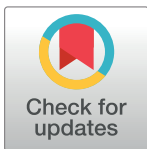


RESEARCH ARTICLE

Human exome and mouse embryonic expression data implicate *ZFH3*, *TRPS1*, and *CHD7* in human esophageal atresia

Rong Zhang¹, Jan Gehlen², Amit Kawalia³, Maria-Theodora Melissari⁴, Tikam Chand Dakal⁵, Athira M. Menon⁵, Julia Höfele⁶, Korbinian Riedhammer^{6,7}, Lea Waffenschmidt¹, Julia Fabian¹, Katinka Breuer¹, Jeshurun Kalanithy¹, Alina Christine Hilger⁸, Amit Sharma^{9,10}, Alice Hölscher¹¹, Thomas M. Boemers¹¹, Markus Pauly¹², Andreas Leutner¹³, Jörg Fuchs¹⁴, Guido Seitz¹⁵, Barbara M. Ludwikowski¹⁶, Barbara Gomez¹⁶, Jochen Hubertus¹⁷, Andreas Heydweiller¹⁸, Ralf Kurz¹⁸, Johannes Leonhardt¹⁹, Ferdinand Kosch²⁰, Stefan Holland-Cunz²¹, Oliver Münsterer²², Beno Ure²³, Eberhard Schmiedeke²⁴, Jörg Nesper²⁵, Petra Degenhardt²⁶, Stefanie Märzheuser²⁷, Katharina Kleine²⁸, Mattias Schäfer²⁹, Nicole Spsychalski²⁹, Oliver J. Deffaa³⁰, Jan-Hendrik Gosemann³⁰, Martin Lacher³⁰, Stefanie Heilmann-Heimbach^{1,31}, Nadine Zwick³², Ekkehart Jenetzky^{32,33}, Michael Ludwig³⁴, Phillip Grote⁴, Johannes Schumacher^{1,2}, Holger Thiele³⁵, Heiko Reutter^{1,35}*



OPEN ACCESS

Citation: Zhang R, Gehlen J, Kawalia A, Melissari M-T, Dakal TC, Menon AM, et al. (2020) Human exome and mouse embryonic expression data implicate *ZFH3*, *TRPS1*, and *CHD7* in human esophageal atresia. PLoS ONE 15(6): e0234246. <https://doi.org/10.1371/journal.pone.0234246>

Editor: Regie Lyn Pastor Santos-Cortez, University of Colorado School of Medicine, UNITED STATES

Received: February 11, 2020

Accepted: May 21, 2020

Published: June 5, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0234246>

Copyright: © 2020 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

1 Institute of Human Genetics, Medical Faculty of Bonn, University of Bonn, Bonn, Germany, **2** Institute of Human Genetics, University Hospital of Marburg, Marburg, Germany, **3** Cologne Center for Genomics, University of Cologne, Cologne, Germany, **4** Institute of Cardiovascular Regeneration, Center for Molecular Medicine, University of Frankfurt, Frankfurt am Main, Germany, **5** Department of Biotechnology, Mohanlal Sukhadia University Udaipur, Rajasthan, India, **6** Institute of Human Genetics, Klinikum Rechts der Isar, Technical University of Munich, School of Medicine, Munich, Germany, **7** Department of Nephrology, Klinikum Rechts der Isar, Technical University of Munich, School of Medicine, Munich, Germany, **8** Department of Pediatrics, Clinic for Pediatrics, University Hospital Bonn, Bonn, Germany, **9** Department of Neurology, University Hospital Bonn, Bonn, Germany, **10** Department of Ophthalmology, University Hospital Bonn, Bonn, Germany, **11** Department of Pediatric Surgery and Urology, University Hospital Cologne, Cologne, Germany, **12** Department of Pediatric Surgery, Asklepios Children's Hospital St. Augustin, St. Augustin, Germany, **13** Department of Pediatric Surgery, Medical Center Dortmund, Dortmund, Germany, **14** Department of Pediatric Surgery Children's Hospital, University of Tübingen, Tübingen, Germany, **15** Department of Pediatric Surgery, University of Marburg, Marburg, Germany, **16** Department of Pediatric Surgery and Pediatric Urology, Medical Center for Children and Adolescents AUF DER BULT, Hannover, Germany, **17** Department of Pediatric Surgery, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University, Munich, Germany, **18** Department of Pediatric Surgery, University Hospital Bonn, Bonn, Germany, **19** Department of Pediatric Surgery, Children's Hospital Braunschweig, Braunschweig, Germany, **20** Department of Pediatric Surgery, Städtisches Klinikum Karlsruhe, Karlsruhe, Germany, **21** Department of Pediatric Surgery, University Children's Hospital Basel (UKBB), Basel, Switzerland, **22** Department of Pediatric Surgery, University Medicine Mainz, Mainz, Germany, **23** Center of Pediatric Surgery Hannover, Hannover Medical School, Hannover, Germany, **24** Clinic for Paediatric Surgery and Paediatric Urology, Klinikum Bremen-Mitte, Bremen, Germany, **25** Department of Pediatric Surgery, General Hospital, Chemnitz, Germany, **26** Department of Pediatric Surgery, Ernst von Bergmann Hospital, Potsdam, Germany, **27** Department of Pediatric Surgery, Campus Virchow Clinic, Charité University Hospital Berlin, Berlin, Germany, **28** Department of Pediatric Surgery, Evangelisches Krankenhaus Oberhausen, Germany, **29** Department of Pediatric Surgery and Urology, Cnopf'sche Kinderklinik, Nürnberg, Germany, **30** Department of Pediatric Surgery, University of Leipzig, Leipzig, Germany, **31** Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany, **32** Department of Child and Adolescent Psychiatry and Psychotherapy, Johannes-Gutenberg University, Mainz, Germany, **33** Institute of Integrative Medicine, Witten/Herdecke University, Herdecke, Germany, **34** Department of Clinical Chemistry and Clinical Pharmacology, University of Bonn, Bonn, Germany, **35** Section of Neonatology and Pediatric Intensive Care, Clinic for Pediatrics, University Hospital Bonn, Bonn, Germany

* These authors contributed equally to this work.

* reutter@uni-bonn.de

Funding: F.K. was supported by a stipend of the University of Bonn, BONFOR (O-149.0115.1; <https://www.medfak.uni-bonn.de/de/forschung/foerderung/interne-foerderung/bonfor>). H.R., J.S., M.L., and P.G. are supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) (BE 3910/6-1, RE 1723/4-1, Exc147-2; <https://www.dfg.de/>). H.R., J.S., H.T., and E.J. are further supported by a grant of the Else-Kröner-Fresenius-Stiftung (EKFS, 2014_A14; <https://www.ekfs.de/>). The Exome analysis was performed on CHEOPS, a high performance computer cluster of the regional data center (RRZK) of the University of Cologne, funded by the DFG (215828658). The transcriptome analysis were performed on the de.NBI cloud, a national infrastructure supported by the German Federal Ministry of Education and Research (FKZ 031A532-0331A540 and 031L0101-0310108). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Introduction

Esophageal atresia with or without tracheoesophageal fistula (EA/TEF) occurs approximately 1 in 3.500 live births representing the most common malformation of the upper digestive tract. Only half a century ago, EA/TEF was fatal among affected newborns suggesting that the steady birth prevalence might in parts be due to mutational *de novo* events in genes involved in foregut development.

Methods

To identify mutational *de novo* events in EA/TEF patients, we surveyed the exome of 30 case-parent trios. Identified and confirmed *de novo* variants were prioritized using *in silico* prediction tools. To investigate the embryonic role of genes harboring prioritized *de novo* variants we performed targeted analysis of mouse transcriptome data of esophageal tissue obtained at the embryonic day (E) E8.5, E12.5, and postnatal.

Results

In total we prioritized 14 novel *de novo* variants in 14 different genes (*APOL2*, *EEF1D*, *CHD7*, *FANCB*, *GGT6*, *KIAA0556*, *NFX1*, *NPR2*, *PIGC*, *SLC5A2*, *TANC2*, *TRPS1*, *UBA3*, and *ZFHX3*) and eight rare *de novo* variants in eight additional genes (*CELSR1*, *CLP1*, *GPR133*, *HPS3*, *MTA3*, *PLEC*, *STAB1*, and *PPIP5K2*). Through personal communication during the project, we identified an additional EA/TEF case-parent trio with a rare *de novo* variant in *ZFHX3*. *In silico* prediction analysis of the identified variants and comparative analysis of mouse transcriptome data of esophageal tissue obtained at E8.5, E12.5, and postnatal prioritized *CHD7*, *TRPS1*, and *ZFHX3* as EA/TEF candidate genes. Re-sequencing of *ZFHX3* in additional 192 EA/TEF patients did not identify further putative EA/TEF-associated variants.

Conclusion

Our study suggests that rare mutational *de novo* events in genes involved in foregut development contribute to the development of EA/TEF.

Introduction

Esophageal atresia with or without tracheoesophageal fistula (EA/TEF) occur approximately 1 in 3.000 to 3.500 live births representing the most common malformation of the upper digestive tract [1; 2; 3; 4]. According to the “European network of population-based registries for the epidemiological surveillance of congenital anomalies (EUROCAT) EA/TEF account for 1% of all birth defects in Europe every year (<https://eu-rd-platform.jrc.ec.europa.eu/eurocat>). Hence, with 5.075 million babies born in the EU in 2017, 1.237 babies have been born with EA/TEF.

In about 40–50% of cases, EA/TEF occurs within the context of additional anomalies mostly belonging to the VATER/VACTERL association (OMIM #192350) spectrum. This acronym

refers to the rare, nonrandom co-occurrence of the following component features (CFs): vertebral defects (V), anorectal malformations (A), cardiac defects (C), tracheoesophageal fistula with or without esophageal atresia (TE), renal malformations (R), and limb defects (L) [5]. Only half a century ago, EA/TEF was fatal among affected newborns suggesting that the steady birth prevalence might in parts be due to mutational *de novo* events in genes involved in fore-gut development. Support for this hypothesis comes from early reports of chromosomal *de novo* aberrations present in 6–10% of syndromic EA/TEF cases [6]. Furthermore, using copy number variation (CNV) analysis in 375 EA/TEF patients we identified eight rare CNVs in six patients, all of which occurred *de novo*, including one CNV previously associated with EA/TEF [7]. Hence, 1.55% of isolated EA/TEF patients and 1.62% of patients with additional congenital anomalies carried *de novo* CNVs. Moreover, several monogenic EA/TEF associated syndromes are caused by smaller *de novo* changes comprising single nucleotides or small indels e.g. *N-MYC* in Feingold syndrome (OMIM #164280), *GLI2* in Pallister-Hall syndrome (OMIM #146510), *CHD7* in CHARGE syndrome (OMIM #214800), and *SOX2* in AEG syndrome (OMIM #206900) [8; 9; 10; 11].

To further explore the involvement of small genetic *de novo* events in the etiology of EA/TEF, we profiled 30 case-parents trios using exome sequencing (ES). Prior to ES chromosomal microarray analysis was negative in all cases [7; 12]. All confirmed *de novo* variants were prioritized using *in silico* prediction tools. To investigate the embryonic role of genes harboring prioritized *de novo* variants we performed targeted analysis of mouse transcriptome data of esophageal tissue obtained at embryonic day (E) E8.5, E12.5, and postnatal.

Materials and methods

Patients and DNA isolation

In 2011, the authors JS and HR founded the scientific network “great” (genetic risk for esophageal atresia; www.great-konsortium.de). The “great network” was founded in order to initiate a nationwide investigation into the genetic causes of EA/TEF. Prior to the commencement of recruitment, the network partners generated a unique standardized case report form (CRF). The CRF comprises an epidemiological questionnaire and a clinical assessment battery. The epidemiological questionnaire is based on: (i) the National Birth Defect Prevention Study questionnaire of the U.S. Centers of Disease Control and Prevention (www.nbdpn.org); and (ii) the questionnaire of the European Surveillance of Congenital Malformations (EUROCAT) network (www.eurocat-network.eu). The clinical assessment battery comprises the classification system of the EA/TEF phenotype according to Gross (1953), and the ICD10 coding with the British Pediatric Association one digit extension (www.eurocat-network.eu/content/EUROCAT-Guide-1.3.pdf) for classification of additional congenital anomalies. The great cohort is being recruited with the support of pediatric surgical departments across Germany, and the German self-help organization for patients and families with EA/TEF (KEKS e.V.; www.keks.org). KEKS e.V. is the largest self-help organization for EA/TEF families in Europe, and supports both the ongoing great investigations and the present proposal.

The here described study fulfilled the requirement of the Declaration of Helsinki and ethical approval was obtained from the local ethic committee of the Medical Faculty of Bonn (Lfd. Nr. 073/12). Every participating family provided written informed consent. The 30 here reported case-parent trios as well as the EA/TEF cohort for resequencing of *ZFHX3*, were recruited through the efforts of the scientific network “great”. In 14 of the 30 case-parent trios, EA/TEF occurred isolated/nonsyndromic. In the remaining case-parent trios EA/TEF co-occurred with additional phenotypic features (syndromic cases) mostly belonging to the VATER/VACTERL spectrum (S1 Table). From each case-parent trio, EDTA blood samples were obtained.

Genomic DNA was isolated using the Chemagic DNA Blood Kit special (Chemagen, Baesweiler, Germany). Through personal communication we identified another patient with EA/TEF as part of his VATER/VACTERL association (patient 750_501).

Exome Sequencing (ES) and data analysis

Exome capture was performed using the NimbleGen SeqCap EZ Human Exome Library v2.0 enrichment kit and sequenced with an Illumina paired end 2x100 bp sequencing (protocol v1.2). Primary data was filtered according to signal purity by the Illumina Realtime Analysis (RTA) software v1.8. Subsequently, reads were mapped to the human genome reference build hg19 using the bwa-aln [13] alignment algorithm. GATK v1.6 [14] was used to mark duplicated reads, for local realignment around short insertions and deletions, to recalibrate the base quality scores and to call SNVs (incorporating variants quality score recalibration) and short indels [15]. Scripts developed in-house at the Cologne Center for Genomics (unpublished) were used to incorporate allele frequencies reported by the ESP6500 database [Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: <http://evs.gs.washington.edu/EVS/>)] and to detect changes in the protein structure. Acceptor and donor splice site mutations were analyzed with a Maximum Entropy model [16]. *De novo* variant calling was performed with the program DeNovoGear (v.0.5.1) [17]. The Varbank GUI (unpublished, <https://varbank.ccg.uni-koeln.de>) was used to filter for high quality (coverage > 15; quality > 25), rare (MAF < 0.005), *de novo* (posterior probability of a *de novo* mutation = PP_DNM > 0.5) variants predicted to alter protein structure or splicing. We also filtered against an in-house database containing all variants from 511 exomes from epilepsy patients to exclude pipeline-related artefacts (MAF < 0.004). Variants with MAF < 0.004 that have been described to occur homozygous in gnomAD were also excluded. Finally, we further excluded all variants with a MAF \geq 0.0003 since the EA/TEF birth prevalence has been reported to be 1 in 3,500 live births (frequency of \approx 0.0003). Hence, (full penetrant) monoallelic variants with a MAF \geq 0.0003 cannot account for the occurrence of EA/TEF.

Variant validation and classification

Variants identified by ES were validated by using polymerase chain reaction (PCR). Automated sequence analysis was carried out using standard procedures. In brief, primers were directed to all variants observed and the resultant PCR products were subjected to direct automated BigDye Terminator sequencing (3130XL Genetic Analyzer, Applied Biosystems, Foster City, California, USA). Both strands from each amplicon were sequenced for the presence of these variants in the respective case-parent trio. In order to further prioritize the identified and confirmed *de novo* variants, we analyzed them using ten different *in silico* prediction tools which are encountered in dbNSFP v3.0 (<https://sites.google.com/site/jpopgen/dbNSFP>): SIFT, LRT, MutationTaster, Mutation Assessor, FATHMM, PROVEAN, MetaSVM, MetaLR, fathmm-MKL coding and CADD [18; 19] (details about these prediction tools are given as supporting information S1 Data).

Re-sequencing of ZFHX3 in EA/TEF patients

All three human *ZFHX3* protein coding transcripts (ENST00000641206.2, ENST00000268489.10, and ENST00000397992.5) listed in 'ensembl database' (www.ensembl.org/ Ensembl Release 98 (September 2019)) were sequenced in 192 unrelated EA/TEF patients. PCR-amplified DNA products (primer sequences available upon request) were subjected to sequencing using a 3130XL Genetic Analyzer (Applied Biosystems, Foster City, USA).

Structural modeling and in-silico analysis of ZFHX3 protein variants

The secondary structure prediction of human OCTs protein sequences was done using PSIPRED in I-Tasser. Three-dimensional protein structural models for ZFHX3 were built using SWISS-MODEL (<https://swissmodel.expasy.org/>). Since, Swiss model cannot handle large protein sequence, for the prediction of ZFHX3 *de novo* changes p.Pro534Arg and p.Ala2126Val we trimmed the sequence of 60 amino acids upstream and 20 downstream of the mutated site. The sequence was subjected to swiss-model based modeling. The structural comparison between wild-type and mutant variant was done in Chimera after superimposing the structure of mutant onto the wild structure using SuperPose using default parameters (superpose.wishartlab.com).

RNA isolation and mRNA library preparation of mouse embryonic esophageal tissue

All animals used in this study were anesthetized by Isoflurane and killed by cervical dislocation. The animals that were used in this study are documented and their usage reported to the local authorities Regierungspräsidium Darmstadt). Embryos from pregnant females of the C57Bl6J strain were harvested at embryonic days (E) E8.5, E12.5, and postnatal. The embryos of the E8.5 litter were determined to be of the developmental Theiler stage 13 (TS13) and the E12.5 embryos TS21. From E8.5 embryos, the pharyngeal pouch containing endoderm and adjacent mesoderm tissue was surgically isolated and transferred into QIAzol[®]. Multiple embryos were pooled for each embryonal timepoint. For the E8.5 stage we pooled biopsies from 5 embryos to prepare the RNA and for the E12.5 and neonates we pooled two each for RNA preparation. From E12.5 and postnatal embryos, the distinct structures of the esophagus and the trachea was surgically isolated, combined and transferred into QIAzol[®]. RNA was isolated from these tissues with the RNEasy Mini Plus Kit (Qiagen) according to the manufacturer's protocols. The transcriptome profile was assessed by RNA-Sequencing with the 3'-mRNASeq Library Preparation Kit from Lexogen, (Lexogen, Vienna, Austria). This protocol generates for each transcript only one single-end strand specific fragment for sequencing at the 3'-end of poly(A)-RNA. Libraries were quality checked on a TapeStation2200 (Agilent, Santa Clara, USA). The sequencing was performed on a HiSeq 2500 (Illumina, San Diego, USA) with two technical replicates of sample.<http://www.bioinformatics.babraham.ac.uk/projects/>

Transcriptome analysis

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2654802/> After demultiplexing with bcl2fastq (Illumina, San Diego, USA), FastQC v0.11.8 (<http://www.bioinformatics.babraham.ac.uk/projects/>) was used for quality control of FASTQ files. Read alignment was performed using STAR_2.6.1d [20] against the primary assembly of murine genome reference build GRCm38 according to the manufacturer's analysis protocol. Read counting was also performed with STAR ("quantMode GeneCounts") using the Ensembl gene annotation (Release 97). Quality metrics were gathered with multiQC [21].

Statistical analyses were performed with the programming language R (R Core Team, 2019) and the DESeq2 R package [22] Technical replicates were combined and differential gene expression considering the embryonal timepoint was assessed with the DESeq2's Wald test as described in Love et al. [23] We required an alpha level of 0.01 and a minimum log₂ fold-change of log₂(1.5). Cumulative expression distributions were calculated for *rlog* normalized

expression values for each timepoint separately. We identified the mouse homologous genes of our human genes of interest using the biomaRt R package [24].

Results

ES analysis

ES analysis identified 25 apparent *de novo* variants in 25 genes in 18 unrelated case-parent trios. Confirmation of these variants using Sanger sequencing validated all of them and confirmed 23 as being *de novo* in patients. 14 of these variants were novel according to the “Genome Aggregation Database (gnomAD; <https://gnomad.broadinstitute.org/>; November 2019)”. In addition, eight of the confirmed *de novo* variants were found to be rare with a minor allele frequency (MAF) between 0.000003–0.00003 (Table 1). One confirmed *de novo* variant in *TPP2* (c.1534G>A, p.Val512Ile, NM_003291.2, rs73578896) has been previously reported in gnomAD with a MAF of 0.002 (303/275.928) and was therefore filtered out. Through personal communication during the project (Dr. Julia Höfele, Institute of Human Genetics, Klinikum Rechts der Isar, Technical University of Munich, School of Medicine, Munich, Germany) we identified an additional EA/TEF case-parent trio (750_501) in which the patient carries a rare *de novo* variant in *ZFHX3* (c.6377C>T, p.Ala2126Val, allele frequency 0.000019) (marked with an asterisks in Table 1).

Table 1. Prioritized *de novo* variants.

Ext-Code	Phenotype	Variant	HGNC	Refseq	gnomAD (MAF)	MutCDNA	MutProt	Mm	Gg	Dr	Xt
4_501	VATER/VACTERL-like association	1	<i>EEF1D</i>	NM_032378.4		c.874C>T	p.Arg292*	K	E		K
		2	<i>CELSR1</i>	NM_014246.1	3/282,594 (0.00001)	c.4357G>A	p.Val1453Ile	V	I		I
21_501	nonsyndromic	3	<i>HPS3</i>	NM_032383.3	10/282,776 (0.00004)	c.1189C>T	p.Arg397Trp	H	R	R	R
27_501	nonsyndromic	4	<i>PIGC</i>	NM_153747.1		c.716C>T	p.Ala239Val	A		G	A
35_501	VATER/VACTERL-like association	5	<i>NFX1</i>	NM_002504.4		c.1723G>A	p.Val575Met	V			V
36_501	nonsyndromic	6	ZFHX3	NM_006885.3		c.1601C>G	p.Pro534Arg	P	P	P	N
41_501	VATER/VACTERL-like association	7	<i>MTA3</i>	NM_020744.2	1/237,600 (0.000004)	c.393C>A	p.Phe131Leu	F		F	
46_501	nonsyndromic	8	FANCB	NM_152633.2		c.782G>A	p.Arg261Gln	R	Q	S	
		9	<i>PLEC</i>	NM_201379.1	17/272,690 (0.00006)	c.6704G>A	p.Arg2394His	R	R	K	R
63_501	VATER/VACTERL-like association	10	<i>PPIP5K2</i>	NM_015216.2	2/247,732 (0.000008)	c.686G>A	p.Arg229Gln	R	R	R	R
88_501	nonsyndromic	11	<i>CLP1</i>	NM_006831.2	1/251,486 (0.000003)	c.814C>A	p.His272Asn	H	H	H	H
		12	<i>GPR133</i>	NM_198827.3	6/282,534 (0.00002)	c.1033G>A	p.Ala345Thr	A			
		13	<i>SLC5A2</i>	NM_003041.3		c.644T>C	p.Leu215Pro	L		L	L
90_501	VATER/VACTERL-like association	14	KIAA0556	NM_015202.2		c.3730C>T	p.His1244Tyr	H	H	H	H
141_501	VATER/VACTERL-like association	15	<i>STAB1</i>	NM_015136.2	9/278,948 (0.00003)	c.6145C>T	p.Arg2049Cys	R		S	
154_501	VATER/VACTERL association	16	<i>GGT6</i>	NM_153338.2		c.1045A>G	p.Ser349Gly	S			
167_501	nonsyndromic	17	CHD7	NM_017780.3		c.4187C>G	p.Ala1396Gly	A	A	A	
172_501	VATER/VACTERL-like association	18	<i>NPR2</i>	NM_003995.3		c.952C>G	p.Arg318Gly	R	K		T
174_501	nonsyndromic	19	<i>UBA3</i>	NM_198195.1		c.1088C>T	p.Ser363Phe	S	S	T	P
181_501	nonsyndromic	20	<i>TANC2</i>	NM_025185.3		c.2357C>T	p.Pro786Leu	P		P	P
288_501	VATER/VACTERL association	21	TRPS1	NM_014112.2		c.1630C>T	p.Arg544*	R		R	R
		22	<i>APOL2</i>	NM_145637.1		c.319G>C	p.Glu107Gln	D			
750_501*	VATER/VACTERL association	23	ZFHX3	NM_006885.3	5/250,880 (0.00002)	c.6377C>T	p.Ala2126Val	A	T	T	A

Annotations marked in bold red represent: “known disease genes” involved in the formation of congenital malformations, variants with truncating consequence, variants in highly conserved regions of the protein, or novel variants (not found in (n.f.i.), gnomAD (MAF)).

<https://doi.org/10.1371/journal.pone.0234246.t001>

Among the novel variants (i) five reside within previously described disease genes (*CHD7*, *FANCB*, *TRPS1*, *KIAA0556*, and *ZFHX3*), (ii) two variants were truncating (c.874C>T (p.Arg292*) in *EEF1D* and c.1630C>T (p.Arg544*) in *TRPS1*), and (iii) three amino acid changes (p.Arg229Gln in *PIIP5K2*; p.His272Asn in *CLP1*; p.His1244Ytyr in *KIAA0556*) reside in highly conserved regions of the respective protein (Table 1). Of the novel *de novo* variants constituting missense variants four amino acid changes (p.Pro534Arg in *ZFHX3*; p.Phe131Leu in *MTA3*; p.Leu215Pro in *SLC5A2*; p.Ala1396Gly in *CHD7*) were called deleterious by at least seven out of nine *in silico* prediction tools (written in bold in Table 2). Similarly, the rare *de novo* variant in *ZFHX3* (c.6377C>T, p.Ala2126Val, allele frequency 0.000019) found in the additional case-parent trio (750_501), was also called deleterious by seven out of nine *in silico* prediction tools (written in bold in Table 2).

One of the novel *de novo* variants (*PIGC*, c.716C>T, CADD score 11.6) and three of the rare *de novo* variants (*CELSR1*, c.4357G>A, CADD score 15.3; *CLP1*, c.814C>A, CADD score 17.4; *ZFHX3*, c.6377C>T, CADD score 19.2) reached CADD scores between 10 and 20 indicating that these variants have been predicted to be among the 10% most deleterious substitutions within the human genome. Nine of the novel *de novo* variants (*EEF1D*, c.874C>T, CADD score 28.5; *NFX1*, c.1723G>A, CADD score 20.8; *ZFHX3*, c.1601C>G, CADD score 22.3; *SLC5A2*, c.644T>C, CADD score 27.6; *CHD7*, c.4187C>G, CADD score 33; *NPR2*, c.952C>G, CADD score 22.2, *UBA3*, c.1088C>T, CADD score 27.8, *TANC2*, c.2357C>T, CADD score 22.7; *TRPS1*, c.1630C>T, CADD score 36) and five of the rare *de novo* variants (*HPS3*, c.1189C>T, CADD score 35; *MTA3*, c.393C>A, CADD score 26; *PLEC*, c.6704G>A, CADD score 26.5; *PIIP5K2*, c.686G>A, CADD score 34; *STAB1*, c.6145C>T, CADD score 24.1) reached CADD scores over 20 indicating that these variants are predicted to be among the 1% most deleterious variants in the human genome (written in bold in Table 2).

Re-sequencing of ZFHX3 in EA/TEF patients

Re-sequencing of *ZFHX3* in 192 EA/TEF patients did not identify additional putative disease-causing variants.

Structural modeling and in-silico analysis of ZFHX3 protein variants

Swiss model employed template id 3wbj.1.A as a template and built the *ZFHX3* amino acid change p.Pro534Arg. Structural models were obtained with sequence identity 14.89%, coverage of 58.75%, and normalized Z-score of -2.90. The respective values are considered as an indicative of correctly folded and good modeled structures close to native structure. For the amino acid change p.Ala2126Val, structural models were obtained with sequence identity of 19.15%, coverage 27.48%, and normalized Z-score of -1.76. From the structural modeling of the *ZFHX3* amino acid changes, we found that the two changes do not have any distortion in the native protein amounting to RMSD change at α -carbon is 0.02 Å and at backbone is 0.03 Å in p.Pro534Arg and RMSD change of at α -carbon is 0.05 Å and at backbone is 0.06 Å in p.Ala2126Val (S1–S4 Figs).

Transcriptome analysis

Evaluation of the transcriptome data showed that all murine genes were expressed at E8.5, E12.5, and postnatal except for some *APOL2* orthologous. Differential gene expression analysis revealed that four out of 24 genes were transcriptome-wide differentially expressed between the time points of E8.5 and E12.5 (*Chd7*: logFC 2.243, p.adj 1.85E-29; *Npr2*: logFC -2.268, p.adj 5.18E-04; *Trps1*: logFC -2.927, p.adj 1.17E-28; *Eef1d*: logFC 1.248, p.adj 6.41E-05), and two between the time points E12.5 and postnatal (*Apol7a*: logFC -9.273, p.adj 5.18E-07; *Plec*: logFC

Table 2. Classification of *de novo* variants using in silico prediction programs.

Ext-Code	Variant	HGNC	MutCDNA	gnomAD (MAF)	SIFT	LRT	Mutation Taster	Mutation Assessor	FATHMM	PROVEAN	Meta SVM	Meta LR	Fathmm MKL_coding	CADD Score
4_501	1	<i>EEF1D</i>	c.874C>T		-	N	A	-	-	-	-	-	N	28,5
	2	<i>CELSR1</i>	c.4357G>A	3/282,594 (0.00001)	T	U	N	N	T	N	T	T	D	15,3
21_501	3	<i>HPS3</i>	c.1189C>T	10/282,776 (0.00004)	D	D	A	M	T	D	D	T	D	35
27_501	4	<i>PIGC</i>	c.716C>T		T	D	D	M	T	N	T	T	D	11,6
35_501	5	<i>NFX1</i>	c.1723G>A		D	N	N	L	T	N	T	T	D	20,8
36_501	6	<i>ZFHX3</i>	c.1601C>G		D	N	D	L	T	N	T	T	D	22,3
41_501	7	<i>MTA3</i>	c.393C>A	1/237,600 (0.000004)	D	D	D	H	D	D	D	D	D	26
46_501	8	<i>FANCB</i>	c.782G>A		T	N	N	N	T	N	T	T	N	7,2
	9	<i>PLEC</i>	c.6704G>A	17/272,690 (0.00006)	D	U	D	N	T	N	T	T	D	26,5
63_501	10	<i>PPIP5K2</i>	c.686G>A	2/247,732 (0.000008)	D	D	D	M	T	D	T	T	D	34
88_501	11	<i>CLP1</i>	c.814C>A	1/251,486 (0.000003)	T	D	D	L	T	N	T	T	D	17,4
	12	<i>GPR133</i>	c.1033G>A	6/282,534 (0.00002)	T	N	N	N	T	N	T	T	N	0,016
	13	<i>SLC5A2</i>	c.644T>C		D	D	D	H	D	D	D	D	D	27,6
90_501	14	<i>KIAA0556</i>	c.3730C>T		D	N	N	L	T	N	T	T	N	1,9
141_501	15	<i>STAB1</i>	c.6145C>T	9/278,948 (0.00003)	T	N	N	M	T	D	T	T	N	24,1
154_501	16	<i>GGT6</i>	c.1045A>G		D	N	N	N	T	D	T	T	N	5,9
167_501	17	<i>CHD7</i>	c.4187C>G		D	D	D	H	T	D	D	D	D	33
172_501	18	<i>NPR2</i>	c.952C>G		T	N	D	L	D	D	T	T	D	22,2
174_501	19	<i>UBA3</i>	c.1088C>T		D	D	D	L	T	D	T	T	D	27,8
181_501	20	<i>TANC2</i>	c.2357C>T		T	D	D	L	T	D	T	T	D	22,7
288_501	21	<i>TRPS1</i>	c.1630C>T		-	D	A	-	-	-	-	-	D	36
	22	<i>APOL2</i>	c.319G>C		T	N	N	N	T	N	T	T	N	0,004
750_501*	23	<i>ZFHX3</i>	c.6377C>T	5/250,880 (0.00002)	D	D	D	L	T	D	D	D	D	19,2

*A: automatic disease causing; D: disease causing; H: high functional; L: non-functional; M: medium functional; N: neutral; T: tolerant. Annotations marked in bold red represent: "variants that are classified to be disease causing by at least eight out of ten *in silico* prediction programs (except for truncating variants) used by dbNSFP v3.0 (<https://sites.google.com/site/jpopgen/dbNSFP>)".

<https://doi.org/10.1371/journal.pone.0234246.t002>

-2.352, p.adj 9.09E-24). Interestingly, most of the candidate genes were highly expressed at each time point (Fig 1, Fig 2, S2 Table). The candidate genes *Zfhx3*, *Ppip5k2*, *Chd7* and *Eef1d* were even expressed above the 95th percentile at E8.5 compared to the expression of all other genes. In addition, the genes *Ppip5k2*, *Trps1*, *Zfhx3* and *Eef1d* were expressed above the 93rd percentile at E12.5.

Discussion

The etiology of EA/TEF is heterogeneous. Previously, disease causing monoallelic mutations of variable genomic size have been reported among EA/TEF patients [7; 25]. Here, we identified 23 single nucleotide *de novo* variants in 23 different genes using 30 unrelated case-parent trios and ES. All confirmed *de novo* variants were prioritized using *in silico* prediction tools.

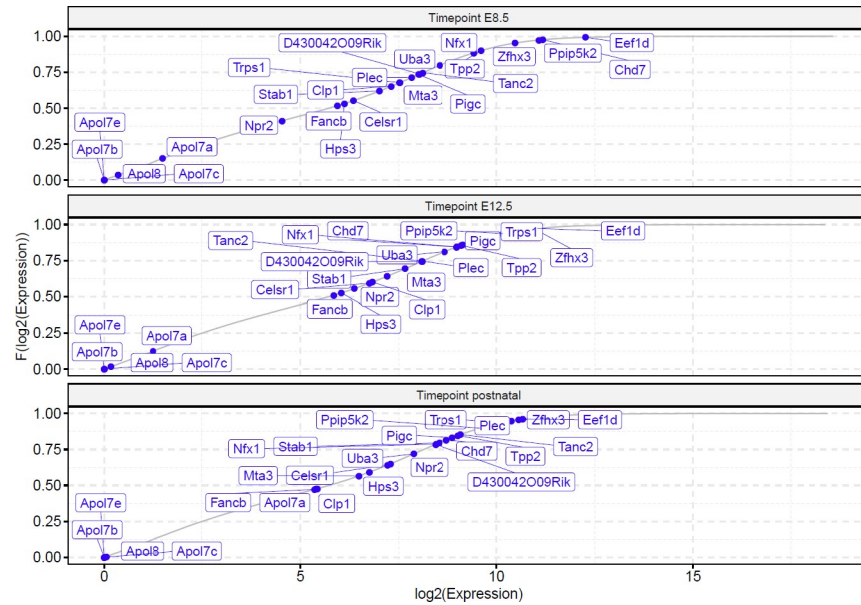


Fig 1. Empirical cumulative distribution of murine candidate gene expression at each timepoint. The empirical cumulative distribution function (F) was calculated from Regularized Log (*rlog*) transformed expression values. E: embryonic day.

<https://doi.org/10.1371/journal.pone.0234246.g001>

The embryonic role of genes harboring prioritized *de novo* variants was further investigated by targeted analysis of mouse transcriptome data of esophageal tissue obtained at E8.5, E12.5, and postnatal.

After prioritization of variants using *in silico* prediction tools, targeted analysis of mouse transcriptome data, and review of the literature we prioritize *TRPS1* and *ZFHX3* as new EA/

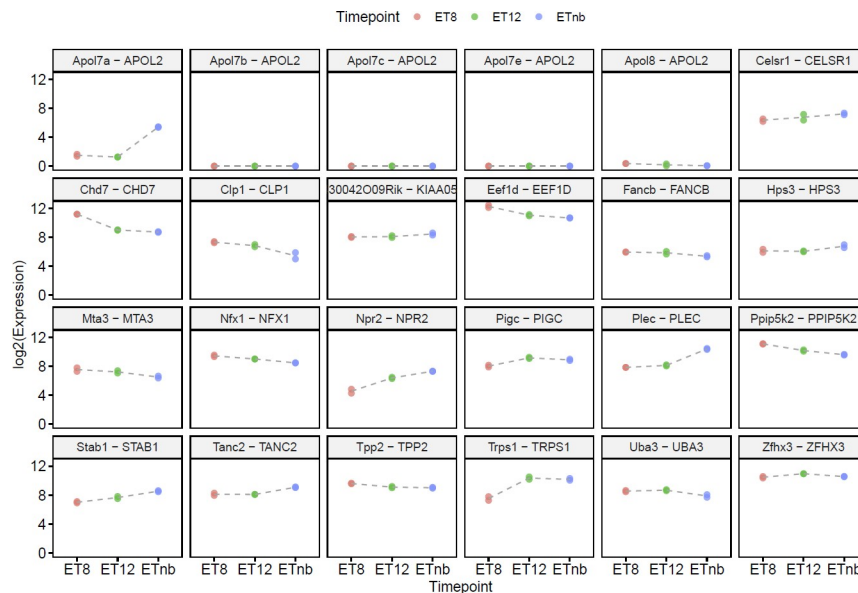


Fig 2. Murine candidate gene expression at three different timepoints. The gene expression is shown as log₂ expression on the y-axis, while the timepoint is shown as categorical variable on the x-axis. The header of each sub figure shows the murine and human gene symbols. E: embryonic day, pn: postnatal.

<https://doi.org/10.1371/journal.pone.0234246.g002>

TEF candidate genes and provide further support for *CHD7* as a key player in esophageal development.

The identified *de novo* amino acid change in *CHD7* has not been previously described. *CHD7* has been established as the major disease gene for CHARGE syndrome (OMIM #214800) [26]. Eight out of nine *in silico* prediction programs used by dbNSFP v3.0 (<https://sites.google.com/site/jpopgen/dbNSFP>) classified this *de novo* amino acid change p. Ala1396Gly in *CHD7* as deleterious. As about 20% of patients with CHARGE syndrome present with EA/TEF [27] we consider the identified variant as disease causing in our patient (167_501, Table 1) even though patient 167_501 did not present with additional congenital anomalies besides EA/TEF that would have suggested the clinical diagnosis of CHARGE syndrome. None of the other identified *de novo* variants resided within a gene that was previously linked to the formation of EA/TEF.

In patient 288_501 we identified a novel *de novo* truncating amino acid change p.Arg544* in *TRPS1* associated with tricho-rhino-phalangeal syndrome I (OMIM #190350). Previously Maas et al. (2015) reported the same truncating variant in three unrelated patients with *TRPS1* [25]. Unlike our patient 288_501, these previously reported patients did not present with any congenital anomaly of esophagus or trachea nor with any congenital anomaly of the heart (personal communication with Dr. Raoul C. Hennekam). Interestingly, in the here generated expression data *Trps1* shows a consistently high expression levels of 67th percentile at E8.5 and a log₂ Foldchange of -2.92 between days E8.5 and E12.5 in mouse embryos in the esophageal area, suggestive of an involvement of *Trps1* during vertebrate foregut development. The latter hypothesis suggests that the here identified *de novo* variant in *TRPS1* might be involved in the expression of EA/TEF in patient 288_501.

In patient 36_501 with nonsyndromic EA/TEF we identified a novel *de novo* variant in *ZFHX3*. Through personal communication during the project we identified another patient with EA/TEF as part of his VATER/VACTERL association (patient 750_501) with a *de novo* variant in *ZFHX3*. While the novel variant p.Pro534Arg resides in a well-conserved region of *ZFHX3* and has not been reported in gnomAD, the variant p.Ala2126Val has been reported five times heterozygous in 250,880 alleles in gnomAD (MAF 0.00002) (Table 1) and resides in a less well conserved region of *ZFHX3*. Prompted by this finding, we re-sequenced *ZFHX3* in 192 additional EA/TEF patients but did not find any further putative EA/TEF associated variant. In order to further analyze the two identified amino acid changes in *ZFHX3* we further performed structural modeling and *in-silico* analysis of *ZFHX3* protein. Here, substitution of C to G at position c.1601 has resulted in substitution of Pro to Arg at position 534 with a RMSD value amounting to 0.02Å at C-alpha carbon and 0.03Å in the protein backbone. Similarly, for the C to T substitution at c.6377 position that resulted into Ala to Val substitution at position 2126 has also recorded a similar change in RMSD value, 0.05Å at C-alpha carbon and 0.06Å in the protein backbone. While the structural modeling suggests that both changes do not cause distortion of the native protein, a possible functional impact of both variants would warrant further functional testing. According to our transcriptome analysis *Zfhx3* is not differentially expressed between either E8.5 and E12.5 or E12.5 and postnatal. However, *Zfhx3* is among the top expressed genes at E8.5 (>95th percentile) and E12.5 (>97th percentile). Interestingly, Thisse and Thisse (2004) reported also expression of *zfhx3* in zebrafish larvae 24 hours post fertilization in the region of the pharyngeal arches representing a series of paired bony or cartilaginous arches that develop along the lateral walls of the foregut, supporting the role of *ZFHX3* in vertebrate foregut development [28]. Taken together, the here detected *de novo* variants in human EA/TEF patients, the high *Zfhx3* expression at 8.5 and 12.5 in embryonic foregut tissue of mouse embryos and the previously reported expression of *zfhx3* in

zebrafish larvae in the region of the pharyngeal arches suggests *ZFHX3* as a putative EA/TEF candidate gene.

Overall, interpretation of the data is limited by the lack of animal models, at least for the findings in *CHD7*, *TRPS1*, and *ZFHX3*. To the best of our knowledge, there has no animal model been described that would have investigated embryonic foregut development, when these genes have been deleted. In order to definitely conclude that our findings respectively *de novo* variants in *CHD7*, *TRPS1*, and *ZFHX3* have been directly causative for the EA/TEF phenotype in the respective patients, *in vivo* experiments including animal models would be necessary, which were beyond the scope of our present study.

Conclusion

In summary, we detected 23 *de novo* mutations in 23 genes in 17 unrelated patients. Human exome and mouse embryonic expression analyses suggest *ZFHX3* and *TRPS1* as putative EA/TEF candidate genes and endorse *CHD7* as a key player for esophageal development.

Supporting information

S1 Fig. Wild and mutant structures of ZFHX3 (c.1601C>G).

(PNG)

S2 Fig. Super imposed structure of ZFHX3 Wild & mutant (c.1601C>G).

(PNG)

S3 Fig. Wild and mutant structures of ZFHX3 (c.6377C>T).

(PNG)

S4 Fig. Super imposed structure of ZFHX3 Wild & mutant (c.6377C>T).

(PNG)

S1 Table. Phenotypes of the patients included in the ES. VATER/VACTERL-like association (vertebral defects (V), anorectal malformations (A), cardiac defects (C), tracheoesophageal fistula with or without esophageal atresia (TE), renal malformations (R), and limb defects (L) [5], Ventricular septal defect (VSD), atrial septal defect (ASD).

(DOCX)

S2 Table. logFC: log2 of foldchange, baseMean: Average gene expression across all time-points, pvalue: nominal P-value, padj: Benjamini-Hochberg corrected P-value, ECDF percentile: Percentile of empirical cumulative distribution function for each timepoint.

(DOCX)

S1 Data.

(DOCX)

Acknowledgments

We thank all patients and their families for their participation, as well as the German self-help organizations for individuals with anorectal malformations (SoMA e.V.) and tracheoesophageal fistula with or without esophageal atresia (TE) (KEKS e.V.) for their assistance with recruitment. We also thank Prof. Raoul Hennekam for fruitful discussion on the manuscript.

Author Contributions

Conceptualization: Jan Gehlen, Tikam Chand Dakal, Alice Hölscher, Phillip Grote, Holger Thiele, Heiko Reutter.

Data curation: Jan Gehlen, Amit Kawalia, Alice Hölscher, Stefanie Heilmann-Heimbach, Nadine Zwink, Ekkehart Jenetzky, Phillip Grote, Holger Thiele, Heiko Reutter.

Formal analysis: Rong Zhang, Jan Gehlen, Amit Kawalia, Maria-Theodora Melissari, Tikam Chand Dakal, Athira M. Menon, Julia Höfele, Korbinian Riedhammer, Lea Waffenschmidt, Amit Sharma, Michael Ludwig, Holger Thiele, Heiko Reutter.

Funding acquisition: Julia Höfele, Ekkehart Jenetzky, Phillip Grote, Johannes Schumacher, Holger Thiele, Heiko Reutter.

Investigation: Rong Zhang, Maria-Theodora Melissari, Tikam Chand Dakal, Athira M. Menon, Lea Waffenschmidt, Julia Fabian, Katinka Breuer, Jeshurun Kalanithy, Alina Christine Hilger, Amit Sharma, Michael Ludwig.

Methodology: Jan Gehlen, Tikam Chand Dakal, Holger Thiele.

Project administration: Nadine Zwink, Ekkehart Jenetzky, Phillip Grote, Holger Thiele, Heiko Reutter.

Resources: Jeshurun Kalanithy, Alice Hölscher, Thomas M. Boemers, Markus Pauly, Andreas Leutner, Jörg Fuchs, Guido Seitz, Barbara M. Ludwikowski, Barbara Gomez, Jochen Hubertus, Andreas Heydweiller, Ralf Kurz, Johannes Leonhardt, Ferdinand Kosch, Stefan Holland-Cunz, Oliver Münsterer, Beno Ure, Eberhard Schmiedeke, Jörg Nesper, Petra Degenhardt, Stefanie Märzheuser, Katharina Kleine, Mattias Schäfer, Nicole Spsychalski, Oliver J. Deffaa, Jan-Hendrik Gosemann, Martin Lacher, Stefanie Heilmann-Heimbach, Nadine Zwink, Ekkehart Jenetzky, Phillip Grote, Johannes Schumacher, Holger Thiele, Heiko Reutter.

Software: Holger Thiele.

Supervision: Michael Ludwig, Phillip Grote, Holger Thiele, Heiko Reutter.

Validation: Rong Zhang, Phillip Grote.

Visualization: Jan Gehlen, Holger Thiele, Heiko Reutter.

Writing – original draft: Rong Zhang, Jan Gehlen, Tikam Chand Dakal, Julia Höfele, Amit Sharma, Michael Ludwig, Phillip Grote, Johannes Schumacher, Holger Thiele, Heiko Reutter.

Writing – review & editing: Rong Zhang, Jan Gehlen, Tikam Chand Dakal, Korbinian Riedhammer, Alina Christine Hilger, Amit Sharma, Markus Pauly, Andreas Leutner, Jörg Fuchs, Guido Seitz, Barbara M. Ludwikowski, Barbara Gomez, Jochen Hubertus, Nadine Zwink, Ekkehart Jenetzky, Phillip Grote, Johannes Schumacher, Heiko Reutter.

References

1. David T.J., and O'Callaghan S.E. (1975). Oesophageal atresia in the South West of England. *J. Med. Genet.* 12, 1–11. <https://doi.org/10.1136/jmg.12.1.1> PMID: 1121014
2. Depaepe A., Dolk H., and Lechat M.F. (1993). The epidemiology of tracheo-oesophageal fistula and oesophageal atresia in Europe. EUROCAT Working Group. *Arch. Dis. Child.* 68, 743–748. <https://doi.org/10.1136/adc.68.6.743> PMID: 8333763

3. Pedersen R.N., Calzolari E., Husby S., Garne E., and EUROCAT Working group (2012). Oesophageal atresia: prevalence, prenatal diagnosis and associated anomalies in 23 European regions. *Arch. Dis. Child.* 97, 227–232. <https://doi.org/10.1136/archdischild-2011-300597> PMID: 22247246
4. Torfs C.P., Curry C.J., and Bateson T.F. (1995). Population-based study of tracheoesophageal fistula and esophageal atresia. *Teratology* 52, 220–232. <https://doi.org/10.1002/tera.1420520408> PMID: 8838292
5. Quan L., and Smith D.W. (1973). The VATER association. Vertebral defects, Anal atresia, T-E fistula with esophageal atresia, Radial and Renal dysplasia: a spectrum of associated defects. *J. Pediatr.* 82, 104–107. [https://doi.org/10.1016/s0022-3476\(73\)80024-1](https://doi.org/10.1016/s0022-3476(73)80024-1) PMID: 4681850
6. 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. <https://doi.org/10.1111/j.1399-0004.2007.00798.x> PMID: 17489843
7. Brosens E., Marsch F., de Jong E.M., Zaveri H.P., Hilger A.C., Choinitzki V.G., et al. (2016). Copy number variations in 375 patients with oesophageal atresia and/or tracheoesophageal fistula. *Eur. J. Hum. Genet. EJHG* 24, 1715–1723. <https://doi.org/10.1038/ejhg.2016.86> PMID: 27436264
8. van Bokhoven H., Celli J., van Reeuwijk J., Rinne T., Glaudemans B., van Beusekom E., et al. (2005). MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome. *Nat. Genet.* 37, 465–467. <https://doi.org/10.1038/ng1546> PMID: 15821734
9. Jongmans M.C.J., Admiraal R.J., van der Donk K.P., Vissers L.E.L.M., Baas A.F., Kapusta L., et al. (2006). CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. *J. Med. Genet.* 43, 306–314. <https://doi.org/10.1136/jmg.2005.036061> PMID: 16155193
10. Motoyama J., Liu J., Mo R., Ding Q., Post M., and Hui C. (1998). Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat. Genet.* 20, 54–57. <https://doi.org/10.1038/1711> PMID: 9731531
11. Williamson K.A., Hever A.M., Rainger J., Rogers R.C., Magee A., Fiedler Z., et al. (2006). Mutations in SOX2 cause anophthalmia-esophageal-genital (AEG) syndrome. *Hum. Mol. Genet.* 15, 1413–1422. <https://doi.org/10.1093/hmg/ddl064> PMID: 16543359
12. Zhang R., Marsch F., Kause F., Degenhardt F., Schmiedeke E., Märzheuser S., et al. (2017). Array-based molecular karyotyping in 115 VATER/VACTERL and VATER/VACTERL-like patients identifies disease-causing copy number variations. *Birth Defects Res.* 109, 1063–1069. <https://doi.org/10.1002/bdr2.1042> PMID: 28605140
13. Li H., and Durbin R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
14. McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytzky A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
15. Kawalia A., Motameny S., Wonzak S., Thiele H., Nieroda L., Jabbari K., et al. (2015). Leveraging the Power of High Performance Computing for Next Generation Sequencing Data Analysis: Tricks and Twists from a High Throughput Exome Workflow. *PLOS ONE* 10, e0126321. <https://doi.org/10.1371/journal.pone.0126321> PMID: 25942438
16. Yeo G., and Burge C.B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J. Comput. Biol.* 11, 377–394. <https://doi.org/10.1089/1066527041410418> PMID: 15285897
17. Ramu A., Noordam M.J., Schwartz R.S., Wuster A., Hurles M.E., Cartwright R.A., et al (2013). DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat. Methods* 10, 985–987. <https://doi.org/10.1038/nmeth.2611> PMID: 23975140
18. Liu X., Jian X., and Boerwinkle E. (2011). dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* 32, 894–899. <https://doi.org/10.1002/humu.21517> PMID: 21520341
19. Liu X., Wu C., Li C., and Boerwinkle E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* 37, 235–241. <https://doi.org/10.1002/humu.22932> PMID: 26555599
20. Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., et al (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* 29, 15–21.
21. Ewels P., Magnusson M., Lundin S., and Käller M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinforma. Oxf. Engl.* 32, 3047–3048.
22. Love M.I., Huber W., and Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281

23. Love M.I., Anders S., Kim V., and Huber W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research* 4, 1070. <https://doi.org/10.12688/f1000research.7035.1> PMID: 26674615
24. 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. <https://doi.org/10.1038/nprot.2009.97> PMID: 19617889
25. Maas S.M., Shaw A.C., Bikker H., Lüdecke H.-J., van der Tuin K., Badura-Stronka M., et al. (2015). Phenotype and genotype in 103 patients with tricho-rhino-phalangeal syndrome. *Eur. J. Med. Genet.* 58, 279–292.
26. Janssen N., Bergman J.E.H., Swertz M.A., Tranebjaerg L., Lodahl M., Schoots J., et al. (2012). Mutation update on the CHD7 gene involved in CHARGE syndrome. *Hum. Mutat.* 33, 1149–1160. <https://doi.org/10.1002/humu.22086> PMID: 22461308
27. Kutiyawala M., Wyse R.K., Brereton R.J., Spitz L., Kiely E.M., Drake D., et al. (1992). CHARGE and esophageal atresia. *J. Pediatr. Surg.* 27, 558–560. [https://doi.org/10.1016/0022-3468\(92\)90445-d](https://doi.org/10.1016/0022-3468(92)90445-d) PMID: 1625119
28. Thisse B., and Thisse C. (2004). Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submiss.