Identification and validation of candidate risk genes in endocytic vesicular trafficking associated with esophageal atresia and tracheoesophageal fistulas

Guojie Zhong,1,2,15 Priyanka Ahimaz,3,15 Nicole A. Edwards,4,15 Jacob J. Hagen,1,3 Christophe Faure,5 Qiao Lu,1,5 Paul Kingma,6 William Middlesworth,7 Julie Khlevner,6 Mahmoud El Fiky,9 David Schindel,10 Elizabeth Fialkowski,11 Adhish Kashyap,4 Sophia Forlenza,6,12 Alan P. Kenny,6,12 Aaron M. Zorn,4,* Yufeng Shen,1,13,* and Wendy K. Chung3,14,*

Summary

Esophageal atresias/tracheoesophageal fistulas (EA/TEF) are rare congenital anomalies caused by aberrant development of the foregut. Previous studies indicate that rare or de novo genetic variants significantly contribute to EA/TEF risk, and most individuals with EA/TEF do not have pathogenic genetic variants in established risk genes. To identify the genetic contributions to EA/TEF, we performed whole genome sequencing of 185 trios (probands and parents) with EA/TEF, including 59 isolated and 126 complex cases with additional congenital anomalies and/or neurodevelopmental disorders. There was a significant burden of protein-altering de novo coding variants in complex cases (p = 3.3 × 10−4), especially in genes that are intolerant of loss-of-function variants in the population. We performed simulation analysis of pathway enrichment based on background mutation rate and identified a number of pathways related to endocytosis and intracellular trafficking that as a group have a significant burden of protein-altering de novo variants. We assessed 18 variants for disease causality using CRISPR-Cas9 mutagenesis in Xenopus and confirmed 13 with tracheoesophageal phenotypes. Our results implicate disruption of endosome-mediated epithelial remodeling as a potential mechanism of foregut developmental defects. Our results suggest significant genetic heterogeneity of EA/TEF and may have implications for the mechanisms of other rare congenital anomalies.

Introduction

Esophageal atresia (EA) is a congenital abnormality of the esophagus, co-occurring with tracheoesophageal fistula (TEF) in 70%–90% cases.1,2 The overall worldwide incidence of EA/TEF is 2.4 per 100,000 births.3 Approximately 55% of individuals with EA/TEF are complex with additional congenital anomalies4 in the cardiovascular, musculoskeletal, urinary, gastrointestinal, or central nervous system. The genetic causes of EA/TEF include chromosome anomalies or variants in genes involved in critical developmental processes that are dosage sensitive.5 Several EA/TEF risk genes include the transcriptional regulators SOX2, MYCN, CHD7, FANCB, and members of FOX transcription factor family.6,7 VACTERAL frequently includes EA/TEF and is frequently of unknown etiology.

Mouse models have demonstrated that precise regulation of the transcription factors Nkx2-1, Sox2, and Foxf1 by WNT, bone morphogenetic protein 4 (BMP4), and Hedgehog signaling pathways is required for patterning of the fetal foregut and separation of the esophagus and trachea.8 Moreover, EFTUD2 haploinsufficiency leads to syndromic EA. EFTUD2 encodes one of the major components of the spliceosome, emphasizing the necessity of mRNA maturation through the spliceosome complex for normal development.10 Recently we have shown that de novo variants are major contributors to EA/TEF genetic risk, especially in genes that are targets of SOX2 or EFTUD2.11 However, it remains unclear how developmental signaling pathways, transcription factors, and RNA metabolism control the cellular behavior of tracheoesophageal morphogenesis.

Despite previous studies of the genetics in several syndromes that include EA/TEF and mouse models, the etiology in most cases of EA/TEF is still unexplained. To identify the genetic etiologies of EA/TEF, we performed whole genome sequencing (WGS) of 185 individuals with EA/TEF, most without a family history of EA/TEF, and their

1Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA; 2Integrated Program in Cellular, Molecular, and Biomedical Studies, Columbia University, New York, NY, USA; 3Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA; 4Center for Stem Cell & Organoid Medicine (CuSTOM), Division of Developmental Biology, Cincinnati Children’s Hospital Medical Center, Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA; 5Division of Pediatric Gastroenterology, CHU Sainte-Justine, Montreal, QC, Canada; 6Division of Neonatology, Cincinnati Children’s Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA; 7Division of Pediatric Gastroenterology, CHU Sainte-Justine, Montreal, QC, Canada; 8Division of Neonatology, Cincinnati Children’s Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA; 9Division of Pediatric Surgery, Columbia University Irving Medical Center, New York, NY, USA; 10Division of Pediatric Gastroenterology, Hepatology and Nutrition, Columbia University Irving Medical Center, New York, NY, USA; 11Pediatric Surgery, Faculty of Medicine, Cairo University, Cairo, Egypt; 12Division of Pediatric Surgery, UT Southwestern School of Medicine Dallas, Texas, USA; 13Division of Pediatric Surgery, Oregon Health and Science University, Portland, OR, USA; 14Division of Pulmonary Biology, Cincinnati Children’s Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA; 15Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA

15These authors contributed equally

*Correspondence: aaron.zorn@cchmc.org (A.M.Z.), ys2411@cumc.columbia.edu (Y.S.), wkc15@columbia.edu (W.K.C.)
biological parents. We confirmed our previous results from a smaller EA/TEF cohort, demonstrating an overall enrichment of de novo coding variants in complex cases. Functional enrichment analysis identified a striking convergence of putative risk genes in biological pathways related to endocytosis, membrane dynamics, and intracellular transport. We then used CRISPR-generated Xenopus mutant models to successfully confirm 13 of 18 candidate risk genes for EA/TEF. Together with recent reports that endosome-mediated membrane remodeling is required for tracheoesophageal morphogenesis in animal models, this suggests that disruptions in endosome trafficking may be a feature of many complex EA/TEF cases.

Methods

Participants recruitment

Individuals with EA/TEF were recruited as part of the CLEAR consortium from Columbia University Irving Medical Center in New York, USA, Center Hospitalier Universitaire Sainte-Justine in Montreal, Canada, Cincinnati Children’s Hospital, in Ohio, USA, Cairo University General Hospital in Cairo, Egypt, University of Texas Southwestern Medical Center in Dallas, USA, and Oregon Health and Science University in Portland, USA. Participants eligible for the study included those diagnosed with EA/TEF without an identified genetic etiology based upon medical record review. All participants provided informed consent. The overall study was approved by the Columbia University institutional review board and each affiliated site. Blood and/or saliva samples were obtained from the probands and both biological parents. A three-generation family history was taken at the time of enrollment, and clinical data were extracted from the medical records and by participant and parental interview.

We performed whole genome sequencing (WGS) on 185 probands without prior sequence-based genetic testing diagnosed with EA/TEF and their parents. DNA from 75 probands was isolated from saliva samples, and DNA from the remaining 110 probands was isolated from blood samples. Individuals with only EA/TEF were classified as isolated cases (59 in total), and individuals with other type of congenital abnormalities or neurodevelopmental disorders were classified as complex cases (126 in total; Table S1).

WGS analysis

We identified de novo coding variants using previously published procedures with heuristic filters augmented with in silico confirmation by DeepVariant (Table S2). We used ANNOVAR and VEP to annotate variants with population allele frequency and gene family were filtered out of further analysis. We identified de novo copy number variants (CNVs) customized pipeline as described in our previous study. Briefly, we applied CNVnator (v0.3.3) with the bin size set as 100 bp to predict CNV segments by read depth evidence and Lumpy v0.2.13 and S深刻的 v0.1.4 to quantify pair-end/split-read (PE/SR) evidence. We only included the CNVs supported by both read depth and PE/SR in downstream analysis. Among the CNVs called in probands with Mendelian errors (that they were not called in any of the parents), we called de novo CNVs by visualization of both normalized read depth and allele fraction of SNP sites. We mapped de novo CNVs on GENCODE v29 protein-coding genes with at least 1 bp in the shared interval. We annotated the genes with variant intolerance metric by ExAC pLI, haploinsufficiency metric by Episcore, haploinsufficiency and triplosensitivity of genes from ClinGen genome dosage map, and CNV syndromes from DECIPHER v11.1.

Burden test

We divided the cohort into two categories based on their phenotypes (isolated and complex) and performed burden tests on both groups and the aggregated group. For each group, we divided de novo coding variants into four types: synonymous, LGD, missense, and protein altering (defined as combination of LGD and missense variants). For each variant type, we calculated the expected number of variants based on a background mutation rate model. We used a single-sided Poisson test to test whether the number of observed de novo variants is significantly higher than expected. We performed the test in all genes, genes intolerant of loss-of-function variants (“constrained genes” based on gnomAD pLI ≥ 0.5), and non-constrained genes. Population attributable risk (PAR) was calculated as follows: $PAR = \frac{N_{1} - N_{2}}{N_{1}}$, where $N_{1}$, $N_{2}$, $N$ are the observed number of individuals with heterozygous protein-altering variants, expected number of individuals with heterozygous protein-altering variants, and the number of all cases, respectively.

Pathway enrichment analysis of de novo protein-altering variants in complex cases

To identify the pathways associated with de novo protein-altering variants, we performed pathway enrichment analysis on the gene ontology (GO) pathways and human phenotype ontology (HPO) terms from GSEA database (version v7.2) in complex cases. We only considered the pathways with at least two protein-altering variants (defined by combination of LGD and missense variants) expected by chance based on background mutation rate model. Based on these criteria, we selected a total of 907 pathways for downstream analysis. We performed a one-sided Poisson test of observed variants versus expectation in each pathway. Since many pathways have shared genes, we performed simulations under the null hypothesis to estimate the family-wise error rate (FWER) for a given p value. In each round, we randomly generated de novo LGD or missense variants based on the background mutation rate and calculated p values for each gene. Based on simulation results, we estimated FWER as follows:

$$FWER(p_{0}) = \frac{S(p \leq p_{0})}{N}$$

where $S(p \leq p_{0})$ is the total number of pathways that have p values smaller or equal to $p_{0}$ in all simulations, and $N$ is the number of simulations. We used both Jaccard index and correlation to show the overlap of two pathways. For each pair of pathways, the Jaccard index was defined as the aggregated mutation rate of overlapping genes divided by aggregated mutation rate of all genes, and correlations were calculated as the Pearson correlation during
We tested protein interactions of \textit{de novo} protein-altering variants in complex cases using STRING(v11.0) with default settings and default interaction sources. Edges were filtered by STRING score $\geq 0.4$ and visualized by Cytoscape. \textsuperscript{31} Proteins that were not connected to any other genes after interaction filtration were removed from the network. Network layout was generated by “Prefuse Forced Directed OpenCL Layout” algorithm in Cytoscape. For each gene, Degree was calculated as the sum of all StringDB scores.

### Protein-protein interaction analysis

We tested protein interactions of \textit{de novo} protein-altering variants in complex cases using STRING(v11.0) \textsuperscript{30} with default settings and simulation. Network layout is generated by “Prefuse Forced Directed OpenCL Layout” algorithm in Cytoscape.

### F0 \textit{Xenopus tropicalis} CRISPR-Cas9 mutagenesis screen

\textit{Xenopus} experiments were performed using guidelines approved by the CCHMC Institutional Animal Care and Use Committee (IACUC 2019-0053). \textit{Xenopus tropicalis} adult frogs were purchased from NASCO (USA) or raised in house and maintained in the CCHMC vivarium under normal housing conditions. \textit{Xenopus} embryos were obtained by \textit{in vitro} fertilization or natural mating as previously described.\textsuperscript{32,33} Germ line sox2$^{-/-}$ embryos (F2 generation) were obtained by mating sox2$^{-/-}$ adults obtained from the National \textit{Xenopus} Resource (NKR, USA; RRID: SCR_013731).

For F0 CRISPR-Cas9 indel mutagenesis, guide RNAs (gRNAs) were designed using CRISPRScan\textsuperscript{34} based on the \textit{Xenopus tropicalis} v9.1 genome assembly on Xenbase.\textsuperscript{35} gRNAs were designed to generate either null mutations (early in the coding sequence) or in the coding region similar to the corresponding mutation in our human cohort. \textit{In vitro} transcribed gRNAs were synthesized using MEGAscript T7 Transcription Kit (ThermoFisher, USA) according to manufacturer’s instructions, or purchased as AltR-cRNA (Integrated DNA Technologies, USA). CrRNAs were annealed with AltR-tracrRNA prior to embryo injections according to manufacturer’s guidelines. Guide RNAs (500–700 pg) were complexed with recombinant Cas9 protein (1 ng, PNA Biosciences) to manufacturer’s guidelines, or purchased as AltR-crRNA (Integrated DNA Technologies, USA). CrRNAs were annealed with AltR-tracrRNA prior to embryo injections according to manufacturer’s guidelines. Guide RNAs (500–700 pg) were complexed with recombinant Cas9 protein (1 ng, PNA Biosciences) and injected into \textit{X. tropicalis} embryos at the one- or two-cell stage. For negative controls, a gRNA designed targeting tyrosinase (\textit{tyr}) was injected to calculate a baseline percentage of defective tracheoesophageal development in \textit{Xenopus} (~2%).

Three-day-old injected tadpoles (stage NF44) were fixed and processed for wholemount immunostaining as previously described\textsuperscript{12} using the following primary antibodies: mouse anti-SOX2 (Abcam, ab79351, 1:1,000), goat anti-FOXF1 (R&D systems, AF4798, 1:300), and rabbit anti-NKX2-1 (SCBT, sc-13040X, 1:300). Imaging was performed using a Nikon A1 inverted LUNA confocal microscope with constant laser settings for all embryos. Image analysis was performed using NIS Elements (Nikon, USA). After image analysis, each embryo was genotyped by PCR amplification of the target region followed by Sanger sequencing. Since F0 CRISPR-mutagenesis is mosaic, different cells can have different mutations, so we used the Synthego ICE software tool\textsuperscript{36} to deconvolute the proportion and sequence of each indel mutation in each embryo (Figure S1). Genotyping primers and gRNA sequences are in Table S3.

We only scored phenotypic data from embryos that had >40% mutation rate. For each gene, CRISPR-mutagenesis experiments were independently repeated at least twice in different batches of embryos, analyzing 5–15 individual mutant tadpoles per experiment. A candidate risk gene was determined to be likely causative if more than 10% of mutant tadpoles had a tracheal or esophageal defect (LTEC, occluded esophagus, failed separation), compared to the baseline rate of <2% in control injected tadpoles.

### Results

A total of 185 individuals with EA/TEF were enrolled into the study, including 102 (55%) male and 83 (45%) female...
Table 2. Burden of de novo variants in all cases

| Variant type   | All cases (n = 185) | Isolated cases (n = 59) | Complex cases (n = 126) |
|----------------|---------------------|------------------------|------------------------|
|                | Obs | Exp | Fold | p value | Obs | Exp | Fold | p value | Obs | Exp | Fold | p value |
| Synonymous     | 58  | 61.6| 0.94 | 0.7     |      |      |      |         |      |      |      |         |
| LGD            | 23  | 19.2| 1.20 | 0.2     | 3    | 6.1 | 0.49 | 0.9     | 20  | 13.1| 1.53 | 0.045   |
| Missense       | 168 | 137.3| 1.22 | 0.0062  | 44   | 43.8| 1.0  | 0.51    | 124 | 93.5| 1.33 | 0.0015  |
| Protein altering (LGD + missense) | 191 | 156.5| 1.22 | 0.0042  | 47   | 49.9| 0.94 | 0.68    | 144 | 106.6| 1.35 | 3.3 x 10^{-4} |

Burdens were calculated in all cases, isolated cases, and complex cases. Protein-altering variants were defined as LGD and missense variants. LGD is likely gene disrupting. Obs is observed. Exp is expected.

probands. Proband were between the ages of 2 days and 54.5 years with an average of 8.2 years old at enrollment (Table 1). The majority (52.4%) were type C EA/TEFs. Fifty-nine probands had isolated EA/TEF, and 126 probands had neurodevelopmental delay and/or at least one additional congenital anomaly and were classified as non-isolated. Of the non-isolated cases, the most common associated anomalies were congenital heart defects (65; 51.5%), skeletal defects (48; 38%), and renal defects (40; 31.7%). Other congenital anomalies included genitourinary defects (non-renal) (16; 12.7%), laryngotracheal defects (13; 10.3%), gastrointestinal defects (9; 7%), limb defects (7; 5.5%), neural tube defects (5; 3.9%), craniofacial defects (5; 3.9%), and other anomalies were seen in 12 probands (9.5%). Twenty-five probands (19.8%) had neurodevelopmental delay. Fifty-five of the cases (30%) previously had mental delay. Fifty-five of the cases (30%) previously had developmental disorders. We identified 13 de novo CNVs variants in 134 probands, including two individuals with heterozygous deletions of 21q11. None of the CNVs overlapped with any of the genes with de novo sequence variants (Table S5).

We performed a burden test for enrichment of de novo coding variants in all cases, isolated cases, and complex cases respectively (Table 2). The number of synonymous variants is close to expectation (fold = 0.94, p = 0.7). Overall, there is a significant burden of de novo protein-altering variants (LGD or missense) (fold = 1.22, p = 4.2 x 10^{-3}). The burden is almost entirely observed in complex cases (fold = 1.35, p = 3.3 x 10^{-4}), as there is no evidence of de novo burden in isolated cases (fold = 0.94, p = 0.68). In complex cases (Table 3), the burden of LGD variants is mostly in genes that are intolerant of loss-of-function variants (defined as gnomAD pLI ≥ 0.5, “constrained genes”; fold = 2.8, p = 2.3 x 10^{-3}), similar to other developmental disorders. The burden of de novo missense variants is also higher in constrained genes compared to non-constrained genes (fold = 1.57 versus 1.22), although it is marginally significant in both constrained and non-constrained gene sets (p = 3.3 x 10^{-3} and 0.045, respectively). We estimate that about 38 genes carrying these variants in the complex cases are risk genes. Overall, de novo protein-altering variants explain about 30% of PAR of complex EA/TEF.

We assessed de novo protein-coding variants for pathogenicity using the ACMG criteria (Table 4 and Table S6). Of the 185 cases, only two clearly had a molecular diagnosis consistent with the phenotype (EFTU2 and MYCN associated with mandibulofacial dysostosis, Guion-Almeida type [OMIM: 610536] and Feingold syndrome [OMIM: 164280],

Table 3. Burden of protein-altering de novo variants in complex cases stratified by gene variant intolerance

| Gene group          | Type of variants | Obs | Exp | Fold | p value |
|---------------------|------------------|-----|-----|------|---------|
| Constrained genes   | LGD              | 11  | 3.9 | 2.82 | 0.0023  |
|                     | missense         | 44  | 28.1| 1.57 | 0.0033  |
|                     | protein altering | 55  | 32.0| 1.72 | 1.4 x 10^{-4} |
| Non-constrained genes | LGD          | 9   | 9.2 | 0.98 | 0.57    |
|                     | missense         | 80  | 65.5| 1.22 | 0.045   |
|                     | protein altering | 89  | 74.6| 1.19 | 0.057   |
Table 4. *De novo* LGD variants. LGD are likely gene disrupting

| Gene   | Variant   | Protein                      | Variant type | CADD score | gnomAD pLI | OMIM             | Individual phenotype                                                                                   | ACMG variant class |
|--------|-----------|------------------------------|--------------|------------|------------|------------------|--------------------------------------------------------------------------------------------------------|-------------------|
| CAMK2B | c.558del  | p.R187Afs*16                 | LGD          | .          | 0.74       | autosomal dominant mental retardation (607707) | EA + TEF type C, long gap, extra ribs, congenital scoliosis, developmental delay                        | pathogenic        |
| GTF2I  | c.761_762del | p.Q254Rfs*5                  | LGD          | .          | 1          | none            | EA + TEF, atrial septal defect, bilateral inguinal hernia                                              | VUS               |
| AMER3  | c.2236C>T  | p.R746*                      | LGD          | 35         | 0.62       | non-OMIM gene                                           | EA + TEF, club feet, pyelectasis, atrial septal defect, developmental delay                           | VUS               |
| EFTUD2 | c.2419del  | p.Q807Rfs*21                 | LGD          | .          | 1          | mandibulofacial dysostosis, Guion-Almeida type (610536) | EA + TEF, club feet, pyelectasis, atrial septal defect, developmental delay                           | pathogenic        |
| ARHGAP21| c.1711C>T  | p.R571*                      | LGD          | 31         | 1          | none            | EA + TEF type C, multicystic dysplastic left kidney, patent ductus arteriosus                          | VUS               |
| ARHGAP17| c.499C>T   | pra9021                      | LGD          | 37         | 0.02       | none            | EA + TEF type C                                                                                       | VUS               |
| MYCN   | c.153_154insC | p.K52Qfs*3                   | LGD          | .          | 0.89       | Feingold syndrome (164280)                             | EA + TEF type C, microcephaly, clinodactyly, developmental delay                                    | pathogenic        |
| USP9X  | c.4775del  | p.G1592Vfs*4                 | LGD          | .          | 1          | X-linked mental retardation (300968)                   | EA, extra thumbs and dysmorphic features, rectus abdominis diastasis, severe laryngomalacia, seizures, hypotonia, intellectual disability | pathogenic        |
| ADRM1  | c.214-28_223del | .                           | LGD          | .          | 1          | none            | EA + TEF type C, atrial septal defect, Ventricular septal defect, developmental delay                  | VUS               |
| ADD1   | c.1A>G     | p.M1?                        | LGD          | 25.1       | 0.99       | none            | EA + TEF type C, vertebral anomalies, extra ribs, patent ductus arteriosus, horseshoe kidney, bilateral radial hypoplasia, thumb anomaly, imperforate anus | pathogenic        |
| FBXO10 | c.1419+1G>A | ..                           | LGD          | 26         | 0          | none            | EA + TEF type C, vertebral anomaly, coarctation of aorta                                            | VUS               |
| CHERP  | c.1306-1G>A | ..                           | LGD          | 23.5       | 1          | none            | EA + TEF type C, renal ectopia, atrial septal defect, scoliosis                                       | VUS               |
| IL32   | c.450_451insC | p.G151Rfs*13                 | LGD          | .          | 0          | none            | EA + TEF long gap, duodenal atresia, small hole in heart                                              | VUS               |
| RASA2  | c.82del    | p.D28Tfs*32                  | LGD          | .          | 0          | none            | EA + TEF type C, extra ribs, congenital scoliosis, developmental delay                                 | VUS               |
| AMACR  | c.197dup   | p.R67Afs*75                  | LGD          | .          | 0.03       | alpha-methylacyl-CoA racemase deficiency (AR-614307); bile acid synthesis defect (AR-214950)          | EA + TEF type C and developmental delay                                                             | pathogenic        |
| HACE1  | c.805C>T   | p.R269*                      | LGD          | 40         | 0          | spastic paraplegia and psychomotor retardation with or without seizures (AR-616756)                   | EA + TEF type D, ventricular septal defect, and atrial septal defect                                | pathogenic        |
| MANB1  | c.72C>G    | p.Y24*                       | LGD          | 34         | 0.03       | none            | EA + TEF type C, short gap                                                                           | VUS               |

(Continued on next page)
respectively). One individual with a de novo p.G365S SMAD6 with a CADD score of 32 has a phenotype partially overlapping with conditions associated with SMAD6 and may represent an expansion of the phenotypes associated with SMAD6. One individual with a de novo p.T647I variant in GLS with a CADD score of 27.1 has a phenotype that at the age of 2 does not overlap with OMIM: 618339 with infantile cataracts, skin abnormalities, and intellectual disability. Of the 24 cases with de novo LGD variants, 21 were associated with complex phenotypes (Table 4).

Protein-altering variants in complex cases are involved in endosome trafficking and developmental pathways

While complex cases have a significant burden of de novo variants, no one gene harbors more than one LGD or missense de novo variant, making it impossible to identify individual risk genes with sufficient statistical support.

To investigate the aggregate properties of risk genes, we performed pathway enrichment analysis on protein-altering de novo variants in complex cases (n = 126). We focused on GO pathways and HPO terms. To ensure sufficient statistical power, we only considered the pathways that are expected to have at least two protein-altering variants by chance in 126 subjects. We compared the observed variants in each pathway to the expected number of variants estimated from background mutation rate and tested the enrichment using a Poisson test. We corrected the multi-testing p values to FWER based on simulations. Eight GO pathways and five HPO terms are enriched with protein-altering de novo variants with FWER ≤ 0.05 (Figure 1A, Table S7). The enriched GO pathways are related to autophagy processes, membrane regulation, and intracellular transport and localization, while the HPO terms are related to other developmental disorders (Figure 1B). A total of 86 genes are involved in at least one significant pathway. Fifty-five genes are involved in endocytosis and transcytosis pathways. Forty-five genes are involved in pathways related to other developmental disorders. The enrichment in GO pathways is mostly driven by de novo missense variants, whereas the enrichment in HPO terms is driven by both LGD and missense variants (Figure 1A). These results remain consistent if we exclude the two cases with the 22q11 deletion (Table S8). These findings are consistent with animal model studies in which pleiotropic signaling pathways and endosome-mediated epithelial remodeling are required for tracheoesophageal morphogenesis.²

We also investigated the functional interactions among the genes (n = 143) with protein-altering de novo variants in complex cases. Based on StringDB (v11.0),³ the number of protein-protein interactions is significantly larger than expected (PPI enrichment p value = 0.0021; Figure 2).

CRISPR mutation of candidate risk genes in Xenopus disrupts trachea-esophageal morphogenesis

The underlying biology of trachea and esophageal development is conserved between humans and other terrestrial...
vertebrates, and animal models have proven effective in assessing candidate risk variants from human affected individuals.\(^2\) We therefore turned to the rapid functional genomics possible in the amphibian *Xenopus*, which is increasingly being used to model human developmental disorders\(^{39,40}\) including tracheoesophageal birth defects.\(^{12}\)

We tested candidate risk variants by CRISPR-Cas9 mutagenesis of the orthologous genes in *Xenopus tropicalis*, assaying F0 mutant embryos rather than establishing multi-generational lines\(^{11}\) since this is faster and more closely mimics the *de novo* mutations in human EA/TEF individuals. F0 mutagenesis results in embryos with a range of mosaic indel mutations. We found that F0 mutagenesis of *sox2*, a known EA/TEF risk gene in humans, resulted in a trachea-esophageal phenotype indistinguishable from F2 *sox2*\(^{+/−}\) germline mutants with a failure of the foregut to

---

**Figure 1. Pathway enrichment analysis**

(A) Volcano plot. Each dot represents a pathway. X axis represents the enrichment rate in log scale, and Y axis is the Poisson test p value in log10 scale. The horizontal dashed line marks family-wise error rate (FWER) of 0.05. Significant pathways (FWER < 0.05) are colored by the percentage of LGD variants, and other pathways are colored gray.

(B) Pathway overlaps. Each circle represents a pathway with FWER < 0.05. Circle size is proportional to the number of observed *de novo* variants in the pathway; circle color represents the FWER; edge width is determined by the Jaccard index between two pathways, and edge color represents the correlation coefficient of the two pathways under the null in simulations.
separate into distinct esophagus and trachea (Figures 3B and 3C). Moreover, unlike human EA individuals with heterozygous $SOX2^{+/−}$ mutations, heterozygous mouse and *Xenopus sox2$^{+/−}$* mutants do not exhibit tracheoesophageal defects.\(^8\),\(^{42}\)

We prioritized and selected 18 candidate risk genes to test based on (1) the likelihood that the variant in affected individual was damaging, (2) expression in the *Xenopus* and mouse fetal foregut, and (3) the predicted function focusing on genes implicated in endosome trafficking or signaling pathways that pattern the fetal foregut (Table 5). gRNAs were designed to generate loss-of-function (null) mutations or in a few cases where early embryonic lethality was predicted, an affected individual-like mutation targeting a conserved sequence near the corresponding variant. We genotyped each CRISPR-injected embryo and assessed the trachea-esophageal phenotype in embryos with $>40\%$ damaging indel mutations. At 3 days of development (stage NF44), when the trachea and esophagus have normally separated (Figure 3A), tadpoles were fixed and assessed by confocal immunostaining for (Figure 3).

Thirteen of 18 genes screened exhibited defective trachea-esophageal development in $>10\%$ of mutated tadpoles (Table 5, Figure 3). The most common phenotype was an LTEC where the trachea and esophagus failed to separate near the larynx (e.g., sox2, eftud2, itsn1) (Figures 3C–3E), or a disorganized esophageal epithelium, likely leading to EA later in development (e.g., arhgap21 and disp1) (Figures 3F and 3G). This failure to separate the embryonic foregut is a typical manifestation commonly observed in both mouse and *Xenopus* embryos with mutations in known EA risk genes.\(^7\) Interestingly, several of the gene mutations also resulted in co-occurring defects in other organ systems like the EA/TEF human cases including microphthalmia, microcephaly, and craniofacial malformations. Notably, five genes are implicated in signaling pathways known to regulate foregut patterning (amer3, apc2, celsr2, disp1, smad6), while five other genes are implicated in endocytosis and/or intracellular trafficking (abra, arhgap21, ap1g2, itsn1, rapgef3) (Table 5).

**Discussion**

In this study, we identified 249 de novo coding variants in 185 EA/TEF individuals, including 23 LGD variants and 168 missense variants. Only two cases were associated with pathogenic variants in genes previously established to cause EA/TEF, suggesting that most of our findings are identifying genetic associations with EA/TEF in genes not previously associated with EA/TEF. Protein-altering de novo variants are enriched in complex cases. Consistent with previous studies of congenital anomalies, those variants showed greater enrichment in constrained genes. Pathway analysis showed that endocytosis, membrane regulation, and intracellular trafficking-related processes are enriched with protein-altering variants. Considering recent findings in mouse and *Xenopus* that endosome-mediated epithelial remodeling acts downstream of Hedgehog-Gli signaling to regulate tracheoesophageal morphogenesis,\(^12\) it is possible
that disruption in endocytic vesicular trafficking may be a common mechanism in many EA/TEF individuals.

Endocytic vesicular trafficking is regulated by small GTPases (Rab/Rho) that link endocytosis of membrane-bound vesicles to the actin intracellular transport machinery, which moves vesicles to different subcellular compartments: to lysosomes in the case of autophagy, to different membrane domains in the case of recycling endosomes, from the Golgi and ER to the cell surface for maturation of membrane proteins, and from basal to apical membranes in the case of transcytosis. Endocytic trafficking can influence morphogenesis in many ways: by changing cell shape, by dynamic remodeling of cell adhesion and junctional complexes, and by regulation of cell migration or cell signaling. Moreover, one of the candidate genes that we tested, ITSN1, encodes a multidomain adaptor protein that coordinates the intracellular transport of endocytic vesicles. ITSN1 is also an autism risk loci and consistent with the neurodevelopmental disorders also present in the EA/TEF individual, IIts1 is required for neural dendrite formation in rodents, where it physically interacts with core endocytic protein Dmm2 acting as a Cdc42-GEF to promote actin-mediated endosome transport. Thus, the finding that several candidate genes validated in Xenopus are implicated in endocytosis or GTPase activity (abra, arhgap21, ap1g2, itsn1, rapgef3, rab3gap2) suggests that the EA/TEF in the individuals may have been due to disrupted foregut morphogenesis.

Our analysis also revealed that LGD and missense variants in complex cases are involved in other developmental disorders, suggesting disruptions to pleiotropic pathways with roles in multiple organ systems. Indeed, Xenopus mutagenesis validated several genes implicated in signaling pathways known to regulate foregut patterning as well as the development of other organ systems including amer3, apc2, and celsr2 in the Wnt pathway, smad6 and sox2 in the BMP pathway, and disp1 required for secretion of Hedgehog ligands. In the future, as more functional data are collected on EA/TEF risk variants, it may be possible to link distinct signaling pathways or
cellular mechanisms such as endocytosis to different co-occurring anomalies in specific organ systems.

Overall, the genetics of EA/TEF is heterogeneous. With 126 complex cases that are overall significantly enriched with de novo protein-altering variants, we did not find a gene with such variants in multiple cases. This indicates that the number of risk genes contributing through de novo variants is large. A sustained effort to expand the cohort with genome sequencing is critical to improve statistical power to identify risk genes in humans.

EA/TEF, like most other congenital anomalies, does not yet have a ClinGen expert panel and has not yet had a formal ClinGen evidence review to establish gene-disease validity for the phenotype of EA/TEF. Some syndromes have been assessed by the syndromic disorders expert panel, but none of the assessed conditions is frequently or consistently associated with EA/TEF. Given the apparent genetic heterogeneity and the small number of genomic studies of EA/TEF, it will likely be some time before there is sufficient evidence to assess any genes beyond perhaps those associated with Fanconi anemia as having more than limited evidence. However, functional data such as that we present in this manuscript add significantly to the evidence review once there are six or more independent de novo predicted loss-of-function individuals with a similar phenotype.

One interesting observation is that all CRISPR-generated Xenopus mutants had severe tracheoesophageal clefts rather than atresia or fistulas. We expect that this is because the CRISPR editing strategy results in high mutagenesis rates and often loss-of-function alleles resulting in more severe tracheoesophageal phenotypes, in contrast to the individuals who have heterozygous variants. Indeed, in all most all reported cases where EA/TEF risk alleles have been modeled in mouse or Xenopus, heterozygous variants do not result in an EA/TEF phenotype, whereas null mutations exhibit a cleft with a single undivided foregut.2 This difference could be due to hypomorphic human variants versus null alleles in animals. In humans, null alleles in pleiotropic developmental genes are likely to be embryonic lethal and may not be viable to term. An additional factor is likely to be the fact that animal models are inbred, whereas the humans have diverse genetic backgrounds, likely associated with modifying alleles. In the future it will be important to test these possibilities with the exact affected individual alleles in animal models to obtain a better assess the genotype-phenotype relationship of these conditions.

### Table 5. EA/TEF candidate genes screened in Xenopus.

| Gene   | Function          | Xenopus TED frequency (n) | Co-occurring defects | % indels | Mutation type     |
|--------|-------------------|---------------------------|----------------------|----------|-------------------|
| sox2   | transcription factor | 100% (14)                |                      | 100% (germline) | null              |
| sox2   | transcription factor | 65% (17)                 | microphthalmia       | 91%      | null              |
| disp1  | Hedgehog signaling | 71% (14)                 |                      | 62%      | null              |
| anfr3  | Wnt signaling     | 62% (21)                 |                      | 57%      | null              |
| eftud2 | mRNA splicing     | 55% (22)                 | microphthalmia       | 92%      | affected individual-like |
| abra   | Rho signaling     | 45% (20)                 | craniofacial         | 92%      | null              |
| itsn1  | endocytosis       | 47% (19)                 | microphalmy         | 71%      | null              |
| itsn1  | endocytosis       | 42% (43)                 |                      | 72%      | affected individual-like |
| apc2   | Wnt signaling     | 37% (19)                 |                      | 87%      | null              |
| smad6  | BMP signaling     | 32% (31)                 | craniofacial heart looping | 68% | affected individual-like |
| arhgap21 | Rho signaling      | 30% (23)                 | craniofacial         | 86%      | null              |
| itgb4  | integrin          | 29% (14)                 | heart looping        | 76%      | null              |
| aptg2  | endocytosis       | 24% (17)                 | microphthalmia       | 92%      | null              |
| rapgef3 | Ras signaling     | 20% (5)                  |                      | 91%      | null              |
| celsr2 | Wnt/PCP signaling | 17% (24)                 |                      | 79%      | null              |
| ptpn14 | RTK signaling     | 13% (23)                 |                      | 94%      | null              |
| add1   | cytoskeleton      | 8% (24)                  | craniofacial         | 89%      | null              |
| map4k3 | MAPK signaling    | 8% (13)                  |                      | 71%      | null              |
| rab3gap2 | endocytosis       | 6% (18)                  |                      | 81%      | null              |
| arhgap17 | Rho signaling    | 0% (11)                  | gut looping           | 88%      | null              |
| pcdh1  | cell-cell adhesion | 0% (7)                  |                      | 47%      | null              |
| tyr (control) | pigmentation | 2% (71)                  |                      | n/d      | null              |

TED, tracheoesophageal defect.
Limitations of this study
Our study had limited statistical power to identify individual risk genes of EA/TEF based solely upon the human genetic studies due to the limited sample size. Collaborative human genetic studies of EA/TEF will be necessary to increase those sample sizes and better understand the spectrum of phenotypes associated with each gene. If somatic mutations play a significant role in disease pathogenesis, genetic analyses of blood or saliva may be insufficient to detect these genetic variants.

However, even with a modest human sample size with only one single human with a \textit{de novo} variant in the gene, we demonstrate the ability to effectively select disease-causing variants and functionally confirm the majority of the candidate genes using a moderate throughput F0 mutagenesis system. The combination of human genetics and model organism modeling is powerful for rare human genetic conditions associated with morphological defects. By examining pathways common across genes, we implicate endocytosis, membrane regulation, and intracellular trafficking in tracheoesophageal development, and these same processes are likely related to other congenital anomalies and neurodevelopmental disorders.

Data and code availability
The code generated during this study is available at GitHub: https://github.com/ShenLab/pathways pathways, which contains the code for pathway enrichment analysis of de novo variants with family-wise error rate estimation.

The accession number for the raw whole genome sequencing data reported in this paper is dbGaP:phs002161.

Supplemental information
Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2022.100107.

Acknowledgments
We would like to thank the individuals with EA/TEF and their families who participated in the study and the TOIS UK organization for their support of the study. We thank Patricia Lanzano, Jian-gyuan Hu, Liyong Deng, and Charles LeDuc from Columbia University for technical assistance. We would also like to thank the study coordinators Gentry Wools (UT-Southwest), Amanda Schreibis (Cincinnati Children’s Hospital) and Andrew Mason (Oregon Health and Science University) for their assistance. We also thank Dr. Na Zhu, Dr. Xueya Zhou, and other members of Chung and Shen labs for helpful discussions. The whole genome sequencing data were generated through NIH Gabriella Miller Kids First Pediatric experimental design: N.A.E., A.K. Xenopus experimental: N.A.E., A.K., S.F., A.P.K. Supervision: A.M.Z., Y.S. and W.K.C. The authors read and approved the final manuscript.

Declaration of interests
The authors declare no competing interests.

Received: August 24, 2021
Accepted: April 6, 2022
Corrected online: June 17, 2022

Author contributions
W.K.C., Y.S. and A.M.Z. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: W.K.C., Y.S. and A.M.Z. Acquisition, analysis, or interpretation of data: G.Z., P.A., N.A.E., J.J.H., C.F., P.K., W.M., J.K., M.E.F., D.S., E.F., A.K., S.F., A.P.K., A.M.Z., Y.S. and W.K.C. Drafting of the manuscript: G.Z., P.A., N.A.E., A.M.Z., Y.S. and W.K.C. Critical revision of the manuscript for important intellectual content: G.Z., P.A., N.A.E., J.J.H., C.F., P.K., W.M., J.K., M.E.F., D.S., E.F., A.K., S.F., A.P.K., A.M.Z., Y.S. and W.K.C. Statistical analysis: G.Z., P.A., N.A.E., A.M.Z., and Y.S. Xenopus experimental design: N.A.E., A.K. Xenopus experiment: N.A.E., A.K., S.F., A.P.K. Supervision: A.M.Z., Y.S. and W.K.C. The authors read and approved the final manuscript.

References
1. Nassar, N., Csaky-Szunyogh, M., Feldkamp, M.L., Khoshnood, B., Landau, D., Leong, N., Lopez-Camejo, J.S., Lowry, R.B., McDonnell, R., Merlob, P., et al. (2012). Prevalence of esophageal atresia among 18 international birth defects surveillance programs. Birth Defects Res. A. 94, 893–899. https://doi.org/10.1002/bdra.23067.
2. Edwards, N.A., Shacham-Silverberg, V., Weitz, L., Kingma, P.S., Shen, Y., Wells, J.M., Chung, W.K., and Zorn, A.M. (2021). Developmental basis of trachea-esophageal birth defects. Dev. Biol. 477, 85–97. https://doi.org/10.1016/j.ydbio.2021.05.015.
3. van Lennep, M., Singendonk, M.M.J., Dall’Oglio, L., Gottrand, F., Krishnan, U., Terheggen-Lagro, S.W.J., Omari, T.I., Benninga, M.A., and van Wijk, M.P. (2019). Oesophageal atresia. Nat. Rev. Dis. Primers 5, 26. https://doi.org/10.1038/s41572-019-0077-0.
4. Stoll, C., Alembik, Y., Dott, B., and Roth, M.P. (2017). Associated anomalies in cases with esophageal atresia. Am. J. Med. Genet. A. 173, 2139–2157. https://doi.org/10.1002/ajmg.a.38303.
5. Shaw-Smith, C. (2010). Genetic factors in esophageal atresia, tracheo-esophageal fistula and the VACTERL association: roles for FOXF1 and the 16q24.1 FOX transcription factor gene cluster, and review of the literature. Eur. J. Med. Genet. 53, 6–13. https://doi.org/10.3109/13602282.2009.100101.
6. Li, Y., Litingtung, Y., Ten Dijke, P., and Chiang, C. (2007). aberrant Bmp signaling and notochord delamination in the pathogenesis of esophageal atresia. Dev. Dyn. 236, 746–754. https://doi.org/10.1002/dvdy.21075.
7. Que, J., Choi, M., Ziel, J.W., Klingensmith, J., and Hogan, B.L. (2006). Morphogenesis of the trachea and esophagus: current players and new roles for noggin and Bmps. Differentiation 74, 422–437. https://doi.org/10.1111/j.1432-0436.2006.00096.x.
8. Que, J., Okubo, T., Goldentzer, J.R., Nam, K.T., Kurotani, R., Morrissey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L.M. (2007). Multiple dose-dependent roles for Sox2 in the
patterning and differentiation of anterior foregut endoderm. Development 134, 2521–2531. https://doi.org/10.1242/dev.003855.

9. Rankin, S.A., Han, L., McCracken, K.W., Kenny, A.P., Anglin, C.T., Grigg, E.A., Crawford, C.M., Wells, J.M., Shannon, J.M., and Zorn, A.M. (2016). A retinoic acid-hedgehog cascade coordinates mesoderm-inducing signals and endoderm competence during lung specification. Cell Rep. 16, 66–78. https://doi.org/10.1016/j.celrep.2016.05.060.

10. Gordon, C.T., Boute-Benejean, O., Caumes, R., Delobel, B., Dietruch, K., Gaillard, D., Gonzales, M., Lacombe, D., Escande, F., Manouvrier-Hamou, S., et al. (2012). EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia. J. Med. Genet. 49, 737–746. https://doi.org/10.1136/jmedgenet-2012-101173.

11. Wang, J., Ahimaz, P.R., Hashemifar, S., Khlevner, J., Piccaro, J.A., Middlesworth, W., Elfiky, M.M., Que, J., Shen, Y., and Chung, W.K. (2021). Novel candidate genes in esophageal atresia/tracheoesophageal fistula identified by exome sequencing. Eur. J. Hum. Genet. 29, 122–130. https://doi.org/10.1038/s41431-020-0680-2.

12. Nasr, T., Trisno, S.L., Cha, S.W., Wells, J.M., Kofron, M.J., Zorn, A.M., Mancini, P., Rankin, S.A., Edwards, N.A., Agricola, Z.N., et al. (2019). Endosome-mediated epithelial remodeling downstream of hedgehog-gli is required for tracheoesophageal separation. Dev. Cell 51, 665–674.e6. https://doi.org/10.1016/j.devcel.2019.11.003.

13. Qi, H., Lim, F.Y., Crombleholme, T., Cusick, R., Azarow, K., Danko, M.E., Chung, D., Warner, B.W., Mychaliska, G.B., Pöttoka, D., et al. (2018). De novo variants in congenital diaphragmatic hernia identify MYRF as a new syndrome and reveal genetic overlaps with other developmental disorders. PLoS Genet. 14, e1007822. https://doi.org/10.1371/journal.pgen.1007822.

14. Richter, F., Parfenov, M., Homsy, J., Gorham, J.M., Manheim, K.B., Velinder, M., Farrell, A., Marth, G., Schadt, E.E., Kaltman, J.R., et al. (2020). Genomic analyses implicate noncoding de novo variants in congenital heart disease. Nat. Genet. 52, 769–777. https://doi.org/10.1038/s41588-020-0652-z.

15. Poppin, R., Gross, S.S., Dorfman, L., McLean, C.Y., DePristo, M.A., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nat. Biotechnol. 36, 983–987. https://doi.org/10.1038/nbt.4235.

16. Karczewski, K.J., Francioli, L.C., MacArthur, D.G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. https://doi.org/10.1038/s41586-020-2632-1.

17. Lek, M., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291. https://doi.org/10.1038/nature19057.

18. Jaganathan, K., Chow, E.D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S.J., Farh, K.K.H., Kyriazopoulou Papaniokopoulou, S., McRae, J.E., et al. (2019). Predicting splicing from primary sequence with deep learning. Cell 176, 535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015.

19. Qiao, L., Nees, S., Khlevner, J., Lim, F.Y., Crombleholme, T., Cusick, R., Azarow, K., Danko, M.E., Chung, D., Warner, B.W., et al. (2020). Likely damaging de novo variants in congenital diaphragmatic hernia patients are associated with worse clinical outcomes. Genet. Med. 22, 2020–2028. https://doi.org/10.1038/s41436-020-0998-0.

20. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21, 974–984. https://doi.org/10.1101/gr.114876.110.

21. Layer, R.M., Chiang, C., Qinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 15, R84. https://doi.org/10.1186/gb-2014-15-6-r84.

22. Chiang, C., Layer, R.M., Faust, G.G., Lindberg, M.R., Rose, D.B., Garrison, E.P., Marth, G.T., Qinlan, A.R., and Hall, I.M. (2015). SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat. Methods 12, 966–968. https://doi.org/10.1038/nmeth.3505.

23. Teschendorff, A.E., Zhu, T., Breeze, C.E., and Beck, S. (2020). EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. Genome Biol. 21, 221. https://doi.org/10.1186/s13059-020-02126-9.

24. Rehm, H.L., Plon, S.E., Sherry, S.T., Watson, M.S., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Lundrum, M.J., et al. (2015). ClinGen—the clinical genome Resource. N. Engl. J. Med. 372, 2235–2242. https://doi.org/10.1056/nejmrs1406261.

25. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpora, M., Ragain, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am. J. Hum. Genet. 84, 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010.

26. Samocha, K.E., Wall, D.P., MacArthur, D.G., Gabriel, S.B., DePristo, M., Purcell, S.M., Palotie, A., Boerwinkle, E., Buxbaum, J.D., Cook, E.H., et al. (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. 46, 944–950. https://doi.org/10.1038/ng.3050.

27. Ware, J.S., Samocha, K.E., Homsy, J., and Daly, M.J. (2015). Interpreting de novo Variation in Human Disease Using denovolyzerR. Curr. Protoc. Hum. Genet. 87, 7.25.1–7.25.15. https://doi.org/10.1002/0471142905.hg0725s87.

28. Mootha, V.K., Houdist, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34, 267–273. https://doi.org/10.1038/ng1180.

29. Subramanian, A., Mesirov, J.P., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. https://doi.org/10.1073/pnas.0506850102.

30. Szklarczyk, D., Jensen, L.J., Mering, C., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607–D613. https://doi.org/10.1093/nar/gky1131.
31. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303.

32. Lane, M., and Khokha, M.K. (2021). Obtaining Xenopus tropicalis embryos by in vitro fertilization. Cold Spring Harb. Protoc. 2022. pdb.prot106351. https://doi.org/10.1101/pdb.prot106351.

33. Lane, M., and Khokha, M.K. (2021). Obtaining Xenopus tropicalis embryos by natural mating. Cold Spring Harb. Protoc. 2022. pdb.prot106609. https://doi.org/10.1101/pdb.prot106609.

34. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscans: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods 12, 982–988. https://doi.org/10.1038/nmeth.3543.

35. Fortriede, J.D., Lotay, V.S., Ponferrada, V.G., Karimi, K., Zorn, A.M., Vize, P.D., Pells, T.J., Chu, S., Chaturvedi, P., Wang, D., et al. (2020). Xenbase: deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. Nucleic Acids Res. 48, D776–D782. https://doi.org/10.1093/nar/gkj933.

36. Hsiau, T., Conant, D., Rossi, N., Waite, K., Yang, J., Hsiau, T., Conant, D., Rossi, N., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., Enzmann, B.L., et al. (2019). Inference of CRISPR edits from sanger trace data. Preprint at bioRxiv. https://doi.org/10.1101/251082.

37. Satterstrom, F.K., Stevens, C., Reichert, J., Mulhem, M.S., Artonov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 180, 568–584. https://doi.org/10.1016/j.cell.2019.12.036.

38. Richards, S., Voelkerding, K., Rehm, H.L., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the association for molecular pathology. Genet. Med. 17, 405–424. https://doi.org/10.1038/gim.2015.30.

39. Fakhro, K.A., Choi, M., Ware, S.M., Belmont, J.W., Towbin, J.A., Lifton, R.P., Khokha, M.K., and Brueckner, M. (2011). Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. Proc. Natl. Acad. Sci. U S A 108, 2915–2920. https://doi.org/10.1073/pnas.1019645108.

40. Willsley, H.R., Kim, A., Anderson, A.S., Shin, D., Seyler, M., Nowakowski, T.J., Harland, R.M., Willsley, A.J., State, M.W., Exner, C.R.T., et al. (2021). Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. Neuron 109, 788–804.e8. https://doi.org/10.1016/j.neuron.2021.01.002.

41. Bhattacharya, D., Marfo, C.A., Li, D., Lane, M., and Khokha, M.K. (2015). CRISPR/Cas9: an inexpensive, efficient loss of function tool to screen human disease genes in Xenopus. Dev. Biol. 408, 196–204. https://doi.org/10.1016/j.ydbio.2015.11.003.

42. Trisno, S.L., Mandegar, M.A., Wells, S.I., Zorn, A.M., Wells, J.M., Philo, K.E.D., McCracken, K.W., Cata, E.M., Ruiz-Torres, S., Rankin, S.A., et al. (2018). Esophageal organoids from human pluripotent stem cells delineate Sox2 functions during esophageal specification. Cell Stem Cell 23, S01–S15.e7. https://doi.org/10.1016/j.stem.2018.08.008.

43. Grant, B.D., and Donaldson, J.G. (2009). Pathways and mechanisms of endocytic recycling. Nat. Rev. Mol. Cell Biol. 10, 597–608. https://doi.org/10.1038/nrm2755.

44. Naslavsky, N., and Caplan, S. (2018). The enigmatic endosome-sorting the ins and outs of endocytic trafficking. J. Cell Sci. 131, jcs216499. https://doi.org/10.1242/jcs.216499.

45. Serra, N.D., and Sundaram, M.V. (2021). Transcytosis in the development and morphogenesis of epithelial tissues. EMBO J. 40, e101663. https://doi.org/10.15252/embj.2020106163.

46. Lee, J.Y., and Harland, R.M. (2010). Endocytosis is required for efficient apical constriction during Xenopus gastrulation. Curr. Biol. 20, 253–258. https://doi.org/10.1016/j.cub.2009.12.021.

47. Bruser, L., and Bogdan, S. (2017). Adherens junctions on the move-membrane trafficking of E-cadherin. Cold Spring Harb. Perspect. Biol. 9, a029140. https://doi.org/10.1101/cshperspect.a029140.

48. Mathew, R., Rios-Barrera, L.D., Machado, P., Schwab, Y., and Leptin, M. (2020). Transcytosis via the late endocytic pathway as a cell morphogenetic mechanism. EMBO J. 39, e105332. https://doi.org/10.15252/embj.2020105332.

49. Kowalczyk, I., Lee, C., Schuster, E., Hoeren, J., Riedel, L., Gorne, J., Wallingford, J.B., Hammes, A., and Feistel, K. (2021). Neural tube closure requires the endocytic receptor Lrp2 and its functional interaction with intracellular scaffolds. Development 148, dev195008.

50. Yoon, J., Garo, J., Lee, M., Sun, J., Hwang, Y.S., and Daar, I.O. (2021). Rab11fip5 regulates telencephalon development via ephrinb1 recycling. Development 148, dev196527.

51. Hussain, N.K., Yamabhai, M., Ramjaun, A.R., Guy, A.M., Barnes, D., O’Bryan, J.P., Der, C.J., Kay, B.K., and McPherson, P.S. (1999). Splice variants of intersectin are components of the endocytic machinery in neurons and nonneuronal cells. J. Biol. Chem. 274, 15671–15677. https://doi.org/10.1074/jbc.274.22.15671.

52. Yu, Y., Chiu, P.Y., Bowser, D.N., Keating, D.J., Dubach, D., Harper, I., Tkalcic, J., Finkelstein, D.L., and Pritchard, M.A. (2008). Mice deficient for the chromosome 21 or- tholog Itsn1 exhibit vesicle-trafficking abnormalities. Hum. Mol. Genet. 17, 3281–3290. https://doi.org/10.1093/hmg/ddn224.

53. Gryaznova, T., Gubar, O., Burdyniuk, M., Kropyvko, S., and Rynditch, A. (2018). WIP/ITSN1 complex is involved in cellular vesicle trafficking and formation of filopodia-like protrusions. Gene 674, 49–56. https://doi.org/10.1016/j.gene.2018.06.078.