech controls, and no association was determined. In addition, no rs number is available for one associated SNP on chromosome 17q12. In Table S1 the frequency of genotypes is shown for all listed variants in each cohort together with information on Hardy-Weinberg equilibrium (HWE) in Table S2. Of note, none of these SNPs showed a deviation of HWE (all p > 0.16) indicating that copy number variants encompassing the implicated regions are unlikely to be present in a considerable proportion of patients. The high imputation quality of all listed SNPs in all samples can be obtained from Table S3 (all SNPs with an imputation quality score > 0.9). Furthermore, for all listed SNPs marginal significant heterogeneity of association across cohorts were observed only for two SNPs on chromosome 17 (Table S4). Table S5 lists all associations after applying genomic control to the GWAS meta-analysis dataset. While the associations on chromosome 16q24 and 17q12 remain genome-wide significant after this control, the association on chromosome 10q21 is slightly above the threshold of genome-wide significance (p = 7.00 × 10^-8 for rs187017665).
(outside of the relevant developmental time frame) \((p = 5.53 \times 10^{-4})\) (Figure S3). On human chromosome 16q24 only Foxc2 of the FOX gene cluster showed a significant differential expression at embryonic stages \((p = 5.17 \times 10^{-3})\). In addition, we observed a slightly less significant differential embryonic expression of Irf8 at this locus \((p = 1.32 \times 10^{-4})\) (Figure S3). On human chromosome 17q12 the embryonic expression of Hnf1b differed most significantly compared with E21 and other genes at this locus \((p = 1.85 \times 10^{-8})\) (Figure S3). However, also Ddx52 and Tada2a showed a significant differential embryonic expression at this locus \((p = 2.35 \times 10^{-6} \text{ and } p = 6.1 \times 10^{-7})\) (Figure S3). The data further support that the genes CTNNA3, FOXC2, and HNF1B may be involved in EA/TEF development and, thus, represent promising candidates for upcoming functional studies.

Finally, we examined the genetic EA/TEF architecture beyond the single marker level. We used the LD score regression method, which allows the collective analysis of common GWAS variants to estimate the SNP-based heritability. This revealed that a substantial fraction of EA/TEF heritability is polygenic with an estimated SNP-based heritability of 37.64\% \pm 14.17\% standard deviation (SD). The large SD most probably reflects the relatively small study sample. The GWAS data were also used to annotate and prioritize EA/TEF relevant SNPs and genes using FUMA. FUMA is an integrative web-based tool that applies information from multiple biological resources to facilitate functional annotation of GWAS results and prioritization of disease genes. However, none of the FUMA findings using our EA/TEF GWAS data were significant (data not shown), most likely because FUMA mainly uses data derived from adult tissues. As in research into other developmental diseases—for example, non-syndromic cleft lip with or without cleft palate (nsCL/P)—this study generated comparatively high heritability estimates and genome-wide significant associations using a relatively small GWAS sample. The limited amount of tissue affected in developmental diseases and the fact that they arise within a narrow embryonic time-period suggest that the genetic architecture of developmental conditions is less complex than that of other multifactorial disorders. To test this hypothesis we applied GENESIS, which uses GWAS data and allows the estimation of effect-size distribution for genetic risk variants underlying complex genetic phenotypes. GENESIS has shown before that there is a wide diversity in the degree and nature of polygenicity across different complex genetic traits, with major depressive disorder (MDD) as a psychiatric disease being most polygenic and Crohn’s disease (CD) as an inflammatory...
The analysis with our EA/TEF GWAS data and GWAS data on MDD and CD provided by GENESIS showed that EA/TEF is indeed comparatively less polygenic (Figure 2). The comparison of the number of SNPs in the tail regions of effect-size distributions showed that EA/TEF has distinctly larger numbers of SNPs with moderate-to-large effects than MDD as a psychiatric disease. Finally, we analyzed another GWAS data set to determine whether other developmental disorders are also comparatively less polygenic. For this we used GWAS data for nsCL/P (399 cases, 1,318 controls). We found that this developmental disorder is also less polygenic than other complex genetic phenotypes (Figure 2). In summary, we present the first GWAS in EA/TEF that led to the identification of three genome-wide significant associated disease loci. The lead risk variant on chromosome 10 is located within CTNNB1, which is a direct interaction partner of CTNNB1 on the protein level. Notably, CTNNB1 inactivation leads to TEF in mice, which represents one of the most studied EA/TEF animal models. The EA/TEF risk locus on chromosome 16 harbors the FOX gene cluster and on chromosome 17 HNF1B, all of these genes play an important role in foregut development. Moreover, de novo deletions and/or duplications involving these genes have been already implicated in EA/TEF development. Based on these findings, functional studies are now required to identify the underlying disease mechanisms and downstream pathways. Furthermore, our data show that a substantial fraction of EA/TEF heritability is polygenic with an estimated SNP-based heritability of around 37% ± 14% SD. Given our GWAS sample size, this is a comparatively high estimate suggesting that EA/TEF as a developmental disorder is less polygenic than other complex genetic traits, which we could confirm using GENESIS.

Ethical statement

The authors state that their study complies with the Declaration of Helsinki, that the locally appointed ethics committees have approved all research protocols and that informed written consent has been obtained from all study participants or their parents prior to the inclusion of subjects into the study.

Data availability

The datasets and codes supporting this study have not been deposited in a public repository but are available from the corresponding author on request.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100093.

Acknowledgments

The authors thank all study participants and the German self-help organization KEKS e.V. (Kinder und Erwachsene mit Erkrankungen der Speiseröhre) for their assistance with recruitment. The authors also thank the Heinz Nixdorf Foundation for access to control subjects from the cohort of the Heinz Nixdorf Recall Study (HNR). In addition, the authors thank the following research funding organizations: Else Kröner-Fresenius Foundation (grant 2014_A14), Deutsche Forschungsgemeinschaft (DFG) (grants LU 731/3-1, LU 1944/3-1, RE 1723/1-1, RE 1723/2-1, RE 1723/1-3, TH 1327/1-1, WO 1732/4-2, Exs147-2), the LOEWE Center for Cell and Gene Therapy Frankfurt (LOEWE-CGT), the German Cancer Aid (Deutsche Krebshilfe) (grant 70113577), and the Sophia Foundations for Scientific Research (grant SW0013-09). Swedish controls were collected as part of the Anorexia Nervosa Genetics Initiative (ANGI), which is an initiative of the Klarmann Family Foundation.

Declaration of interests

The co-author C.M.B. declares the following interests: Shire (grant recipient, Scientific Advisory Board member), Idorsia (consultant), Lundbeckfonden (grant recipient), Pearson (author, royalty recipient), and Equip Health Inc. (Clinical Advisory Board). All other co-authors declare no competing interests.

Received: October 25, 2021
Accepted: January 20, 2022

Web resources

CADD, https://cadd.gs.washington.edu/
FUMA, https://fuma.ctglab.nl/
GENESIS, https://github.com/yandorazhang/GENESIS
gnomAD, https://gnomad.broadinstitute.org/
HaploReg, https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php.
METAL, https://genome.sph.umich.edu/wiki/METAL.
OMIM, http://omim.org/
PLINK, https://zzz.bwh.harvard.edu/plink/
TOPMed Imputation server, https://imputation.biodatacatalyst.nhlbi.nih.gov/
LocusZoom, http://locuszoom.org/

References

1. Torfs, C.P., Curry, C.J., and Bateson, T.F. (1995). Population-based study of tracheoesophageal fistula and esophageal atresia. Teratology 52, 220–232.
2. Spitz, L. (2007). Oesophageal atresia. Orphanet J. Rare Dis. 2, 24.
3. David, T.J., and O’Callaghan, S.E. (1975). Oesophageal atresia in the South West of England. J. Med. Genet. 12, 1–11.
4. Geneviève, D., de Pontual, L., Amiel, J., Sarnacki, S., and Lyonnet, S. (2007). An overview of isolated and syndromic oesophageal atresia. Clin. Genet. 71, 392–399.
5. Brown, A.K., Roddam, A.W., Spitz, L., and Ward, S.J. (1999). Oesophageal atresia, related malformations, and medical problems: a family study. Am. J. Med. Genet. 85, 31–37.
17q12 microdeletion including HNF1B: a new associated malformation in this syndrome. Am. J. Med. Genet. A 164A, 3076–3082.

37. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 47, D886–D894.

38. Moll, P., Ante, M., Seitz, A., and Reda, T. (2014). QuantSeq 3’ mRNA sequencing for RNA quantification. Nat. Methods 11, 972.

39. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191.

40. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.K., ReproGen Consortium; Psychiatric Genomics Consortium; and Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241.

41. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826.

42. Mangold, E., Ludwig, K.U., Birnbaum, S., Baluardo, C., Ferrian, M., Herms, S., Reutter, H., de Assis, N.A., Chawa, T.A., Mattheisen, M., et al. (2010). Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. Nat. Genet. 42, 24–26.

43. Zhang, Y., Qi, G., Park, J.H., and Chatterjee, N. (2018). Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat. Genet. 50, 1318–1326.