Klar et al. BMC Medical Genomics (2020) 13:85
https://doi.org/10.1186/s12920-020-00737-6

BMC Medical Genomics

RESEARCH ARTICLE Open Access

# Whole genome sequencing of familial isolated oesophagus atresia uncover shared structural variants

Joakim Klar[1,2*], Helene Engstrand-Lilja[1,2], Khurram Maqbool[1,2], Jonas Mattisson[1,2], Lars Feuk[1,2] and Niklas Dahl[1,2]

## Abstract

**Background:** Oesophageal atresia (OA) is a life-threatening developmental defect characterized by a lost continuity between the upper and lower oesophagus. The most common form is a distal connection between the trachea and the oesophagus, i.e. a tracheoesophageal fistula (TEF). The condition may be part of a syndrome or occurs as an isolated feature. The recurrence risk in affected families is increased compared to the population-based incidence suggesting contributing genetic factors.

**Methods:** To gain insight into gene variants and genes associated with isolated OA we conducted whole genome sequencing on samples from three families with recurrent cases affected by congenital and isolated TEF.

**Results:** We identified a combination of single nucleotide variants (SNVs), splice site variants (SSV) and structural variants (SV) annotated to altogether 100 coding genes in the six affected individuals.

**Conclusion:** This study highlights rare SVs among candidate gene variants in our individuals with OA and provides a gene framework for further investigations of genetic factors behind this malformation.

**Keywords:** Oesophagus atresia, Whole genome sequencing

## Background

Oesophageal atresia (OA) is the most common congenital anomaly of the oesophagus with an incidence of around 1 in 3500 births [1, 2]. The malformation is characterized by a distal tracheoesophageal fistula, classified as Gross type C, in 85% of all cases [3]. Isolated OA occurs in approximately 50% of cases whereas the remaining are syndromic [4–7]. Although the genetic basis for syndromic OA has been identified in a proportion of syndromic cases, the genetics behind isolated forms remains elusive. The recurrence risk for isolated OA is estimated to approximately 1% and twin studies have shown a concordance rate of 2.5% [8–10]. The few familial cases of isolated OA that are reported suggest an autosomal dominant inheritance with reduced penetrance [8, 11, 12] whereas epidemiological studies on isolated OA indicate a multifactorial aetiology, with a contribution from both gene variants and environmental factors [4]. A prior effort to unravel genetic factors behind isolated OA using SNP microarray analysis identified two distinct de novo CNVs in a cohort consisting of 129 cases [13]. When including syndromic OA, no consistent genomic region was identified. While the study provided important information, it is still unclear whether structural variants (SVs), small insertions/deletions (indels) and single nucleotide variants (SNVs), below detection limit of the DNA microarray used, are associated with the malformation.

Genetic variation in humans can be everything from rare to common, where variants usually associated with Mendelian traits tend to be rare whereas more frequent

* Correspondence: joakim.klar@igp.uu.se
[1]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala, Sweden
[2]Department of Women's and Children's Health, Section of Pediatric Surgery, Uppsala University, SE-75185 Uppsala, Sweden

allele variants contribute risk for complex disease phenotypes [14]. The genetic architecture of both rare Mendelian diseases and multifactorial disorders such as birth defects may be explained by a continuum based primarily on the frequencies of the relevant variant alleles. Furthermore, incomplete penetrance for one or several common variants has inferred a shift beyond the classical concept of Mendelian inheritance [15, 16]. Complex traits can be seen as omnigenic and thereby driven by large numbers of variants of small effects and, in this context, we also need to identify modifier loci that contribute to penetrance of variation [17]. We therefore set out to analyse the genomic sequences of three families with recurrent cases of a low tracheoesophageal fistula (TEF), the most common form of congenital OA. In order to decipher gene variants in individuals with the malformation we performed whole genome sequencing of affected members and their parents. Bioinformatic analysis using a set of variant calling algorithms revealed a large number of candidate SNVs, indels, and SVs shared by affected and obligate carriers. Our data provide a selected set of candidate genetic variants to further identify, prioritize and test for genetic mechanisms associated with OA.

## Methods

### Patients

We identified three families with recurrent and isolated OA treated in our tertiary paediatric surgical centre between 1994 and 2014 among a cohort of in total 55 cases with isolated OA Gross type C. Each of the three families comprised two individuals born with isolated OA. All the patients had an isolated OA with a lower tracheoesophageal fistula (Gross type C). None of them had other malformations or dysmorphic features, including VACTERL association or CHARGE syndrome (Fig. 1a).

In Family 1, the mother and her son were born with OA. In Family 2, two affected girls were second cousins and their fathers, respectively, were first cousins. In Family 3, a girl and a boy with OA were second cousins. The father of the girl and the mother of the boy were first cousins (Fig. 1b).

### Genetic analysis

Whole genome sequencing (WGS) was performed on peripheral blood DNA from the affected individuals ($n$ = 6) and the presumed obligate and healthy carriers ($n$ = 4). Sequencing libraries were prepared using the TruSeq PCRfree DNA sample preparation kit and sequenced using HiSeqX and paired-end 150 bp read length, v2.5 sequencing chemistry (Illumina). Quality control measurements are gathered using Qualimap v2.2. For data analysis we used Piper, a pipeline system developed and maintained at the National Genomics Infrastructure build on top of GATK Queue. For more information and the source code visit: www.github.com/National-GenomicsInfrastructure/piper. Data has been aligned to the GRCh37.75 reference using BWA-MEM v0.7.12 [18] and we identified single nucleotide variants (SNVs) and indels using the Genome Analysis Toolkit (GATK v3.3). For deduplication PicardMarkDuplicates, available in Picard that is bundled with GATK (broadinstitute.github.io/picard). For the analysis of structural variants (SVs), we used Manta, Delly, CNVnator and Tiddit as described previously [19–23]. Manta and Delly calls SVs using combination of paired and split-reads and report breakpoints for SVs with strong evidence from split-reads whereas CNVnator uses read depth technique. TIDDIT is mainly designed to identify larger SVs (> 1 kb) using coverage and insert size distribution to identify SVs based on discordant read pairs. We collapsed SVs who had regions where both starts and ends were within 10 bp, as they likely represent the same called SV. All variants were annotated using ENSEMBL Variant Effect Predictor (VEP) [24]. Variants that potentially affect

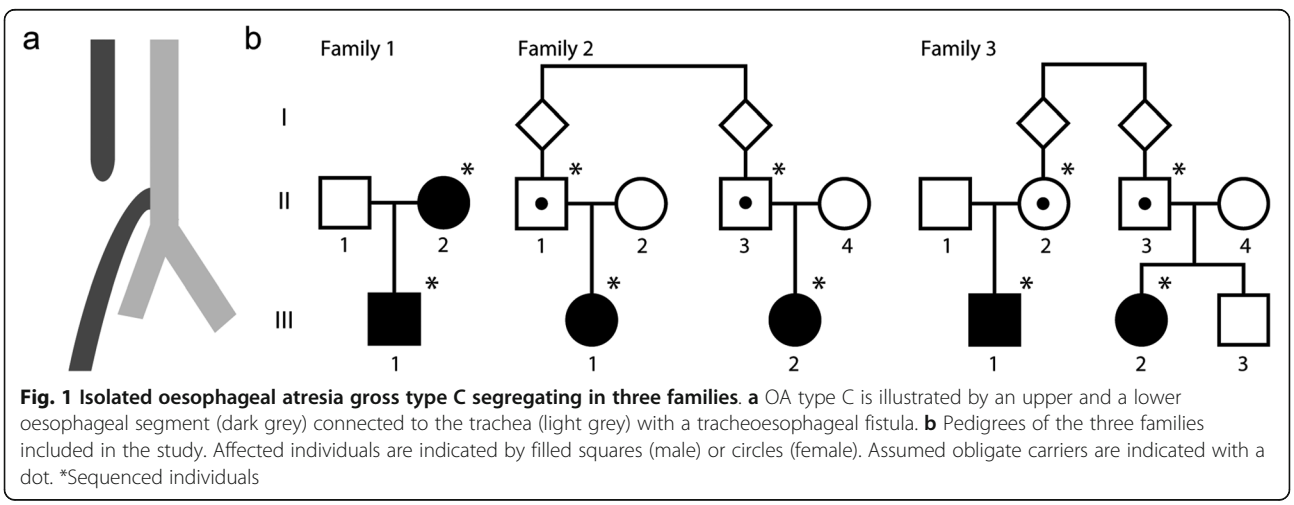


**Fig. 1 Isolated oesophageal atresia gross type C segregating in three families**. **a** OA type C is illustrated by an upper and a lower oesophageal segment (dark grey) connected to the trachea (light grey) with a tracheoesophageal fistula. **b** Pedigrees of the three families included in the study. Affected individuals are indicated by filled squares (male) or circles (female). Assumed obligate carriers are indicated with a dot. *Sequenced individuals

splicing were identified using SpliceAI (Score cut off 0.2 (high recall/likely pathogenic), 0.5 (recommended/pathogenic), and 0.8 (high precision/pathogenic)) [25]. We filtered SNVs and indels for frequency and sequencing errors using the Moon (www.diploid.com) software and against GnomAD [26]. The SVs were filtered against 1000 Genomes and the SweGen dataset using the aforementioned SV callers [23]. The SweGen dataset is the largest available cross-section of variation detected by the same WGS pipeline as we used in the Swedish population and can be seen as a population matched control dataset for our families. We annotated identified SVs with allele frequencies from 1000G, GnomAD and the SweGen dataset for all four SV callers.

Tissue expression of genes was investigated using the Genotype-Tissue Expression (GTEx; gtexportal.org) Project Portal of 05/08/19. We did pathway analysis using EnrichR (amp.pharm.mssm.edu/Enrichr). We investigated associations between candidate genes identified in our study and oesophageal disease using GeneDistiller [27]. A candidate gene list was generated in GeneDistiller by adding OMIM entries using the keyword 'esophageal' and manually adding genes previously associated with OA or similar phenotypes (complete reference gene list: *ALDH2, C2orf40, CHD7, COL4A6, CTAG1B, DEC1, DLEC1, EFTUD2, FANCB, FGF8, FOXC2, FOXF1, FOXL1, GAEC1, GER, GLI3, HOXD13, HSN1B, LZTS1, MTHFSD, MYCN, PTEN, RFX6, SOX2, SPG9, TMPRSS11A,* and *ZIC3*). Associations identified by GeneDistiller included Gene Reference into Function (GeneRIF) and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; string-db.org).

## Results

### Whole genome sequencing

Quality control measurements show that the mean coverage for the 10 samples was 32.6X (min 21.9X; max 39.9X) with a median insert size of 394 (min 371; max 419). The total number of reads was 1,057,991,431 (min 794,472,217; max 1,279,622,464) per sample and the number of aligned reads was 1,052,483,855 (min 791,300,824; max 1,270,865,080) corresponding to 99.5% aligned reads (min 99.3%; max 99.7%). The duplication rate was 35% (Standard deviation = 5.4%; min 28%; max 42%). We excluded sample bias by looking at the GC-content distribution compared to a pre-calculated distribution for the reference genome (hg19; Additional file 1).

### Single nucleotide variants (SNVs) and small insertions/deletions (indels)

After an initial filtering for uncommon variants (< 2% frequency in GnomAD) shared by the affected members of each family, we identified in total 345,156 (6604 coding) variants in Family 1, 199,780 (3450 coding) variants in Family 2 and 207,691 (3549 coding) SNVs variants in Family 3. We also performed a combined analysis using all three families and identified 109,712 (2098 coding) SNVs and indels shared by all six affected individuals. Since no shared homozygous, or compound heterozygous, variants were identified with an allele frequency less than 0.001 in GnomAD, we considered dominant acting variants and filtered for rare coding heterozygous variants present in all sequenced individuals. This approach revealed in total 47 heterozygous variants in coding sequences and splice sites. However, when filtering against the Moon database, which includes sequencing errors, all 47 variants showed an allele frequency of > 0.029. These high frequencies suggest that the variants are either benign or common sequencing errors. We therefore performed a re-analysis of each family separately. By applying all abovementioned filtering steps, we identified 59 variants in Family 1, three variants in Family 2 (in *KATNB1, ZMYND15* and *DHX33*) and three variants in Family 3 (in *BCAP29*, *DOCK4* and *PPP1R3A*) (Fig. 2a). All identified variants were heterozygous compatible with a dominant model (Additional file 2). We further analysed our data using SpliceAI in a similar fashion and identified 40 heterozygous splice-site variants in Family 1, 5 variants in Family 2 and 6 variants in Family 3. Four variants were shared by all 10 sequenced individuals (Fig. 2a). The variants have a predicted effect on splicing of the *PRIM2*, *FAM182B*, *MAP2K3* and *CCDC144NL* genes (Additional file 3). The shared variants identified in Family 2 and Family 3, respectively, were present also in the obligate carrier parents of affected cases. No variants identified was inherited from the non-sequenced parent, indicating co-segregation from the obligate carrier parents in family 2 and 3 and from the affected parent in family 1.

### Structural variants (SVs)

We then analysed our sequencing data for SVs. In total, we identified 5777 deletions (DEL), 1480 duplications (DUP) and 104 inversions (INV) that were detected in at least one sample. Filtering for SVs shared by all six affected individuals was performed as for SNVs and small indels (AF < 0.003 in SweGen which equals at most two individuals carrying the SV) resulting in a total of 52 SVs, i.e. 43 DEL and 9 DUP (Additional file 4). Analysis of SV shared by the two affected individuals in each separate family resulted in the identification of 339 DEL, 47 DUP and 3 INV in Family 1, 172 DEL, 19 DUP and 2 INV in Family 2 and 135 DEL, 17 DUP and 1 INV in Family 3 (Fig. 2a; Additional files 5, 6 and 7). All identified variants were found also in the parents of each family confirming segregation.

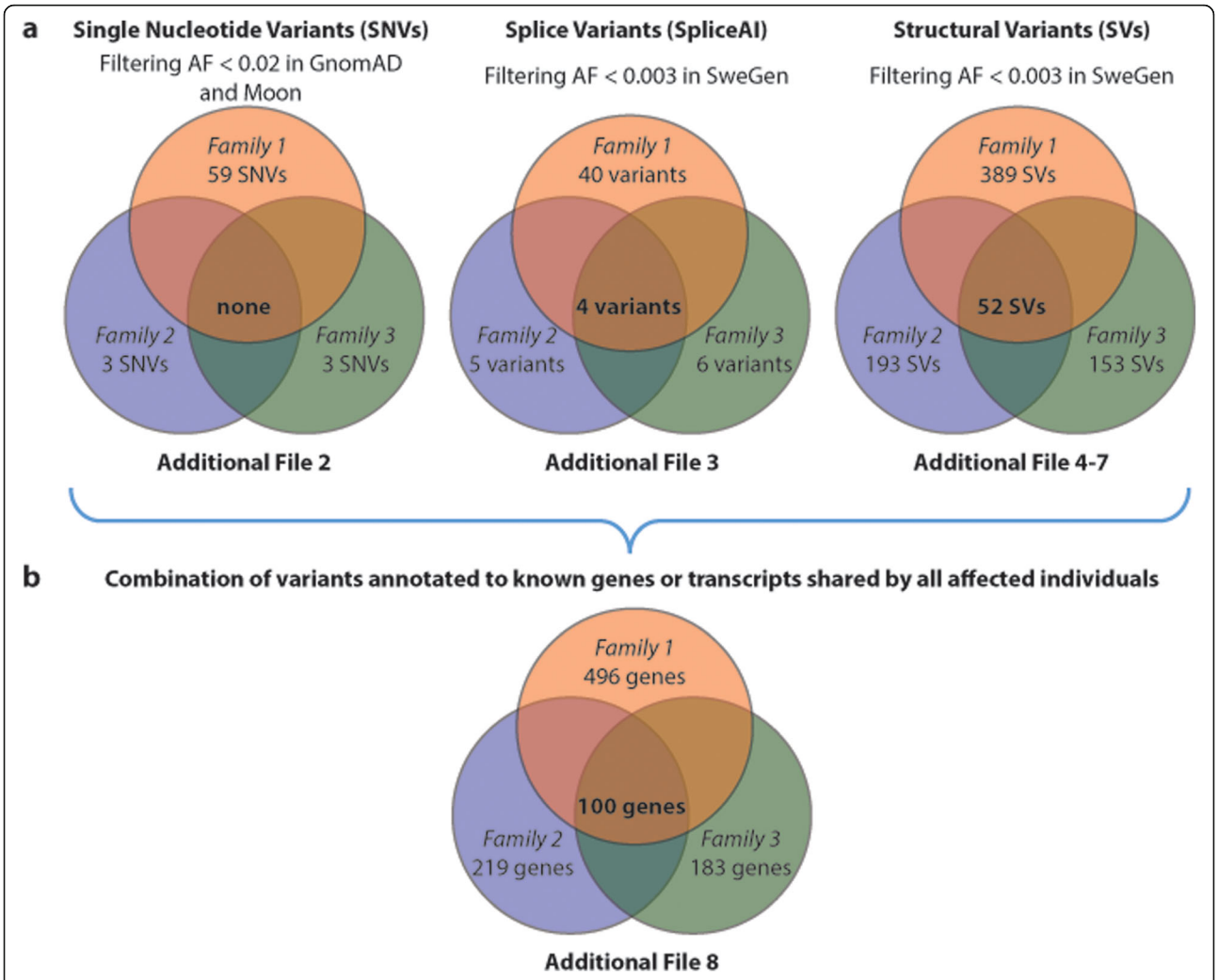


**Fig. 2 Overview of variant calling in WGS data. a** Number and type of gene variants called in each of the three families segregating OA with number of shared variants. For SNVs and small indels, we identified 59 heterozygous variants in Family 1, 3 heterozygous variants in Family 2 and 3 heterozygous variants in Family 3 compatible with a dominant model. None of these variants were shared between all three families. For splice variants identified using SpliceAI, we identified 40 heterozygous variants in Family 1, 5 variants in Family 2 and 6 variants in Family 3. Four splice variants were shared by all three families. For structural variants we identified 389 SVs (339 DEL, 47 DUP and 3 INV) in Family 1, 193 SVs (172 DEL, 19 DUP and 2 INV) in Family 2 and 153 SVs (135 DEL, 17 DUP and 1 INV) in Family 3. In total, 52 SVs (43 DEL and 9 DUP) were shared by all three families. **b** Number of genes with annotated variants in each family and the number of genes shared by all affected individuals. When combining our data sets (**a**) we identified rare SNVs, splice variants or SVs that were annotated to in total 100 different genes in all three families. AF; allele frequency

### Combination of variants in genes shared by all affected individuals

When combining our data sets we identified rare SNVs, splice variants or SVs that were annotated to in total 100 genes in all three families (Fig. 2b). The major proportion of variants consisted of SVs (involving 95 genes) followed by splicing (involving 4 genes). A single gene (*DOCK4*) showed a SNV (NM_014705.3:c.717G > C) in Family 3 in combination with a splice variant (acceptor gain) in Family 1 and a SV (intronic deletion of 33 bp) in Family 2 (Additional file 8). The 100 genes were then subject analysis by GeneDistiller to identify genes associated to the oesophagus. Two genes, *DEFB4A* and *PDS5B*, are associated (GeneRIF) with Esophageal squamous cell carcinoma (ESCC). A third gene, *PRKCZ*, is indirectly associated to the same disease in that it interacts (STRING) with the ESCC susceptibility gene *LZTS1*. The 100 shared genes are apparently diverse in function as we could not identify any significantly enriched category using EnrichR. Similarly, if we consider all genes from each family regardless whether they are shared (595 genes) we similarly did not find any significantly enriched category.

## Discussion

Epidemiological studies indicate that the causes of isolated OA are multifactorial similar to many other birth defects. We identified three families segregating isolated OA and performed whole genome sequencing on samples from family members in search for candidate gene variants with a relatively high penetrance for the malformation. The clinic was very similar between all three families described as OA Gross type C. However, all families was genetically analysed separately. Due to the fact that familial cases of OA is rare and that the patients came from the same population (Swedish), we also considered shared aetiology. A possible pattern of inheritance was autosomal dominant, with reduced penetrance in two families and, retrospectively, only genes with heterozygous variants were shared between individuals in each family. We hypothesized that shared rare variants in the two extended families 2 and 3, each one with affected second cousins, would bring stronger support for associated gene variants given the number of meiosis separating the cases. Due to the large number of variants identified in each of the small families (> 400 per family), we decided to focus the discussion on variants in genes shared by all three families. Bioinformatic analysis of WGS data disclosed 100 genes affected by a combination of SNVs, splice site variants and SVs in all six affected individuals (Fig. 2b). Our initial analysis of rare protein coding SNVs or small indel (filtered for an AF < 0.001) did not show any shared variant among all six cases. However, further analysis using SpliceAI revealed four heterozygous variants, none of them present in GnomAD, in the sequenced family members. The variants predict perturbed splicing of the *PRIM2*, *FAM182B*, *MAP2K3* and *CCDC144NL* genes, respectively. Notably, variants in *PRIM2* have been associated with a specific microbiome community of the oesophagus and *CCDC144NL* show differential expression in drug resistant oesophageal carcinoma cells (ESCC) [28, 29]. No connection could be found for the genes *FAM182B* and *MAP2K3* to oesophageal disease, although *MAP2K3* show a high level of expression in oesophageal tissue.

Importantly, the largest group of called variants that were shared among the six individuals with isolated OA and the four presumed carrier parents consists of SVs. Identification of SVs from short-read sequencing data is challenging as it yields a large amount of both false positive and false negative results. This is mainly due to repetitive elements that frequently result in ambiguities when trying to assign short read sequences to reference sequences. To this end, we combined four different callers, with their advantages and disadvantages, for the identification of SVs. This resulted in the identification of in total 52 shared SVs in all three families. Variants were identified in two gene regions with association to the oesophagus, namely *DEFB4A* and *PDS5B* previously associated with ESCC [30, 31] (Additional file 4). The *DEFB4A* gene (previously called *HBD2*) has an almost oesophagus specific expression and promotes both growth and invasion of oesophageal cancer [30, 32]. Moreover, allelic loss and altered expression of the *PDS5B* gene has been associated in ESCC [31]. A third gene, *PRKCZ* has a weak association to the oesophagus. This gene encodes for Protein kinase C zeta, a protein that interacts with Leucine zipper tumour suppressor 1 encoded by the gene *LZTS1* [33]. Somatic variants in *LZTS1* have been associated primary oesophageal cancer [34]. Interestingly, young adults with OA have a 100-fold increased prevalence of ESCC compared to the general population and a shared mechanism between the two disorders is plausible [35]. Cancer is usually associated with loss of function variants, while in our cases non-coding SVs may alter expression of genes that are required for proper specification and elongation of the oesophagus during embryogenesis, ultimately resulting in OA [36]. However, the high incidence of gastroesophageal reflux disease (GERD) in patients after OA repair may be an additional contributing factor to ESCC [37]. Recessive variants in the gene *TRAP1* have previously been associated with VACTERL, therefore it is possible that the heterozygous variant identified in this gene in Family 1 is a possible modifier of disease [38].

Our study adds to the few families reported with recurrent isolated OA and suggests contributing genetic factors behind the malformation. Furthermore, we applied a combination of bioinformatic tools on WGS data from affected families highlighting a large number of rare variants, primarily SVs, in individuals with the malformation in our cohort. Genome wide association studies (GWAS) of other non-syndromic congenital malformations such as cleft-lip-palate (CLP) and congenital heart disease have reported associations to numerous SNVs with significant effect sizes [39]. These risk variants occur frequently in the population and each one confers a modest effect size. In CLP, a GWAS study suggested a genomic region that is also vulnerable to rare variants [40]. However, risk loci identified by GWAS account for only a small fraction of the heritability of congenital malformations [41] and this is likely also for OA. While previous GWAS studies of different types of congenital malformations have provided important information on contributing genetic factors, they have not clarified the potential role of specific SVs in the aetiology of isolated birth defects and for some of the missing heritability [39]. Our study shows that the identification of rare SVs, whether in coding or non-coding regions, is possible using whole genome sequencing and novel bioinformatic pipe-lines.

## Conclusion

In summary, we explored the possible association between familial forms of isolated OA and rare gene variants using whole genome sequencing. In two out of the three families, the affected members were second cousins, whereas one family consisted of an affected mother and son. Among all variants (SNVs, splice sites and SVs) in 100 genes shared by the six individuals from the three families, we identified variants in three genes associated with oesophageal disease, namely *CCDC144NL, DEFB4A* and *PDS5B*, bringing further support for a molecular link between OA and oesophageal cancer. Furthermore, our study provides a data set for further computational and functional analyses of both isolated and syndromic OA. We anticipate that the identification of additional families segregating isolated OA, in combination with our findings, will facilitate the identification of specific gene variants contributing to this serious malformation.

## Supplementary information

**Supplementary information** accompanies this paper at https://doi.org/10.1186/s12920-020-00737-6.

**Additional file 1 Quality control of GC-content distribution.** The distribution of GC content of mapped reads for the samples (orange) indicate expected distribution compared to a pre-calculated GC distribution for the reference genome (hg19; blue). The bars indicate standard deviation (SD) for the samples ($n = 10$).

**Additional file 2 Variants identified using a dominant model.** All rare variants (< 2% frequency in GnomAD) identified following a dominant segregation pattern in the families.

**Additional file 3 SpliceAI variants identified using a dominant model.** All rare predicted splice variants (< 2% frequency in GnomAD) identified following a dominant segregation pattern in the families.

**Additional file 4 SVs identified.** All rare SVs (AF < 0.003 in SweGen) shared by the affected individuals in the families.

**Additional file 5 SVs identified in Family 1.** All rare SVs (AF < 0.003 in SweGen) shared by the affected individuals in Family 1.

**Additional file 6 SVs identified in Family 2.** All rare SVs (AF < 0.003 in SweGen) shared by the affected individuals in Family 2.

**Additional file 7 SVs identified in Family 3.** All rare SVs (AF < 0.003 in SweGen) shared by the affected individuals in Family 3.

**Additional file 8 Summary of the type of variants in shared genes between all families.** Rare SNVs, splice variants or SVs that were annotated to in total 100 genes in all three families.

### Abbreviations

OA: Oesophageal atresia; TEF: Tracheoesophageal fistula; SNVs: Single nucleotide variants; SSV: Splice site variants; SV: Structural variants; indels: Small insertions/deletions; WGS: Whole genome sequencing; DEL: Deletions; DUP: Duplications; INV: Inversions; ESCC: Esophageal squamous cell carcinoma; STRING: Search Tool for the Retrieval of Interacting Genes; gnomAD: The Genome Aggregation Database; GERD: Gastroesophageal reflux disease; GWAS: Genome wide association studies; CLP: Cleft-lip-palate

### Acknowledgements

We are grateful to the patients and their families for their cooperation. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

### Authors' contributions

HE-L, LF and ND designed the study. HE-L and ND interpreted the patient data regarding the disease. JK wrote the manuscript with the assistance of ND. JK, KM and JM performed the different genetic analyses. All authors read and approved the final manuscript.

### Funding

This study was supported by the Swedish Research Council 2015–02424, Science for Life Laboratory at Uppsala University, and Uppsala University Hospital. The funders had no role in the design of the study nor in the collection, analysis, interpretation of data and writing of the manuscript. Open access funding provided by Uppsala University.

### Availability of data and materials

The datasets during and/or analysed during the current study available from the corresponding author on reasonable request. The datasets generated during the current study submitted to The European Genome-phenome Archive (EGA; ID EGAS00001004394). GRCh37.75 reference is available from ENSEMBL (ftp://ftp.ensembl.org/pub/release-75).
The SweGen Variant Frequency Dataset is available to the scientific community through the website swefreq.nbis.se and the National Bioinformatics Infrastructure Sweden (NBIS) DOI repository (doi:https://doi.org/10.17044/NBIS/G000003) upon registration and agreement to terms and conditions for data download.
Allele frequencies from GnomAD and 1000G was retrieved using the ENSEMBL Variant Effect Predictor (VEP). The GnomAD data are available from gnomad.broadinstitute.org (direct link: SNV frequency file from storage.googleapis.com/gnomad-public/release/2.1.1/vcf/genomes/gnomad.genomes.r2.1.1.sites.vcf.bgz and SV frequency file from storage.googleapis.com/gnomad-public/papers/2019-sv/gnomad_v2.1_sv.sites.vcf.gz). The 1000G data is available from internationalgenome.org/data (direct link: frequency file from ftp://ftp.ensembl.org/pub/grch37/current/variation/gvf/homo_sapiens/1000GENOMES-phase_3.gvf.gz).

### Ethics approval and consent to participate

The study was approved by the Regional Ethics Review Board of Uppsala (Dnr 2004:M-270). Written informed consent was obtained from the single adult patient and from all parents to the affected children included in the study.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2020 Accepted: 8 June 2020
Published online: 26 June 2020

### References

1. Pedersen RN, Calzolari E, Husby S, Garne E. Oesophageal atresia: prevalence, prenatal diagnosis and associated anomalies in 23 European regions. Arch Dis Child. 2012;97(3):227–32.
2. Oddsberg J, Lu Y, Lagergren J. Aspects of esophageal atresia in a population-based setting: incidence, mortality, and cancer risk. Pediatr Surg Int. 2012;28(3):249–57.
3. Gross RE: The surgery of infancy and childhood. Its principles and techniques. By Robert E. Gross, M.D., D.Sc., and William E. Ladd, Professor children's surgery, The Harvard Medical School etc. 7 × 10 in. Pp. 1000 + xxiv, with 567 illustrations, the drawings by Etha Piotti. 1953. Philadelphia and London: W. B. Saunders Co. 80s. BJS 1953, 41(165):112–112.
4. Shaw-Smith C. Oesophageal atresia, tracheo-oesophageal fistula, and the VACTERL association: review of genetics and epidemiology. J Med Genet. 2006;43(7):545–54.
5. Felix JF, Tibboel D, de Klein A. Chromosomal anomalies in the aetiology of oesophageal atresia and tracheo-oesophageal fistula. Eur J Med Genet. 2007;50(3):163–75.

6. Brosens E, de Jong EM, Barakat TS, Eussen BH, D'Haene B, De Baere E, Verdin H, Poddighe PJ, Galjaard RJ, Gribnau J, et al. Structural and numerical changes of chromosome X in patients with esophageal atresia. Eur J Hum Genet. 2014;22(9):1077–84.
7. Genevieve D, de Pontual L, Amiel J, Sarnacki S, Lyonnet S. An overview of isolated and syndromic oesophageal atresia. Clin Genet. 2007;71(5):392–9.
8. Van Staey M, De Bie S, Matton MT, De Roose J. Familial congenital esophageal atresia. Personal case report and review of the literature. Hum Genet. 1984;66(2–3):260–6.
9. Robert E, Mutchinick O, Mastroiacovo P, Knudsen LB, Daltveit AK, Castilla EE, Lancaster P, Kallen B, Cocchi G. An international collaborative study of the epidemiology of esophageal atresia or stenosis. Reprod Toxicol. 1993;**7**(5):405–21.
10. Orford J, Glasson M, Beasley S, Shi E, Myers N, Cass D. Oesophageal atresia in twins. Pediatr Surg Int. 2000;16(8):541–5.
11. Casteels K, Devlieger H, Lerut T, Eggermont E. Familial occurrence of esophageal atresia. Tijdschrift voor kindergeneeskunde. 1993;61(3):100–2.
12. Choinitzki V, Zwink N, Bartels E, Baudisch F, Boemers TM, Holscher A, Turial S, Bachour H, Heydweiller A, Kurz R, et al. Second study on the recurrence risk of isolated esophageal atresia with or without trachea-esophageal fistula among first-degree relatives: no evidence for increased risk of recurrence of EA/TEF or for malformations of the VATER/VACTERL association spectrum. Birth Defects Res A Clin Mol Teratol. 2013;97(12):786–91.
13. Brosens E, Marsch F, de Jong EM, Zaveri HP, Hilger AC, Choinitzki VG, Holscher A, Hoffmann P, Herms S, Boemers TM, et al. Copy number variations in 375 patients with oesophageal atresia and/or tracheoesophageal fistula. Eur J Hum Genet. 2016;24(12):1715–23.
14. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356–69.
15. Timberlake AT, Choi J, Zaidi S, Lu Q, Nelson-Williams C, Brooks ED, Bilguvar K, Tikhonova I, Mane S, Yang JF, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. eLife. 2016;5:e20125.
16. Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M, Shen Y, Yu G, Liu J, Xie H, et al. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. N Engl J Med. 2015;372(4):341–50.
17. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to Omnigenic. Cell. 2017;169(7):1177–86.
18. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
19. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220–2.
20. Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A. TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. F1000Research. 2017;6:664.
21. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21(6):974–84.
22. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333–9.
23. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kahari AK, Lundin P, Che H, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet. 2017;25(11):1253–60.
24. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. Genome Biol. 2016;17(1):122.
25. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176(3):535–548.e524.
26. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019:531210.
27. Seelow D, Schwarz JM, Schuelke M. GeneDistiller--distilling candidate genes from linkage intervals. PLoS One. 2008;3(12):e3874.
28. Deshpande NP, Riordan SM, Castano-Rodriguez N, Wilkins MR, Kaakoush NO. Signatures within the esophageal microbiome are associated with host genetics, age, and disease. Microbiome. 2018;6(1):227.
29. Yang LX, Li BL, Liu XH, Yuan Y, Lu CJ, Chen R, Zhao J. RNA-seq reveals determinants of sensitivity to chemotherapy drugs in esophageal carcinoma cells. Int J Clin Exp Pathol. 2014;7(4):1524–33.
30. Wyględowska-Kania M, Gola J, Uttecht-Pudełko A, Wcisło-Dziadecka D, Kapral M, Strzałka-Mrozik B, Kruszniewska-Rajs C, Tkacz M, Mazurek U, Brzezińska-Wcisło L. Defensin DEFB4A transcript level in the differentiation of keratoacanthoma, squamous and basal cell carcinomas, vol. 28; 2015.
31. Zhang Y, Huang X, Qi J, Yan C, Xu X, Han Y, Wang M. Correlation of genomic and expression alterations of AS3 with esophageal squamous cell carcinoma. J Genet Genomics. 2008;35(5):267–71.
32. Shi N, Jin F, Zhang X, Clinton SK, Pan Z, Chen T. Overexpression of human beta-defensin 2 promotes growth and invasion during esophageal carcinogenesis. Oncotarget. 2014;5(22):11333–44.
33. Fujita T, Ikuta J, Hamada J, Okajima T, Tatematsu K, Tanizawa K, Kuroda S. Identification of a tissue-non-specific homologue of axonal fasciculation and elongation protein zeta-1. Biochem Biophys Res Commun. 2004;313(3):738–44.
34. Ishii H, Baffa R, Numata SI, Murakumo Y, Rattan S, Inoue H, Mori M, Fidanza V, Alder H, Croce CM. The FEZ1 gene at chromosome 8p22 encodes a leucine-zipper protein, and its expression is altered in multiple human tumors. Proc Natl Acad Sci U S A. 1999;96(7):3928–33.
35. Vergouwe FWT, Hanneke IJ, Biermann K, Erler NS, Wijnen RMH, Bruno MJ, Spaander MCW. High Prevalence of barrett's esophagus and esophageal squamous cell carcinoma after repair of esophageal atresia. Clin Gastroenterol Hepatol. 2018;16(4):513–521.e516.
36. Katz J, Malik A, Basit H. Embryology, esophagus. In: StatPearls. Treasure Island: StatPearls Publishing StatPearls Publishing LLC; 2019.
37. Vergouwe FW, Gottrand M, Wijnhoven BP, Hanneke IJ, Piessen G, Bruno MJ, Wijnen RM, Spaander MC. Four cancer cases after esophageal atresia repair: time to start screening the upper gastrointestinal tract. World J Gastroenterol. 2018;24(9):1056–62.
38. Saisawat P, Kohl S, Hilger AC, Hwang D-Y, Yung Gee H, Dworschak GC, Tasic V, Pennimpede T, Natarajan S, Sperry E, et al. Whole-exome resequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association. Kidney Int. 2014;85(6):1310–7.
39. Webber DM, MacLeod SL, Bamshad MJ, Shaw GM, Finnell RH, Shete SS, Witte JS, Erickson SW, Murphy LD, Hobbs C. Developments in our understanding of the genetic basis of birth defects. Birth Defects Res A Clin Mol Teratol. 2015;103(8):680–91.
40. Butali A, Suzuki S, Cooper ME, Mansilla AM, Cuenco K, Leslie EJ, Suzuki Y, Niimi T, Yamamoto M, Ayanga G, et al. Replication of genome wide association identified candidate genes confirm the role of common and rare variants in PAX7 and VAX1 in the etiology of nonsyndromic CL(P). Am J Med Genet A. 2013;161a(5):965–72.
41. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13(2):135–45.


## Publisher's Note