A functional operon delineates an extracellular pathway that controls body asymmetry only in animals with a ciliated left-right organizer

Emmanuelle SZENKER-RAVI[1,2,$,*], Tim OTT[3,$], Muznah KHATOO[1,2], Anne MOREAU DE BELLAING[4], Wei Xuan GOH[1], Yan Ling CHONG[5,30], Anja BECKERS[6,7], Darshini KANNESAN[1], Guillaume LOUVEL[8], Priyanka ANUJAN[5,31], Vydianathan RAVI[5], Carine BONNARD[1], Sébastien MOUTTON[9], Patric SCHOEN[10], Mélanie FRADIN[11], Estelle COLIN[12], André MEGARBANE[13,14], Linda DAOU[15], Ghassan CHEHAB[16], Sylvie DI FILIPPO[17], Caroline ROORYCK-THAMBO[18], Jean-François DELEUZE[19], Anne BOLAND[19], Nicolas ARRIBARD[20], Rukiye EKER[21], Sumanty TOHARI[5], Alvin Yu Jin NG[5], Marlène RIO[22], Chun Teck LIM[23], Birgit EISENHABER[23], Frank EISENHABER[23], Byrappa VENKATESH[5,24], Jeanne AMIEL[22,25], Hugues ROEST CROLLIUS[8], Christopher T. GORDON[25], Achim GOSSLER[6,7], Sudipto ROY[5,26], Tania ATTIE-BITACH[22,27], Martin BLUM[3,*], Patrice BOUVAGNET[28,*] and Bruno REVERSADE[1,2,5,24,29,*]

1. Institute of Medical Biology (IMB), A*STAR, Singapore
2. Genome Institute of Singapore (GIS), A*STAR, Singapore
3. University of Hohenheim, Garbenstr. 30, 70599 Stuttgart, Germany
4. Laboratoire de Cardiogénétique, Groupe Hospitalier Est, Hospices Civils de Lyon, Lyon, France
5. Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore
6. Institute for Molecular Biology, OE5250, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany
7. REBIRTH Cluster of Excellence, Hannover, Germany
8. Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, PSL Research University, 75005, Paris, France
9. Service de Génétique, CHU Dijon, 14, rue Paul Gaffarel, 21079, Dijon, France
10. Praxis Dr Patric SCHÖN, Feierabendstrasse 53, 85764 OBERSCHLEISSHEIM, Germany
11. Service de Génétique Médicale, Hôpital Sud, CHU de Rennes, 35203, Rennes cedex, France
12. Service de Génétique, CHU Angers, 4, rue Larrey, 499933 Angers, France
13. CEMEDIPP-Centre Médico-Psychopédagogique, Beirut, Lebanon
14. Institut Jérôme LEJEUNE, Paris, France
15. Unité de Cardiologie Pédiatrique, Hôtel Dieu de France, Boulevard Alfred Naccache, Achrafieh, Beirut, Lebanon
16. Department of Pediatrics, Lebanese University, Faculty of Medical Sciences, Hadath, Greater Beirut, Lebanon
17. Service de Cardiologie Pédiatrique, Groupe Hospitalier Est, Hospices Civils de Lyon, 69677 Bron, France
18. Service de Génétique, CHU de Bordeaux, place Amélie Rabba-Léon, 33076, Bordeaux cedex, France
19. Centre National de Génotypage, Institut de Génomique, Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Evry, France
20. Service de Cardiologie Pédiatrique, Hôpital Universitaire des Enfants Reine Fabiola (HUDERF), Brussels, Belgium
21. Pediatric Cardiology Division, Pediatrics Department, Istanbul Medical Faculty, Istanbul University, 34093 Istanbul, Turkey

22. Fédération de Génétique, Hôpital Necker-Enfants Malades, Assistance Publique Hôpitaux de Paris, 75015 Paris, France
23. Bioinformatics Institute (BII), A*STAR, Singapore
24. Department of Paediatrics, National University of Singapore (NUS), Singapore
25. Université de Paris, Imagine Institute, Laboratory of Embryology and Genetics of Malformations, INSERM UMR 1163, 75015 Paris, France
26. Department of Biological Sciences, National University of Singapore (NUS), Singapore
27. Université de Paris, Imagine Institute, Laboratory of Genetics and development of the cerebral cortex, INSERM UMR 1163, 75015 Paris, France
28. CPDPN, Hôpital MFME, CHU de Martinique, BP632, 97200 Fort de France, France
29. Medical Genetics Department, Koç University School of Medicine (KUSOM), Istanbul, Turkey
30. Present address: Department of Pathology, National University Hospital, Singapore
31. Present address: Institute of Reproductive and Developmental Biology, Hammersmith Hospital, Imperial College, London, UK

$ These authors contributed equally to this work

* Authors to whom correspondence should be addressed:

Emmanuelle Szenker-Ravi (emmanuelle.szenker@reversade.com)
Martin Blum (martin.blum@uni-hohenheim.de)
Patrice Bouvagnet (patrice.bouvagnet@chu-martinique.fr)
Bruno Reversade (bruno@reversade.com)

Abstract

In deuterostomes, the left-right (LR) axis is specified during embryogenesis by a transient organ known as the left-right organizer (LRO). Most vertebrate animals including fish, amphibians, rodents and humans deploy motile cilia in the LRO to break bilateral symmetry thanks to a cilia-driven leftward flow in the extracellular space. Other vertebrates such as reptiles, birds, even-toed mammals and cetaceans have non-ciliated LROs and employ alternative mechanisms for LR patterning. In an attempt to identify genes whose loss during vertebrate evolution follows a similar pattern, we delineated a list of five extracellular proteins, of which two were hitherto unknown. We focused on a novel metalloproteinase, herein named TOUT-DE-TRAVERS (TDT), which is specifically expressed in ciliated LROs. Using CRISPR/Cas9 genome-editing in zebrafish and *Xenopus*, we demonstrate that TDT is solely required on the left side of the LRO to control organ laterality. Together with MMP21, another protease, TDT constitutes part of a proteolytic cascade that is downstream of leftward flow but upstream of *DAND5*, the first asymmetrically expressed gene. This pathway is clinically relevant, as we report 21 human patients with loss-of-function *TDT* mutations presenting with recessive *situs* anomalies with congenital heart defects. Our findings posit the existence of a functional operon that is instrumental for distinguishing left from right specifically in species that rely on motile cilia in the LRO.

Introduction

The three embryonic axes of vertebrate embryos are established around gastrulation. This involves the transduction of conserved signalling pathways initiated by extracellular ligands: the Anterior-Posterior (AP) axis is mediated by the canonical WNT pathway, the Dorsal-Ventral (DV) axis is driven by the BMP pathway, while the Left-Right (LR) axis is under the control of the NODAL signaling cascade[1–7]. Bilateral symmetry is broken by the directional right-to-left flow of extracellular fluids across the ciliated epithelium of the LRO, resulting in the asymmetric expression of conserved genes in the LRO and in the lateral plate mesoderm (LPM).

The current model of flow-dependent symmetry breaking at the ciliated vertebrate LRO involves several well-defined steps, which have been experimentally and genetically defined in the various model organisms, particularly zebrafish, *Xenopus* and mouse[7–9]. Following canonical WNT-dependent specification of the LRO precursor tissue in superficial organizer cells during gastrulation, which is marked by *Foxj1*, the LRO forms as a single-layer epithelium in the dorsal midline of the archenteron (amphibians) or remnants thereof (Kupffer's vesicle in zebrafish and ventral node in mouse). Cilia on central LRO cells become motile and polarize to the posterior pole of cells in a Wnt/PCP-dependent manner. In *Xenopus* and mouse, the LRO in fact constitutes the widened posterior part of the notochordal plate, which at this stage is exposed to the archenteron lumen and flanked by endodermal cells. Lateral LRO cells harbor immotile and non-polarized cilia which express both NODAL/GDF1 as well as the NODAL inhibitor DAND5. A right-to-left flow of extracellular fluids in the immediate vicinity of the apical cell surface of central LRO cells sets in during early neurula stages. This leftward flow is then sensed by immotile cilia on lateral LRO cells in a poorly defined process that involves the ion channel PKD2, which forms a complex with PKD1L1. As a consequence of flow-sensing, *Dand5* mRNA gets reduced specifically on the left side, relieving the repression of NODAL, which signals and transfers to the left LPM. There, NODAL induces the transcription of the so-called NODAL cascade

genes, *NODAL* itself, its feedback inhibitor *Lefty* as well as the homeobox transcription factor *Pitx2*. Together, these genes determine proper asymmetric organ morphogenesis and placement. This evolutionary-conserved program ultimately orchestrates the correct positioning of the heart and other visceral organs such as the lungs, spleen and stomach i.e., *situs solitus*[10]. Defects in these processes cause heterotaxy, the abnormal formation and arrangement of organs across the left-right axis, that can range from complete inversion of symmetry (*situs inversus totalis*) to the selective misarrangement of organs (*situs ambiguus*). Heterotaxy is often associated with complex congenital heart defects (CHDs)[11]. Laterality defects have an estimated prevalence of 1.1/10 000 live births in humans[12] and account for approximately 3% of all CHDs[11]. *Situs inversus totalis* can also occur in the context of syndromic ciliopathies such as the primary ciliary dyskinesia syndrome, a sinopulmonary disease (PCD; MIM244400) caused by mutations in structural components of motile cilia[13,14].

Genes implicated in heterotaxy (HTX) in humans include components of the NODAL pathway such as the ligands themselves, namely *NODAL* (HTX5; MIM601265)[15] and *GDF1* (CHTD6; MIM613854)[16], their cognate cell surface receptor *ACVR2B* (HTX4, MIM602730)[17] and co-receptor *CFC1* (HTX2, MIM605194)[18], and the dedicated extracellular modulators *LEFTY2*[19] and *DAND5*[20]. Other factors not directly related to the NODAL pathway have also been reported to be mutated in Mendelian forms of heterotaxy[21]. These include structural components of the motile cilia including *CCDC11* (HTX6; MIM614779)[22,23], signaling components that function on immotile cilia that sense motile cilia-induced fluid flow such as *PKD2* (PKD2; MIM613095) and *PKD1L1* (HTX8; MIM616749)[24], as well as downstream effectors including the transcription factor *ZIC3* (HTX1; MIM306955)[25]. More recently, recessive mutations in the gene coding for the Matrix MetalloProtease 21 (*MMP21)* have also been reported to cause heterotaxy (HTX7; MIM616749)[26–28], but how this metalloprotease fits into the NODAL pathway has remained enigmatic.

In this study, we identify a group of genes encoding extracellular proteins that directly contribute to establishing the LR axis in animal species that have kept cilia in LROs. Through

142 phylogenetic analyses, we uncovered a hitherto uncharacterized metalloprotease, which we
143 named TOUT-DE-TRAVERS (TDT, from the French expression meaning “all upside down”).
144 Like *MMP21*, and three other genes namely *PKD1L1*, *DAND5* and *ALENDROIT* (*ALED*,
145 a.k.a. *C1orf127*), we find that *TDT* is specifically and solely expressed in the LRO in
146 zebrafish, *Xenopus* and mice. Genetic inactivation of Tdt results in fully penetrant *situs*
147 anomalies, downstream of leftward flow, but upstream of DAND5-mediated control of
148 NODAL propagation to the left LPM. Through rescue, epistasis and biochemical
149 experiments, we propose that TDT, MMP21, ALED, PKD1L1, and DAND5 form a module
150 with interdependent function in LR axis specification. Because their requirements are
151 interlocked and strictly limited to this early developmental process, this “functional” operon
152 has twice been lost during evolution of the vertebrates, first in the ancestor of sauropsids
153 (birds/reptiles) and then once again in even-toed mammals and cetaceans (cetartiodactyla).

154

Results

At least 5 extracellular proteins have been lost in species without a ciliated LRO

MMP21 has been shown to have a singular pattern of gene loss across vertebrate species[27]. While it is present in humans and mice, its homologues in cetartiodactyla have acquired biallelic loss-of-function mutations, indicating that no selective pressure exists to maintain *MMP21* in these mammalian species (**Figure S1**). C. Gordon and colleagues have proposed that early vertebrates, including fish and amphibians, possessed *MMP21*, but that the gene had been specifically lost in all bird and reptile species (**Figure S1**)[27]. This peculiar and repeated *MMP21* gene disappearance/inactivation across diverse phyla appears to coincide with the loss of motile cilia in central cells of the LRO in animals that utilize asymmetric cell movement rather than cilia-driven flow to impart LR polarity to their developing embryos (**Figure 1a,b**)[29,30].

Based on this initial observation, we hypothesized that *MMP21* must represent, and belong to, a larger group of genes which, together, constitute a pathway specifically needed for motile cilia-dependent LR patterning. We thus searched for genes that, like *MMP21*, are present in fish, frogs, primates, rodents and odd-toed ungulates but are absent or inactivated in birds, reptiles and even-toed mammals (**Figure 1a**). In order to focus our attention on extracellular proteins, which may include possible MMP21 substrates, we restricted our query to protein-coding genes that are annotated to be containing a signal peptide targeted for cell surface retention or secretion (**Figure 1 and Figure S1**). This search produced 4 additional genes (**Figure 1b**). Two of these, *PKD1L1* and *DAND5* are already known to be needed for LR patterning in various model organisms, but had not been previously reported to have been lost in birds, reptiles and cetartiodactyla. Two novel genes were identified, both encoding secreted proteins, which we named Tout-de-Travers (*TDT*) meaning “all upside down”, and Alendroit (*ALED*) which means “right side out” in French. Like *MMP21*, *TDT* and *ALED* are both present in fish, amphibians and odd-toed mammals including rodents and

humans, but are absent or mutated in birds, reptiles and cetartiodactyla (**Figure 1a,b and Figure S1**).

TDT codes for a Zn^{2+}-dependent metalloproteinase belonging to the M8 peptidase family of leishmanolysin-like metzincins (a.k.a. LMLN2). As no more than 2 ESTs exist for this gene, the exon-intron structure of *TDT* is incompletely annotated in UCSC, Ensembl or NCBI. We determined that *TDT* in fact spans 14 exons, and consists of 4 domains: a signal peptide, a zinc catalytic domain containing the consensus His-Glu-x-x-His...His (HExxH...H) active site motif and a Met-turn methionine, a cysteine-rich domain and a C-terminal transmembrane domain (**Figure 1c and Figure S2**). Transient expression of human or zebrafish *Tdt* in HEK293T cells showed that the encoded protease is membrane-bound and does not involve a GPI-anchor (**Figure S3a**). This is in contrast to the TDT homolog GP63 in the *Leishmania* promastigote, which is GPI-anchored[31]. The zebrafish TDT protein was readily secreted and is heavily post-translationally modified by sugar moieties which can be removed by PNGase F and to a lesser extent by Endo H treatment (**Figure S3b**). Upon DSS-crosslinking, TDT assembled into a high-molecular weight oligomer consistent with a discrete quaternary structure in solution (**Figure S3c**).

ALED, currently annotated as *C1orf127*, encodes a protein with a signal peptide, a DUF4556 domain, and a C-terminal proline-rich domain. A rewarding validation of the legitimacy of our evo-devo screen is the report that an ENU-induced splice mutant c.26+1G>A in the mouse homologue of *ALED* (NM_001085505, Gm572) displays heterotaxy when bred to homozygosity (MGI Direct Data Submission, ID:175213). Moreover, a single case of heterotaxy in humans has recently been associated with a homozygous p.Arg113* nonsense *ALED* mutation[32]. These results encouraged us to focus our attention on *TDT*, given that no publication has ever mentioned its existence, let alone a possible role in LR specification.

Tdt is specifically expressed in the LRO in zebrafish, Xenopus and mouse embryos

We hypothesized that if *TDT* is to play a role during LR patterning, its developmental expression should coincide with the formation and function of the LR organizer. Using whole

mount *in situ* hybridization, we examined its spatio-temporal expression in three distinct vertebrate species that possess a ciliated LRO.

In zebrafish, *tdt* was found to be specifically and solely expressed from 70% epiboly to the 3-somite stage in dorsal forerunner cells (DFCs), which coalesce to form the Kupffer's vesicle (KV), the zebrafish LRO (**Figure 2a and Figure S4a**). By double fluorescent *in situ* hybridization, *tdt* mRNA was confirmed to colocalize with that of the DFC marker *sox17*[33] (**Figure 2b**). *Tdt* expression in the DFCs was found to be regulated by Foxj1a, a master regulator of the motile ciliogenetic program[34,35]. Indeed, *foxj1a* depletion by morpholino oligomer injections[34] completely abolished *tdt* expression in the DFCs (**Figure 2c**), although ubiquitous *foxj1a* overexpression using a heat shock-inducible transgenic *hs::foxj1a* line[36], did not result in ectopic *tdt* expression (**Figure 2c**). This reveals that *foxj1a* is essential, but is not sufficient, to control endogenous *tdt* expression in zebrafish DFCs and KV precursors. We next assessed whether *Xenopus laevis,* an amphibian species which also possesses a homologous *tdt* gene, would similarly express it during gastrulation. *In situ* hybridization revealed a restricted expression pattern initiated after the mid-blastula transition at the commencement of zygotic transcription. *Tdt* was first expressed at stage 10.5 in the involuting marginal zone (IMZ) along a dorsal-high ventral-low gradient (**Figure 2d**). Dorsal IMZ cells include the LRO precursors, which are internalized during gastrulation with the endo- and mesodermal cell mass. Early neurulae (stage 14) showed strong *tdt* expression in cells of the circumblastoporal collar (CBC), a structure through which involuting cells pass as they enter the LRO (**Figure 2d**). Cellular transition from the CBC into the LRO is accompanied by a dampened *tdt* signal. This spatiotemporal expression is analogous to what is observed in zebrafish, and the expression in the *Xenopus* CBC closely resembles that of *bicc1* which is known to be involved in LR patterning in *Xenopus* embryos[37].

Lastly, we performed an *in situ* hybridization in an odd-toed mammalian species, *Mus musculus,* which based on our evo-devo analysis should also display LRO-specific *Tdt* expression. The analysis of mouse embryos across gastrulation/early neurulation stages revealed strong *Tdt* mRNA levels in the ventral node from the early bud to the 4-somite stage

(**Figure 2e**). Endogenous *Tdt* expression was enriched in the peri-nodal crown cells and became asymmetric at the 3-somite stage, suggesting a downregulation in the prospective left side in response to leftward flow (**Figure 2e**). Notably, *Tdt* expression was entirely abolished in mutant embryos lacking *Noto*, a transcription factor required for proper LRO formation upstream of FOXJ1[38] (**Figure 2f**). Altogether, these expression analyses in three distinct animal models with functional ciliated LROs revealed that *Tdt* was specifically transcribed under the control of the transcription factors FOXJ1 and NOTO. This singular spatio-temporal regulation is consistent with TDT playing a role in the process of left-right patterning during embryonic development of these species. Reptiles, birds or even-toed mammals could not be examined for expression of *Tdt* homologues as these species have either lost the gene altogether or have accumulated deleterious mutations, suggestive of an ongoing process of gene deactivation (**Figure 1 and Figure S1**).

Knockout of tdt in zebrafish and Xenopus results in organ laterality defects

A partial sequence of zebrafish *tdt* can be found in the EST XM_002662823, but it lacks a proper signal peptide sequence. We noted the presence of a possible upstream start codon, which would lead to the transcription of a full-length mRNA with an additional 168 nucleotides (**Figure S4b,** light gray). We used CRISPR/Cas9-mediated gene editing to induce frameshift mutations, upstream of the catalytic domain of Tdt in order to generate knockout zebrafish lines (**Figure S4b**). Five germline alleles with distinct frameshift mutations all leading to early stop codons were obtained, outcrossed and bred to homozygosity. Despite the absence of measurable nonsense-mediated decay (NMD) of the mutant *tdt* transcripts (**Figure S4c**), all homozygous *tdt* knockout (KO) fish presented with a classical phenotype of heterotaxy (**Figure 3a-c**). Zygotic (Z) null *tdt*$^{-/-}$ fish were fertile, allowing us to generate 100% maternal zygotic (MZ) *tdt*$^{-/-}$ fish which had normal viability. The developing heart is one of the first organs to undergo asymmetric LR morphogenesis. While control embryos showed a heart tube looping to the right (D-looping) at 48 hpf, Z and MZ *tdt*$^{-/-}$ embryos showed cardiac looping randomization by visual inspection or using the *Tg(myl7:eGFP)* transgenic reporter

which marks cardiomyocytes[39] (**Figure 3a**). To confirm that these laterality defects seen in half of *tdt*$^{-/-}$ embryos were not the consequence of Cas9 off-target effects, we performed rescue experiments. The injection of 250 pg of wildtype *tdt* mRNA from human or zebrafish did not induce any notable phenotype in control embryos (**Figure S5a-c and data not shown**), but was sufficient to rescue the percentage of normal cardiac looping in 87% in MZ *tdt*$^{-/-}$ embryos (**Figure 3a**). Interestingly, injection of a catalytic-dead (CD) *tdt* mRNA, harboring three missense mutations in its catalytic site (HExxH→AAxxA, p.His247Ala, p.Glu248Ala and p.His251Ala) was unable to rescue the heart looping defects in larvae bereft of *tdt* (**Figure 3a**). These results demonstrate the specificity of the observed phenotype, highlight the importance of the Zn^{2+}-binding site for the catalytic activity of Tdt, and suggest that generalized overexpression of *tdt* is not sufficient to cause LR anomalies. Triple *in situ* hybridization allowed us to better assess the LR asymmetry defects in embryos lacking *tdt*. Brain asymmetry is observed by expression of the *leftover (lov)* gene in the left habenulae, the leftward heart looping can be scored by *myl7* expression, while the correct positioning of the pancreas to the right side is marked by *insulin (ins)* expression (**Figure 3b**). Only 25% of MZ *tdt*$^{-/-}$ embryos show a normal positioning of these three organs (*situs solitus*), while more than half displayed defects in the position of at least one organ (*situs ambiguus*). The remaining 25% presented with a complete reversal of organ positioning (*situs inversus*) (**Figure 3b**). Interestingly, no other ciliopathy-related phenotypes such as body curvature, kidney cysts or hydrocephalus were observed in *tdt*$^{-/-}$ mutant embryos at 72 hpf (**Figure 3c**), or in the MZ null adults that are fully viable and fertile. These results are consistent with the exclusive expression of *tdt* in the KV, suggesting a highly specialized function of Tdt during a narrow window of time for LR asymmetry establishment in this species.

We next sought to confirm the role of Tdt for LR patterning in *Xenopus* embryos. Bilateral co-injection of *tdt* CRISPR gRNA with Cas9 protein (**Figure S4d**) resulted in crispants presenting with *situs* defects, as assessed by heart and intestine looping as well as the

position of the gallbladder (**Figure 3d**). Owing to holoblastic development, this amphibian species permits the assignment of gene activity requirements in left versus right lineages[40,41]. To determine whether Tdt is specifically required on the left or on the right side of the LRO, we generated *tdt* crispants by injecting either the left or right blastomeres at the 4-cell stage. Injections on the left side of the embryos resulted in similar penetrance of laterality defects, while no phenotypes were observed when *tdt* was inactivated specifically on the right side (**Figure 3d**). These lineage-restricted gene inactivations demonstrated that Tdt requirement for proper LR asymmetry patterning is restricted to the left side of the embryo. Similar to zebrafish mutants, *Xenopus* crispants did not present with any additional phenotype (**Figure 3e**), revealing the specificity of Tdt in the control of laterality. Altogether, these knockout studies performed in zebrafish and *Xenopus* firmly establish the role of Tdt for proper LR polarity in these two vertebrate species that possess cilia in their LRO.

Downregulation of dand5 rescues Nodal signaling defects in tdt mutant embryos

To get a better understanding of the *situs* anomalies observed in zebrafish embryos lacking *tdt*, earlier molecular markers of LR asymmetry were inspected. The heart looping defects documented using *myl7 in situ* hybridization at 48 hpf could be traced back to heart jogging irregularities already observable by 30 hpf (**Figure 4a**). The expression of the nodal-related gene *southpaw (spaw)* and its downstream targets *lefty1* and *lefty2*, whose asymmetric expression and involvement in vertebrate LR patterning are evolutionary conserved[42], can serve as early readouts for future organ laterality. In control embryos, s*paw* expression is initiated in the posterior side and remains expressed in the left LPM at the 22-somite stage (**Figure 4b**). Interestingly, bilateral, absent or posterior-restricted expression of *spaw* was observed in MZ *tdt*$^{-/-}$ mutants, suggestive of defective asymmetric nodal signaling (**Figure 4b**). The integrity of the dorsal midline, which serves as a barrier to prevent diffusion of asymmetric signals between the left and right sides[42], was intact as judged by uninterrupted *tbxta* expression at the midline (**Figure S5d**). Expression of the Nodal target gene *lefty1* in the left diencephalon was completely abolished in the absence of Tdt (**Figure 4c**). Similarly,

316 the expression of *lefty2* in the left heart primordium was largely lost in *tdt* mutants (**Figure**
317 **4d**). The inability of *spaw* expression to propagate to the anterior left part of the embryos was
318 likely the cause for defective *lefty1* and *lefty2* expression, which in turn might account for the
319 observed randomization of brain asymmetry and heart looping, respectively.
320 Similar analyses were conducted in *Xenopus tdt* crispants. Early LR asymmetry defects
321 could be documented using *in situ* hybridization for *pitx2,* a homeobox protein that acts
322 downstream of Nodal and Lefty to induce situs-specific morphogenesis on the left side of the
323 embryo[43–46]. While *pitx2* was restricted to the left in control embryos, it was mostly lost or
324 misexpressed on the right side of *tdt* crispants (**Figure 4e**). This confirms that Tdt is
325 important for proper Nodal signaling, a view which was further supported by the observation
326 that the expression of *nodal1* in the left LPM was lost in *tdt* crispants (**Figure 4f**).
327 We next assessed whether Tdt acts downstream or upstream of Dand5, a key secreted
328 Nodal inhibitor which is downregulated on the left of wildtype embryos in response to the
329 leftward flow[47,48]. The MO-mediated knockdown of *dand5* on the right side of control embryos
330 effectively triggered bilateral *pitx2* expression and *situs* anomalies, consistent with the key
331 role of DAND5 in NODAL induction (**Figure 4e**), while left-sided MO-injections had no effect,
332 as described previously[47]. However, depletion of *dand5* on the left side of *tdt* crispants, i.e.
333 on the side in which Tdt is required, was sufficient to rescue the asymmetry defects observed
334 in *tdt* crispants (**Figure 4e**). This important result illustrates that derepressing Nodal
335 signalling by depleting endogenous Dand5 is able to compensate for the loss of Tdt activity.
336 Altogether, these data reveal that Tdt likely functions upstream of Dand5, where its role is to
337 enhance Nodal signalling for LR asymmetry establishment in vertebrate embryos that
338 possess a ciliated LRO.

339 ***Tdt is required for asymmetric dand5 expression but not for ciliogenesis***

340 In order to more precisely delineate how TDT acts vis-a-vis endogenous Nodal signaling, we
341 carefully recorded *dand5* expression in 8-somite stage zebrafish embryos. Compared to
342 controls, *dand5* expression in MZ *tdt*$^{-/-}$ zebrafish was not repressed on the left side (**Figure**

5a). This apparent altered response to KV fluid flow could either be the result of defective KV formation, improper ciliogenesis, erratic cilia-driven flow, or poor sensing of flow on the left side. Using immunofluorescence for acetylated alpha-tubulin and gamma-tubulin, we found that MZ *tdt*$^{-/-}$ embryos presented with a normal-sized KV containing motile cilia with normal beating patterns (**Movies S1 and S2**). Cilia length and numbers did not significantly differ from that of control embryos (**Figure 5b**). This suggests that ciliogenesis *per se* and fluid flow in the LRO are not appreciably affected by the absence of Tdt, as verified by the normal pattern of movement of endogenous particles that remain suspended in KV fluid (**Movies S3 and S4**). Furthermore, we found that *tdt* inactivation had no obvious consequence on DFC migration, compaction and KV formation as revealed by *sox17* staining (**Figure 2a-b and Figure S5e**). Likewise, Tdt was not required for *foxj1a* expression in the DFCs or KV (**Figure S5f**).

A more granular analysis in *Xenopus tdt* crispants revealed that the expression domains of the lateral LRO markers *nodal1* and *dand5* were significantly smaller in the absence of Tdt (**Figure 5c,d**). The asymmetric expression of *dand5,* which is normally repressed on the left side in response to Nodal signaling, was not recorded when embryos lacked functional Tdt (**Figure 5d**). Careful analyses of various LRO parameters such as cell surface areas, cilia length and polarization as well as flow velocity and directionality (following fluorescent bead application to LRO explants) revealed that no single parameter was statistically different between control and *tdt* crispants (**Figure 5e,f and Movie S5**). These results argue that Tdt, like in zebrafish, is also dispensable for ciliogenesis and cilia-driven flow in the LRO of *Xenopus* embryos. Altogether, these analyses in zebrafish and *Xenopus* place Tdt as a key regulator of LR development downstream of leftward flow but upstream of *dand5* repression.

Tdt acts upstream of Mmp21

To begin to analyze the relationship between MMP21 and TDT, we propagated and characterized several *mmp21* KO zebrafish lines (**Figure S6a**). As previously reported, 50% of zygotic and MZ null *mmp21* fish displayed randomization of their LR axis[26,27]. To test

whether these two genes work in parallel or epistatic pathways, we generated double heterozygous *tdt*$^{+/-}$;*mmp21*$^{+/-}$, *tdt*$^{-/-}$;*mmp21*$^{+/-}$, *tdt*$^{+/-}$;*mmp21*$^{-/-}$ and double KO *tdt*$^{-/-}$;*mmp21*$^{-/-}$ fish. Double heterozygous fish did not show any LR defects, and no discernible increase or decrease in penetrance or severity of heterotaxy was observed in double KO compared to the simple KO of *tdt* or *mmp21* alone (**Figure 6a**). This result suggests that when one of these two genes is mutated, the absence of the other may become irrelevant, and lends support to the notion that the two proteins might function in the same pathway rather than in parallel ones. Importantly, overexpression of *tdt* mRNA that was used to rescue *tdt*$^{-/-}$ mutant fish could not rescue the *situs* defects of *mmp21* mutants (**Figure 6b**), suggesting that Mmp21 most likely functions downstream of Tdt since it is required for overexpressed Tdt to be biologically active.

A careful analysis of the temporal transcription of these two proteases during zebrafish development revealed that they are sequentially turned on in the LRO, with *tdt* expression preceding that of *mmp21,* which persists several hours after *tdt* is silenced (**Figure 6c**). Using *in situ* hybridization and RT-qPCR, we found that the endogenous expression of *mmp21* was significantly increased in *tdt* KO fish (**Figure 6d,e and Figure S6b**), while the reverse was not true (data not shown). This phenomenon was independently confirmed in double KO *tdt*$^{-/-}$;*mmp21*$^{-/-}$ zebrafish where the NMD of *mmp21* mutant transcripts normally observed in single *mmp21* KO embryos was overcome by transcriptional upregulation when Tdt function was also absent (**Figure 6e**). All of these data suggest that a one-way transcriptional feedback operates between Tdt and Mmp21, and that they might be part of a linear epistatic pathway controlling LR specification. Taken together, our data suggest a working model for the action of TDT in LR axis formation (**Figure 6f**). TDT works in concert with at least 4 partner proteins DAND5, PKD1L1, ALED and MMP21 to form an extracellular module needed to sense and propagate the NODAL/GDF1 signalling cascade. This functional operon has been twice lost in vertebrates that have forgone the use of beating and sensing cilia in the LRO to distinguish left from right. In the ciliated LRO, TDT acts downstream of the flow which is sensed by PKD1L1. TDT is required on the left side to

ensure that *DAND5* expression is inhibited for asymmetric NODAL signalling to be propagated to the left lateral plate mesoderm. Together with MMP21 which acts after TDT, these two proteases form a proteolytic cascade downstream of the cilia-driven flow to prevent *situs* anomalies in fish, frogs and humans.

TDT is often mutated in patients with recessive heterotaxy

As TDT is required for proper left-right asymmetry patterning in zebrafish, *Xenopus* and presumably mice, we hypothesized that loss-of-function mutations in *TDT* could exist and account for cases of heterotaxy in humans. To test this, we screened for germline *TDT* mutations using targeted sequencing in a cohort of 186 index cases with congenital heart defects (CHDs) (**Table S1**). These patients either presented with heterotaxy (extra-cardiac and/or cardiac laterality defects such as dextrocardia, n=108), or without heterotaxy (including isolated tetralogy of Fallot or atrial septal defect (ASD), n=78). In this cohort, we identified germline homozygous *TDT* variations in 7 cases of CHD associated with heterotaxy from 7 different families (Families 1-7) (**Figure 7, Table 1 and Figure S7**). We then identified *TDT* mutations by whole-exome sequencing or targeted sequencing in five additional families segregating heterotaxy (Families 8-12) (**Figure 7, Table 1 and Figure S7**). All germline variations, homozygous patients from 11 families and compound heterozygous in family 12, were confirmed by Sanger sequencing in available family members, and found to fully segregate with the disease according to an autosomal recessive mode of inheritance (**Figure 7a and Table 1**). With these twelve families of distinct geographical origin, we report an allelic series of 9 *TDT* germline mutations in 21 patients with congenital heterotaxy (**Figure 7 and Table 1**). Among the nine germline variants identified, five were missense variants affecting highly conserved residues. Three of those, p.Cys296Ser (Family 8), p.Ser329Leu (Family 1) and p.Arg334Ile (Families 4 and 7) are within the catalytic domain of TDT which could affect the enzymatic activity of the protease.

The two others missense variants p.Ser31Phe (Families 2, 5, 11 and 12) and p.Leu414Pro (Family 6), as well as the deletion of a single conserved residue p.Phe401del (Family 9) are situated in other domains which may affect the half-life or quaternary structure of TDT (**Figure 7c,d and Table 1**). The three last variants are classical knockout alleles including two nonsense mutations p.Arg194* (Family 12) and p.Trp291* (Family 3) and a frameshift mutation p.Gly382Aspfs*8 (Family 10) leading to early stop codons, which we suspect will trigger NMD of mutant *TDT* transcripts **(Figure 7c,d and Table 1**). Of the 21 cases presenting with congenital heterotaxy, 8 had *situs solitus* with or without isolated dextrocardia, 5 had *situs ambiguus*, 8 had *situs inversus totalis*, and nearly all had congenital heart defects (CHD). As is the case for mutant *tdt* fish and frogs, there is no other ciliopathy-related phenotype arguing that TDT plays no other function beyond early LR specification (**Figure 7 and Table 1**). We also note that all homozygous carriers of *TDT* mutations were affected, suggesting a 100% penetrance, which is in contrast to the 25% of normal embryos observed in *tdt* mutant fish and frogs. We suspect that humans will be found to have *situs solitus* despite being mutated in *TDT*. This cohort of patients with heterotaxy curated by P. Bouvagnet had also been consulted to gauge how frequently *MMP21* is mutated[27]. Hence, we can estimate that *TDT* mutations account for up to 6.5% of CHD human patients with non-syndromic heterotaxy which is comparable to the 5.9% that was calculated for *MMP21*[27]. This clinical study lends strong support to the notion that biallelic loss of TDT in *Homo sapiens*, like in *Danio rerio* and *Xenopus laevis,* is a significant contributor for severe birth defects associated with heterotaxy. Taken together, our results lend credence to the notion that Tdt is essential for patterning the LR axis in these three vertebrate species that utilize motile cilia in the LRO.

Discussion

In this study, we identify, characterize and define the clinical importance of *TDT*, a gene which was never studied before. *TDT*, which encodes a 70 kDa metalloprotease, is poorly annotated in human genome and transcriptome databases due to its very restricted spatial

and temporal expression pattern. In an organism's entire lifespan, *TDT*'s transcription and physiological requirement appears to be limited to a finite process which only lasts a couple of hours during early embryogenesis. This unique feature has precluded obtaining sufficient ESTs to properly map its exon-intron borders. As such, *TDT* has not been properly covered by commercially-available exome platforms and has thus escaped clinical investigation, despite our data showing that it is frequently mutated in recessive forms of heterotaxy involving CHDs. By contrast, MMP21 which has a comparable temporal expression pattern and requirement for LR axis formation, is also expressed in other adult cell lineages including the skin[49–52], and its high expression has been associated with poor overall survival of patients with various forms of epithelial cancers[53,54]. To the best of our knowledge, *TDT* stands out as the single gene in LR patterning with no other function than the one ascribed here.

What are the substrates of TDT and MMP21?

Based on our phylogenetic analysis, we anticipate that the substrates of these two proteases might also be important for LR specification. We cannot exclude the possibility that, while being cleaved by MMP21 or TDT near the LRO, these substrates may have alternate roles beyond LR specification, and thus, continue to exist in some animals that do not use cilia in the LRO. Alternatively, their substrates might be part of the module that we have begun to delineate this far. We note that 4 of the 5 identified genes encode proteins which have been reported to be post-translationally cleaved. Beyond the necessary removal of the signal peptide, TDT likely harbors a prodomain with a conserved cysteine-switch which needs to be removed for full activity[55], PKD1L1 has been shown to be cleaved within the third intracellular loop of its transmembrane region[56], and DAND5, like its close homologue CERBERUS, may undergo proteolytic cleavage[57,58]. Future investigations will need to address if ALED is also cleaved or if other proteins not yet identified in this evo-devo screen might serve as substrates for MMP21 and TDT. Whatever these may be, we propose that an extracellular proteolytic cascade is likely at play during LR specification. This is reminiscent of another

evolutionary-conserved extracellular proteolytic cascade that imparts embryonic polarity. Patterning of the DV axis is mediated by the controlled cleavage of Chordin by the Zinc metalloprotease Tolloid which frees BMP to signal in ventral most regions of the embryo when Sizzled is silenced[59,60].

Evo-devo of LR specification: a functional operon twice disappears

Our phylogenetic screen for genes that have disappeared in vertebrate species as a consequence or cause of de-ciliation of the LRO yielded important evolutionary insights into the developmental programs of LR specification. Indeed, all 5 genes: *TDT*, *ALED*, *MMP21*, *DAND5* and *PKD1L1* were found by us, and others before, to be expressed in the very tissue and at the correct time to play a role in LR axis formation. These genes are not physically linked on the same chromosomal location, and as such escape the traditional definition of an operon[61]. However, they seem to be functionally interlocked in a transcriptionally coordinated group[62], which we propose might be referred to as a "functional operon" for LR patterning. The facultative nature of this "evo-devo module" is striking, since with enough time, all of these genes were lost in reptiles/birds or became pseudogenes in cetartiodactyla. Forming an integrated module for flow sensing and propagation of NODAL/GDF1 signalling downstream of cilia-driven flow, we suspect that like a bidirectional domino-effect, if any given gene becomes inactivated, all other 4 will too become obsolete. We find that this has happened not once but twice during vertebrate evolution. We suspect that this was allowed to happen because NODAL-dependent LR axis formation preceded cilia and flow-dependent symmetry breaking during animal evolution. Lophotrochozoans, such as snails, activate left-asymmetric Nodal signaling independently of cilia, but via chiral cleavage patterns. Such patterns can still be seen in amphibians such as *Xenopus*[63], but have become functionally obsolete during amniote evolution[30,64]. Upon loss of ciliated LROs, cytoskeletal asymmetries such as the ones underlying chiral cleavage may have become functional again, for example as seen in chiral cell migration at the chick Hensen's node[29].

This brings an important question as to whether birds/reptiles and even-toed mammals/cetaceans have found identical ways to cope with the loss of these genes or have instead deployed different strategies. If so, existing pathways that can provide buffering or functional redundancy might have been mobilized in both phyla. This paradigm is of immediate importance since we note that canines and felines (**Figure S1a**), have already begun to lose *ALED* but still harbor what appears to be normal genes for *TDT*, *MMP21, DAND5* and *PKD1L1*. This indicates that an ongoing forfeiture of this functional module might be occurring contemporarily. If so, this should allow to settle the corollary question as to which comes first. Is it the de-ciliation of the LRO that triggers the obsolescence of this functional operon, or is it the loss of one of these 5 genes that specifically turns off the program of ciliogenesis within the LRO? Observing the organizer of canine or feline embryos might help us answer this egg/chicken dilemma and give evolutionary biologists the opportunity to investigate how evo-devo traits become fixed or purged in coeval vertebrate species.

Methods

Phylo-genomic analysis

We scanned the 23904 phylogenetic trees of the Ensembl (version 93) Compara database to identify genes with a presence/absence pattern in extant species similar to the *MMP21* evolutionary profile. We restricted the trees to 75 species with minimal genome assembly contiguity (50% of scaffolds/chromosomes must contain at least 25 genes). To identify MMP21-like patterns, we searched for evolutionary histories from the Amniote ancestor that maximize retention in Euarchontoglires and Zoomata and losses in Artiofabula and Sauria using the following score: $S = P_{Eu} + P_{Zoo} - 3P_{Art} - 2P_{Sau}$. *Eu, Zoo, Art* and *Sau* represent respectively Euarchontoglires (primates and rodents, 39 species), Zoomata (horse and carnivores, 6 species), Artiofabula (even-toed ungulates excluding camelids, 4 species) and Sauria (birds and reptiles, 5 species). *P* represents the proportion of species in each clade that retained at least one copy of the gene in the clades noted in indices. We applied higher weights on gene losses in Artiofabula and Sauria because these are the distinctive features of the desired evolutionary pattern, although with less emphasis on losses in Sauria owing to the greater evolutionary distance to the other groups. All genes should also possess an ortholog in actinopterygian (ray-finned) fish to be considered further. Finally, we only considered genes with proteins containing a predicted N-terminal signal peptide using SignalP version 5.0[65]. Genes displaying a phylogenetic score > 1 and fulfilling these criteria are *MMP21*, *DAND5*, *TNFRSF14*, *PKD1L1*, *ALED (C1orf127)* and *TDT (LMLN2)* (**Table S2**). We considered *TNFRSF14* as false positive because despite its high phylogenetic score, it is present in the lizard genome. As nothing was known about *TDT,* we decided to focus our functional analysis on this gene.

We then specifically analyzed the presence or absence of the *ALED*, *DAND5*, *MMP21*, *PKD1L1* and *TDT* genes in various groups of animals (mammals, birds, reptiles, amphibians, ray-finned fishes, cartilaginous fishes, jawless vertebrates, cephalochordates, echinoderms and insects) using TBLASTN searches against their respective genome assemblies at NCBI

with human protein sequences as queries. The genomic region encompassing the High-scoring Segment Pair (HSP) was extracted and searched against the NCBI NR protein database using BLASTX to confirm the identity of the gene. Each BLAST HSP was then manually verified and checked for the presence of frameshifts/stop codons (if any). Synteny information for each of these genes was obtained from various sources such as the UCSC Genome Browser (http://genome-euro.ucsc.edu/), Genomicus browser (https://www.genomicus.biologie.ens.fr/genomicus-99.01/cgi-bin/search.pl), NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) and Ensembl genome browser (http://asia.ensembl.org/index.html).

Comparative genomics

Comparative analysis of the TDT protein family was performed using the NCBI reference sequences of the following species with modifications: human (*Homo sapiens*; LOC100128908, BC153822, IMAGE:40134016, *LMLN2-204* with addition of the conserved exons 5, 11, 12, 13 and 14, **Figure S2**), rabbit (*Oryctolagus cuniculus*; LOC103351174, XM_008269362.1, XP_008267584.1), gray mouse lemur (*Microcebus murinus;* LOC105864643, XM_012752611.1, XP_012608065.1), ferret (*Mustela putorius furo*; LOC101683375, XM_004755331.2, XP_004755388.1), mouse (*Mus musculus*; LOC101056084, XM_006519864.2, XP_006519927.2), rat (*Rattus norvegicus*; RGD1560492, XM_008770711.2, XP_008768933.1), zebrafish (*Danio rerio*; LOC100331300, XM_002662823, XP_002662869.2 using a START Methionine located 56 amino acids upstream of the predicted one, **Figure S4b**). We used the following coordinates for the full length ORF of human *TDT* (chr14:23,104,920 - 23,099,282): *BC153822/LMLN2-204* with extra exons 5 (23,103,051-23,102,926), 10 (23,101,765-23,101,669), 11 (23,101,564-23,101,292), 12 (23,101,079-23,100,850), 13 (23,100,681-23,100,502) and 14 (23,099,489-23,099,282). To identify the orthologous proteins of TDT, the genome sequence databases were searched using BLASTP and TBLASTN. The TDT protein or deduced amino acid sequences from all 7 species listed above were aligned using ClustalO[68].

Constructs

Full length human TDT (as described above) was synthesized (Genscript) and cloned into the pCS2+ vector. Zebrafish Tdt cDNA (as described above) was cloned from the XM_002662823/ODa12959 clone (Genscript) to which we added the 168 missing nucleotides using a gBlock (IDT) and cloned into the pCS2+ vector. For the zebrafish TDT-ΔTM construct, a STOP codon was introduced before the transmembrane domain. The human CD109 cDNA was also cloned into the pCS2+ vector with a N-terminal FLAG tag.

TDT antibodies

Polyclonal antibodies were raised against the human TDT epitope: CWKKENGFPAGVDNPHGEI, and the zebrafish Tdt epitope: CWIEDNARSGMNEGGGEI within the cysteine-rich domain, and peptide-affinity purified from 2 rabbit sera each (GL Biochem).

HEK293T cell culture and treatments

HEK293T (from ATCC) cells were cultured on plates coated with poly-L-lysine (Sigma P4707) with the following medium: DMEM high glucose (HyClone SH300081.01) with 10% fetal bovine serum (Thermo Scientific SH30070), and 2 mM L-glutamine (ThermoFisher Scientific 25030081). Cells were transfected with DNA plasmid using the FuGENE HD transfection reagent (Promega E2312) in OptiMEM medium (Gibco 31985070). For protein extraction, cells were lysed using an appropriate amount of RIPA buffer (Tris-HCl pH7.5, 50 mM, NaCl 150 mM, NP-40 0.1%, Na2+-deoxycholate 0.05%) supplemented with proteinase inhibitors (Complete, Roche 04693159001). Lysates were centrifuged at 17,000g for 15 min at 4°C to remove cell debris, and the supernatants (protein extracts) were collected. For secretion studies, culture medium was changed 24 h after transfection with a serum-free medium Pro293a-CDM (Lonza 12-764Q) supplemented with L-glutamine. Secretion was allowed for 48 h before collection of the conditioned media. For Phosphatidylinositol-specific Phospholipase C (PI-PLC) treatment, culture medium was changed 48 h after transfection

with serum-free medium without or with 1 U/mL of PI-PLC (Molecular Probes, P-6466), cell culture plates were then placed at 4°C for 20 min rotating before cell lysate and medium extraction. For EndoH (NEB #P0702S) and PNGaseF (NEB #P0704S) treatments, conditioned media were treated as per the manufacturer protocols. For DSS crosslinking experiments, conditioned media were first dialysed overnight in PBS at 4°C and then DMSO or 3 mM final disuccinimidyl suberate (DSS in DMSO) was added to the dialysed conditioned media. The mixture was then incubated at room temperature (RT) for 1 h with maximum shaking of 2000 rpm on a tabletop shaker. The reaction was stopped by the addition of 60 mM final of Tris pH 7.5 and incubated for another 30 min at RT, shaking at 2000 rpm on a tabletop shaker.

Western blotting

For western blotting, samples were electrophoresed with reducing Laemeli loading buffer after denaturation at 95°C for 10 min. The protein ladder (Bio-Rad 161-0377) and denatured and reduced samples were loaded onto 4–20% gradient precast gels (BioRad Criterion 567-1093) in 1x running buffer (25 mM Tris, 200 mM glycine, 0.1% SDS) and ran at 80–180 V until desired separation. Gels were transferred onto 0.2 µm PVDF membranes (BioRad Criterion 170-4157) using the Trans-Blot TurboTM transfer system for 7 min. Membranes were blocked for 1 h at room temperature with 5% milk in TBST (50 mM Tris-HCl pH 7.5, 150 mM NaCl; with 0.05% Tween20). Membranes were incubated with primary antibody diluted in 5% milk in TBST at 4°C overnight (anti-Flag, 1:1,000, Cell Signaling 14793 S; anti-GAPDH, 1:4,000, SantaCruz 47724, anti-zfTdt, 1 µg/mL, in-house; and anti-hTDT, 1 µg/mL, in-house). After washes in TBST, membranes were incubated for 1 h at room temperature with secondary antibodies (Mouse-HRP 71503510 or Rabbit-HRP 711035152, 1:4,000, Jackson Immuno) in 5% milk in TBST. After several washes in TBST, the signal was revealed with the HRP substrate (Thermo Scientific SuperSignal 34080/34076/34096) for 3 min at room temperature. Membranes were then exposed to CL-Xposure films (Thermo Scientific 34091), and developed in a Carestream Kodak developer.

627 *Mouse husbandry and lines*

628 Mice were handled in accordance with the German laws and regulations (Tierschutzgesetz).
629 All procedures were approved by the ethics committee of Lower Saxony for care and use of
630 laboratory animals LAVES. Mice were housed in the animal facility of Hannover Medical
631 School (ZTL) as approved by the responsible Veterinary Officer of the City of Hannover,
632 Germany. Animal welfare was supervised and approved by the Institutional Animal Welfare
633 Officer. *Noto*$^{-/-}$ mutant mice (*Noto*$^{eGFP/eGFP}$) were described previously[69].

634 *Mouse whole mount in situ hybridization (WISH)*

635 Mouse WISH was carried out on E7.0-8.5 (late streak till 6-somite stage) old mouse embryos
636 using standard procedures in detail described in[70]. The DIG-labelled antisense RNA probes
637 were generated using the DIG RNA labelling system (Roche). DIG-labelled probes were
638 detected with BM-Purple AP substrate (Roche). WISH results were documented with the
639 Leica DM5000B microscope with Leica Firecam software. Mouse *Tdt*-probes (Gm29776;
640 ENSMUST00000224691.2) were synthesized with T7 RNA-polymerase from a linearized
641 pCRII-TOPO vector (Invitrogen) containing the following inserts: "*Tdt*-5' part", a 926 bp
642 fragment (exons 3-11) amplified with the forward 5'- AGATTCCAGATGCCCACCTGC-3' and
643 reverse 5'- GCAGCGACTGTGCCTATGGTA-3' primers; or the "*Tdt*-3' part", a 739 bp
644 fragment (exons 11-14) amplified with forward 5'-TACCATAGGCACAGTCGCTGC-3' and
645 reverse 5'-GGAGCATAGCCCGTTTCTGGT-3' primers. Template for PCRs was cDNA from
646 wildtype (CD-1 mouse) E7.5 whole embryos. The cDNA was generated from total RNA using
647 SuperScript II Reverse Transcriptase (Invitrogen) and Oligo(dT) priming.

648 *Zebrafish husbandry and lines*

649 Zebrafish (*Danio rerio*, AB line) were maintained and used according to the Singapore
650 National Advisory Committee on Laboratory Animal Research Guidelines (Institutional
651 Animal Care and Use Committee (IACUC #161172). Zebrafish embryos were grown at
652 28.5°C in egg water as previously described[71]. The *Tg(hsp70::foxj1a)* line was recently

described[36]. For induction of the *foxj1a* transgene, a 1 hr heat shock was administered at 50% epiboly and embryos were fixed at the 90% epiboly stage. The *tdt* mutant line was generated using CRISPR/Cas9 genome editing technology, as previously described[72]. In brief, a custom gBLOCK (Integrated DNA Technologies) was designed incorporating a guide RNA-targeting sequence preceded by a T7 promoter sequence. The targeted sequence on the exon 4 of the *tdt* gene was 5'-GTCCCAACAGTGCTGAGGCC-3' with a guide RNA sequence of 5'-GGCCTCAGCACTGTTGGGAC-3'. The gRNA was synthesized using the MEGA shortscript™ Kit (Applied Biosystems), and was purified using the RNeasy Mini Kit (QIAGEN). Cas9 mRNA was synthesized using the mMESSAGE mMACHINE® Kit (SP6) from a NotI-linearized zebrafish codon-optimized Cas9 construct in pCS2+ (gift from Tom Carney, Singapore). The gRNA and Cas9 mRNAs were mixed together to a concentration of 250 ng/µL each, and 2 nL was injected into the yolk of one-cell stage AB zebrafish embryos. A 5 bp insertion allele resulting in a frameshift and premature stop codon at amino acid 171 was isolated (Line 1, **Figure S4b**). Additional alleles were isolated (Lines 2-5, **Figure S4b**) showing an equivalent phenotype at the homozygous state (data not shown). Mmp21 zebrafish crispants were a gift from Nicholas Katsanis (Duke Med, USA) and were described previously[26]. Three distinct alleles leading to a frameshift with an early stop codon were isolated (**Figure S6a**), all showing an equivalent phenotype (data not shown). Here, we focused on the allele with a complex 10 bp deletion leading to a frameshift at amino acid 38 (Line 3, **Figure S6a**). The heart looping (normal, inverted or no looping) was scored at 48 hours post fertilization (hpf) while embryos were still in their chorion.

Genotyping of zebrafish mutants

The *tdt* mutants could be genotyped by sequencing a 484 bp product containing exon 4 amplified with the following forward 5'-TGTTGGAAACCTGGAACCAT-3' and reverse 5'-ATGGTTCCAGGTTTCCAACA-3 primers. The *tdt* Line 1 was genotyped by allele-specific PCR with the same two primers as well as an internal forward primer annealing specifically to either the wildtype (wt): 5'-AATCCTGAGCATCCAGTCCC-3', or the mutant: 5'-

AATCCTGAGCATCCAGTCAG-3' allele, leading to an additional amplicon of 142 bp or 149 bp, respectively. The *mmp21* mutants were genotyped by sequencing a 240 bp product in exon 1 amplified with the following forward 5'-GAATAAATGCGTGTGGAGTTCA-3' and reverse 5'-ACTGCAAAATACAAAATGTGCG-3' primers.

Zebrafish in situ hybridization

Whole mount *in situ* hybridization (WISH) was carried out using DIG-labelled and fluorescein-labelled antisense probes as previously described[73]. Fluorescein-labelled probes were detected using INT/BCIP and DIG-labelled probes with NBT/BCIP. Probes were synthesized with Sp6 or T7 polymerases from a linearized pGEMT-easy vector containing the following inserts: for *tdt*, a 365 bp fragment between exons 8-10 was amplified with the forward 5'-TGTGCTTCAGGGCTCCATTA-3' and reverse 5'-TCTCTCCGCCTCCTTCATTC-3' primers ; for *lov*, a 631 bp fragment in the single exon was amplified with the forward 5'-ACCCCGAGTGTGGAATATCA-3' and reverse 5'-TCTCCAAACACCTCCTTTGC-3' primers ; for *dand5*, a 609 bp fragment in exon 2 was amplified with the forward 5'-GTCAGTGCAGCTCCATGTTC-3' and reverse 5'-CCAATTAATATTTCGGCCACTT-3' primers ; for *myl7 (cmlc2)*, a 246 bp fragment between exons 2-5 was amplified with the forward 5'-ATACAGGAGTTTAAGGAGGC-3' and reverse 5'-GTCGAACAATTTAAAAGCAG-3' primers ; for *ins*, a 437 bp fragment between exons 1-3 was amplified with the forward 5'-CCATATCCACCATTCCTCGC-3' and reverse 5'-CAAACGGAGAGCATTAAGGC-3' primers ; for *lefty1*, the complete cDNA was amplified with the forward 5'-GCACAGCAGTGACAGCTTCT-3' and reverse 5'-TTCAGTCAGTCTCACATATA-3' primers ; for *lefty2*, a 465 bp fragment in exon 4 was amplified with the forward 5'-ATCCCAAGGCAACTGTAACT-3' and reverse 5'-AAGCTTACATTACATAAAGC-3' primers ; for *tbxta (ntl)*, the complete cDNA was amplified with the forward 5'-CGCTGTCAAAGCAACAGTAT-3' and reverse 5'-GAACTCCTTTTTTTTGACAG-3' primers ; for *sox17*, a 961 bp fragment between exons 1-2 was amplified with the forward 5'-TGAATGAACTGTATGCACTC-3' and reverse 5'-TGCAGGTTTATTGAACTGAG-3' primers ;

and for *foxj1a*, a 634 bp fragment between exons 1-2 was amplified with the forward 5'-CATCAAGCCGCCATACTCAT-3' and reverse 5'-TGAATCCAGTAGAGCGTCCC-3' primers.

Zebrafish mRNA and morpholino injection

Embryos were injected at the one-cell stage according to standard procedures. For overexpression/rescue experiments, SP6-transcribed full length zebrafish *tdt* capped mRNA (mMessage mMachine SP6 transcription kit, Thermofisher #AM1340) was injected into one-cell stage embryos at the indicated concentration. The catalytic dead (cd) Tdt corresponds to p.His247Ala, p.Glu248Ala and p.His251Ala mutations (HExxH to AAxxA). The *mmp21* splice-blocking morpholino oligonucleotide (MO) 5'-GTTGTATATTTGTTCACTGACCCGT-3' has been described previously[27]. The *foxj1a* translation-blocking MO 5′-CATGGAACTCATGGAGAGCATGGTC-3′ was previously described[34].

Zebrafish real-time qPCR

For qPCR experiments, embryos at indicated stages were lysed in the QIAGEN RLT buffer and total RNAs were extracted using the QIAGEN RNeasy Mini kit (74106), including the optional DNase RNase-free treatment. cDNAs were obtained using the iScript reverse transcription supermix (Bio-Rad 170-8841). qPCR were performed with the following primers using the Power SYBR Green Master mix (Applied Biosystems 4367659) on the Applied Biosystems 7900HT Fast Real-Time PCR system. zf tdt-qPCR-F: 5'-CGATGCTGACTTCCTGCTTT -3' ; zf tdt-qPCR-R: 5'-GAACCCGTCTGACAGTGAGC -3' ; zf mmp21-qPCR-F: 5'-GGACCGCAAATCTATCCAGA-3' ; zf mmp21-qPCR-R: 5'-CGTCTGCTCCTTTCTGATCC-3' ; zf actin-qPCR-F: 5'-GATCTTCACTCCCCTTGTTCA-3' ; zf actin-qPCR-R: 5'-GGCAGCGATTTCCTCATC-3'. Data are average of at least three biological triplicates and statistical analyses were done with the PRISM5® software using a one-way ANOVA test with Bonferroni correction for multiple hypothesis testing when more than two groups were compared, or an unpaired t-test with Welch's correction when less than three groups were compared. Non-significant (ns) indicates a p-value of $p>0.05$ and asterisks

indicate p-values of p<0.05(*), p<0.01(**) and p<0.001(***). Error bars indicate standard error of the mean (s.e.m.).

Zebrafish immunofluorescence

Wildtype and *tdt*$^{-/-}$ mutant 10-somite embryos were fixed for 2 hours at room temperature in 4% paraformaldehyde (PFA) in PBS. The embryos were stored in methanol at -20°C, and then rehydrated in a gradient of PBS/methanol. The embryos were permeabilized in acetone, blocked in PBDT (PBS, 1% BSA, 1% DMSO and 1% Triton X-100) plus 5% sheep serum. Embryos were incubated with primary and secondary antibodies and washed in PBDT. The primary antibodies used were anti-acetylated-alpha-tubulin (rabbit monoclonal, Cell Signaling, 5335S), anti-acetylated-tubulin (monoclonal, Sigma, T6793), anti-gamma (γ)-tubulin (monoclonal, Sigma, T6557) and the secondary antibodies used were Alexa fluorophore-labeled anti-mouse and anti-rabbit (Molecular Probes). Immunofluorescence images were captured on an Olympus Fluoview 1000 microscope with 60 x, 1.4 NA oil immersion objectives.

Zebrafish KV movies

For analysis of cilia motility in KV and movement of endogenous particles suspended in KV fluid, appropriately staged zebrafish embryos were embedded in 1.5% low-melting-point agarose on 50 mm glass-bottom dishes. Ciliary motility and particle movement were viewed with a Zeiss 63X water-dipping objective on a Zeiss Axioplan2 (upright) microscope equipped with an ORCA-Flash4.0 V2C11440-22CU camera (Hamamatsu). Processing of videos and was performed with ImageJ 1.44d (NIH, USA).

Xenopus husbandry and injection experiments

Handling of *Xenopus laevis* frogs (Nasco) was in accordance with the German laws (Tierschutzgesetz), and was approved by the Regional Government Stuttgart, Germany. Animal welfare was supervised and approved by the Institutional Animal Welfare Officer. Embryos were injected in Ficoll medium (1x MSBH and 2 % Ficoll) at the 1- or 4-cell stage,

and were further cultivated in 0.1x MSBH. *Xenopus laevis tdt* crispants were created with a gRNA 5'-GGGGAAGTCAGGGATCTAGG-3' designed via CRISPRscan[74] targeting the splice acceptor site of exon 4. gRNA templates harboring a T7 promoter were assembled using oligo extension reaction[75]. The gRNA was transcribed using the MEGAshortscript T7 Transcription Kit (Invitrogen) in combination with the MEGAclear Transcription Clean-Up Kit (Invitrogen). Ribonucleoproteins (RNPs) consisting of gRNA and Cas9 protein (PNA Bio) were pre-assembled at 37°C for 5 min and in total 1 ng Cas9 protein was injected together with 150 pg gRNA. Genome editing was confirmed via direct sequencing of PCR products using the following forward 5'-AAAGTGTTTGGACGTCACAGT-3' and reverse 5'-TCACCTGAGCTGCACATTTCT-3' primers[75]. Efficiency was determined using the Synthego ICE analysis tool[76].

Xenopus in situ hybridization

WISH was performed using DIG-labelled antisense probes in combination with BM-Purple staining solution (Roche) as previously described[77]. All probes were transcribed from linearized plasmids using T7 or SP6 RNA polymerases (Promega). For *tdt,* a 774 bp fragment from exon 1 was amplified with the following forward 5'-ACACCCACTTGTACATCGCT-3' and reverse 5'-GCAGGGCAATCTTCAGTTGT-3' primers.

Xenopus immunofluorescence

Embryos were fixed at stage 17 overnight at 4°C in 4% PFA in PBS^- and stored in fixative. The specimens were permeabilized in PBST (PBS, 1% Triton X-1000) and blocked in CAS-Block (Thermo Fisher). Embryos were subsequently incubated with the primary anti-acetylated-α-tubulin 6-11B-1 antibody (1:700 dilution, Sigma, #T6793) followed by the secondary anti-mouse IgG Cy3 antibody (1:1000 dilution, Merck) in combination with Alexa Fluor 488 Phalloidin (1:200 dilution, Thermo Fisher). Imaging was performed on a Zeiss Observer.Z1 microscope with a Zeiss Plan-Apochromat 20x, 0.8, objective coupled to a Zeiss LSM 600.

Xenopus LRO movies

Leftward flow was analyzed as previously described[78]. In brief, dorsal explants were crafted at stage 17 and bathed in a solution of yellow-green fluorescent 0.5 µm FluoSpheres (1:3000, 1x MSBH, Invitrogen). Explants were placed in a chamber made from duct tape, a slide and a cover slip. 1 min time-lapse movies of the flow were captured on a Zeiss Axioskop 2 mot plus microscope with a Zeiss Plan-Neofluar 20x, 0.5, objective in combination with a Zeiss Hsm Camera at 2 FPS. Particle trajectories were traced with the ImageJ[79] plugin Particle Tracker[80]. Measurements of the trajectories and the visualization of the particle movements in gradient-time-trails (GTTs)[78] was done with a custom-made script written in R[81].

Human study participants

The study included 57 individuals from 12 families, of which 21 individuals were affected with left-right asymmetry defects. Families ethnologically originated from France (Families 5, 8, 11 and 12), Turkey (Families 1, 2, 4, 6 and 7), Italy (Family 3) and Lebanon (Families 9 and 10) (**See Figure 7 and Table 1**). The study protocol was approved by A*STAR IRB (2019-087) and genetic analyses were performed in accordance with bioethics rules of French laws. Written informed consents were obtained from all the participants or parents.

DNA extraction, quantification, and quality control

Genomic DNA of the affected individuals and other available family members was extracted from either whole blood or fetal tissue using a variety of extraction protocols. DNA concentration and quality were assessed using NanoDrop (Thermo Scientific) and Qubit (Life technologies) fluorometers. A260/A280 ratios of 1.8 to 2.0 and A260/A230 ratios >1.5 were accepted. DNA fragmentation was assessed using agarose gel (0.8%) electrophoresis.

Genome-wide genotyping

For Families 1, 2, 3, 6 and 7, minimal amount of 1 microgram at a minimal concentration of 50 ng/mL of each gDNA was used for genotyping on HumanCoreExome-12 v1.0 (WG-353-

1104) arrays (Illumina, San Diego, USA) according to the Illumina protocol 'Infinium HD Ultra Assay Automated EUC (11328108 B)' (Illumina, San Diego, USA). The arrays were scanned on an IScan+ scanner. Genotyping results were obtained using GenomeStudio (Illumina, San Diego, USA). Quality controls were performed on inner house controls, and control DNA. Genotyping and pedigree information were compared (sex, familial relationship, mendelian allele segregation) with Linkdatagen[82,83], a PERL script that generates datasets for linkage analysis, relatedness checking, IBD and HBD inference. Parental relatedness was also checked with Graphical Representation of Relationships (http://csg.sph.umich.edu/abecasis/GRR/index.html)[84]. Linkdatagen created output files for MERLIN[85]. Parametric (disease allele frequency of 0.0001, a fully penetrant disease, and a 0% phenocopy rate, autosomal recessive inheritance) and non-parametric analyses were performed. Haplotypes of the region of interest were prepared with GeneHunter. Genehunter output files were used to visualize haplotypes with the Haplopainter 1.043 software[86]. Two-point analyses (between the disease phenotype and TDT mutations) were carried out with EasyLinkage[87] and Superlink v1.6 (http://bioinfo.cs.technion.ac.il/superlink/) with a disease allele frequency of 0.0001, a phenocopy rate of 0%, an autosomal recessive inheritance and a penetrance value of 100%.

For Family 8, genome-wide linkage scans were performed using the Affymetrix SNP 10K2.1 array (Affymetrix, Santa Clara, CA, USA) in the two affected individuals (V:2 and V:4) and their healthy sister (V:3). These data were analyzed by calculating multipoint lod scores using MERLIN software and assuming a recessive inheritance with complete penetrance thanks to a collaboration with Dr. Emmanuelle GENIN. The region of homozygosity on chromosome 14 was confirmed and refined by haplotyping with microsatellite markers selected on the UCSC browser. The microsatellite markers were amplified in a multiplex manner using the QIAGEN Multiplex PCR kit. The amplification products were finally analyzed on an ABI Prism 3130 analyzer (Applied Biosystems). The haplotyping of the markers has been optimized using Genescan Analysis and Genotyper softwares.

Ampliseq design, library preparation and targeted sequencing

Targeted sequencing was used in Families 1-7, and 11-12. In brief, the 14 exons including UTRs and 20 intronic nucleotides on each side of each exon of the *TDT* gene, along with 2 other genes (total target region: 32450 nucleotides, total target on *TDT*: 4039 nucleotides) were selected to generate primers according to Ampliseq Designer (https://www.ampliseq.com/) with a predicted *TDT* sequencing coverage of 100%. Pools of 6 gDNA samples were prepared by pooling together gDNA extracted with the same protocol. For library construction, 5ng of pooled DNA was amplified using the customized panel (Ampliseq, Thermo Fisher). The amplicons were then partially digested, barcoded and amplified using the Ion Ampliseq™ Library kit 2.0 and Ion Xpress™ barcode adapter kit (Thermo Fisher) according to manufacturer instructions. The library was quantified using the Qubit fluorometer and the BioAnalyzer 2100 (Agilent). Libraries were multiplexed at a final concentration of 15 pM, and 25 μL were clonally amplified on Ion sphere™ particles (ISP) by emulsion PCR performed on the Ion One Touch 2 instrument (Thermo Fisher) according to manufacturer instructions. Quality control was performed using the Ion Sphere™ Quality Control kit (Thermo Fisher) to ensure that 10-30% of templates positive ISP was generated in the emulsion PCR. Finally, the template ISP were enriched, loaded on an Ion 318™ chip v2 and sequenced on a PGM™ sequencer with the Ion PGM™ sequencing 200 kit v2 according to manufacturer instructions. Quality of sequencing was assessed by the PGM sequencer by providing the Ion Sphere Particle (ISP) density, number of total reads, percentage of usable reads, percentages of monoclonal and polyclonal reads, mean read length (AQ17, AQ20 and perfect), percentages of low quality sequences, adapter dimers and aligned bases. Bam files were loaded on Alamut Visual 2.5 (Interactive Biosoftware, Rouen, France) and variant detection was performed manually by setting the variant detection threshold to 0.04. STOP gain, splice site, frameshift and rare missense variants (< 0.1%) were selected for Sanger sequencing of the 6 individual DNA composing the pool carrying each selected variant. Variants of interest were confirmed on a second dilution of DNA and

on an independent sample. Subsequently, relatives with available gDNA samples were tested for variant segregation analysis.

Exome sequencing

Exome sequencing was employed independently for the detection of variants in Family 8 (Trio analysis with individuals IV:1 (mother), V:2 (patient) and V:4 (affected sister)), Family 9 (IV:1 (patient) and IV:3 (affected brother)), and Family 10 (IV:2 (patient)). One microgram of high-quality gDNA was used for exome capture with the ION TargetSeq Exome Kit. The exome library was prepared on an ION OneTouch System and sequenced on an Ion Proton instrument (Life Technologies, Carlsbad, CA, USA) using one ION PI chip. Sequence reads were aligned to the human reference genome (Human GRCh37 (hg19) build) using Torrent Mapping Alignment Program (TMAP) from the Torrent Suite (v5.0.2). The variants were called using the Torrent Variant Caller (TVC) plugin (v5.0.2), and were annotated with the associated gene, location, protein position and amino acid changes, quality-score, coverage, predicted functional consequences using SIFT[88], PolyPhen2[89], Grantham[90] prediction scores, phyloP[91] conservation scores, and 5000 genomes Minor Allele Frequencies. Variants were filtered for common SNPs using the NCBI's "common and no known medical impacts" database (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/), the Exome Aggregate Consortium (ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.2/) and the Exome Sequencing Project (http://evs.gs.washington.edu/EVS/). We next removed variants that were present in greater than 1% of the previously in-house 478 sequenced samples. For the Family 8 trio analysis, a total of 17.5 Gb, 14.8 Gb, and 14.6 Gb were sequenced with an average read length of 183 bp, 175 bp, and 180bp for individuals IV:1, V:2, and V:4, respectively. An average coverage of 212X, 183X, and 171X was achieved over the exome, with 96% of bases covered at least 20X for each individual. A combined total of 46,723 variants were identified across protein-coding exons, UTRs, splice sites and flanking introns. Additional filters were applied to retain variants that were homozygous in both probands and heterozygous in the mother. A final set of 14 variants remained, of which only one was not a

892 common SNP, hg19: chr14:23571855 C>G, which corresponds to a *TDT* c.806G>C;
893 p.C269S mutation. Using the same method, we identified a homozygous *TDT* mutation in the
894 two probands in Family 9 (*TDT* c.1202_1204delTCT ; p.Phe401del) and in the proband of
895 Family 10 (*TDT* c.1144delG ; p.Q381fsX9/splicing defects). Subsequently, relatives with
896 available gDNA samples were tested for variant segregation analysis.

897 *Data availability*

898 The data that support the findings of this study are available from the corresponding author
899 upon reasonable request.

900

Acknowledgments

We are grateful to all members of the Reversade laboratory for support. We thank Ray DUNN (NTU-LKC, Singapore) for the gift of the zebrafish *lefty1/2*, *spaw* and *insulin* plasmids for *in situ* hybridization experiments and critical feedback. We thank Vijay NARASIMHAN for participating in the initial stages of the zebrafish study. E.S-R is supported by a NMRC Open Fund - Young Individual Research Grant (OF-YIRG, #OFYIRG18May-0053). M.B. was supported by DFG grants BL285/xx. T.O. was the recipient of a fellowship of the Landesgraduiertenförderung Baden-Württemberg. B.R. is a fellow of the National Research Foundation (NRF, Singapore) and Branco Weiss Foundation (Switzerland) and an EMBO Young Investigator. This work was funded by a Strategic Positioning Fund for Genetic Orphan Diseases and an inaugural A*STAR Investigatorship from the Agency for Science, Technology and Research in Singapore.

Author contributions

E.S-R, T.O, M.B, P.B, and B.R designed the study. T.A-B, M.R, S.M, P.S, M.F, E.C, A.M, L.D, G.C, S.DF, C.R-T, J-F.D, A.B, N.A, B.T, R.E, and P.B made clinical diagnoses and collected clinical data and samples. Family 8 from T.A-B and M.R first allowed to identify *TDT* as the causative gene. E.S-R, A.MdB, C.B, S.T, AYJ.N and B.V performed WES, homozygosity mapping, high throughput cohort re-sequencing and sequencing analyses. E.S-R, M.K, YL.C, WX.G, D.K, P.A, S.R and B.R performed and supervised the zebrafish functional experiments. T.O and M.B performed and supervised the *Xenopus* functional experiments. A.B and A.G performed and supervised the mouse experiments. V.R, B.E, F.E and B.V performed evolution genomics analyses. E.S-R and B.R wrote the manuscript with input from S.R, M.B, and P.B.

Competing financial interest

The authors declare no competing financial interests.

References

1. Shiratori, H. & Hamada, H. TGFβ signaling in establishing left-right asymmetry. *Semin. Cell Dev. Biol.* **32**, 80–84 (2014).
2. Zhang, H. T. & Hiiragi, T. Symmetry Breaking in the Mammalian Embryo. *Annu. Rev. Cell Dev. Biol.* **34**, 405–426 (2018).
3. Bier, E. & De Robertis, E. M. EMBRYO DEVELOPMENT. BMP gradients: A paradigm for morphogen-mediated developmental patterning. *Science* **348**, aaa5838 (2015).
4. Sokol, S. Y. Wnt signaling and dorso-ventral axis specification in vertebrates. *Current Opinion in Genetics & Development* vol. 9 405–410 (1999).
5. Petersen, C. P. & Reddien, P. W. Wnt Signaling and the Polarity of the Primary Body Axis. *Cell* **139**, 1056–1068 (2009).
6. Niehrs, C. Regionally specific induction by the Spemann–Mangold organizer. *Nature Reviews Genetics* **5**, 425–434 (2004).
7. Blum, M. & Ott, T. Animal left-right asymmetry. *Curr. Biol.* **28**, R301–R304 (2018).
8. Shinohara, K. & Hamada, H. Cilia in Left-Right Symmetry Breaking. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
9. Grimes, D. T. Making and breaking symmetry in development, growth and disease. *Development* **146**, (2019).
10. Grimes, D. T. & Burdine, R. D. Left-Right Patterning: Breaking Symmetry to Asymmetric Morphogenesis. *Trends Genet.* **33**, 616–628 (2017).
11. Sutherland, M. J. & Ware, S. M. Disorders of left-right asymmetry: heterotaxy and situs inversus. *Am. J. Med. Genet. C Semin. Med. Genet.* **151C**, 307–317 (2009).
12. Lin, A. E. *et al.* Laterality defects in the national birth defects prevention study (1998-2007): birth prevalence and descriptive epidemiology. *Am. J. Med. Genet. A* **164A**, 2581–2591 (2014).
13. Mirra, V., Werner, C. & Santamaria, F. Primary Ciliary Dyskinesia: An Update on Clinical Aspects, Genetics, Diagnosis, and Future Treatment Strategies. *Front Pediatr* **5**, 135

(2017).

14. Zariwala, M. A., Omran, H. & Ferkol, T. W. The emerging genetics of primary ciliary dyskinesia. *Proc. Am. Thorac. Soc.* **8**, 430–433 (2011).
15. Mohapatra, B. *et al.* Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations. *Hum. Mol. Genet.* **18**, 861–871 (2009).
16. Karkera, J. D. *et al.* Loss-of-function mutations in growth differentiation factor-1 (GDF1) are associated with congenital heart defects in humans. *Am. J. Hum. Genet.* **81**, 987–994 (2007).
17. Kosaki, R. *et al.* Left-right axis malformations associated with mutations in ACVR2B, the gene for human activin receptor type IIB. *Am. J. Med. Genet.* **82**, 70–76 (1999).
18. Bamford, R. N. *et al.* Loss-of-function mutations in the EGF-CFC gene CFC1 are associated with human left-right laterality defects. *Nat. Genet.* **26**, 365–369 (2000).
19. Kosaki, K. *et al.* Characterization and Mutation Analysis of Human LEFTY A and LEFTY B, Homologues of Murine Genes Implicated in Left-Right Axis Development. *The American Journal of Human Genetics* **64**, 712–721 (1999).
20. Cristo, F. *et al.* Functional study of DAND5 variant in patients with Congenital Heart Disease and laterality defects. *BMC Med. Genet.* **18**, 77 (2017).
21. Sempou, E. & Khokha, M. K. Genes and mechanisms of heterotaxy: patients drive the search. *Curr. Opin. Genet. Dev.* **56**, 34–40 (2019).
22. Perles, Z. *et al.* A human laterality disorder associated with recessive CCDC11 mutation. *J. Med. Genet.* **49**, 386–390 (2012).
23. Narasimhan, V. *et al.* Mutations in CCDC11 , which Encodes a Coiled-Coil Containing Ciliary Protein, Causes Situs Inversus Due to Dysmotility of Monocilia in the Left-Right Organizer. *Human Mutation* **36**, 307–318 (2015).
24. Vetrini, F. *et al.* Bi-allelic Mutations in PKD1L1 Are Associated with Laterality Defects in Humans. *Am. J. Hum. Genet.* **99**, 886–893 (2016).
25. Gebbia, M. *et al.* X-linked situs abnormalities result from mutations in ZIC3. *Nat. Genet.*

17, 305–308 (1997).

26. Perles, Z. *et al.* A human laterality disorder caused by a homozygous deleterious mutation in MMP21. *J. Med. Genet.* **52**, 840–847 (2015).
27. Guimier, A. *et al.* MMP21 is mutated in human heterotaxy and is required for normal left-right asymmetry in vertebrates. *Nat. Genet.* **47**, 1260–1263 (2015).
28. Akawi, N. *et al.* Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat. Genet.* **47**, 1363–1369 (2015).
29. Gros, J., Feistel, K., Viebahn, C., Blum, M. & Tabin, C. J. Cell movements at Hensen's node establish left/right asymmetric gene expression in the chick. *Science* **324**, 941–944 (2009).
30. Blum, M., Feistel, K., Thumberger, T. & Schweickert, A. The evolution and conservation of left-right patterning mechanisms. *Development* **141**, 1603–1613 (2014).
31. Yao, C., Donelson, J. E. & Wilson, M. E. The major surface protease (MSP or GP63) of Leishmania sp. Biosynthesis, regulation of expression, and function. *Mol. Biochem. Parasitol.* **132**, 1–16 (2003).
32. Petrovski, S. *et al.* Whole-exome sequencing in the evaluation of fetal structural anomalies: a prospective cohort study. *Lancet* **393**, 758–767 (2019).
33. Aamar, E. & Dawid, I. B. Sox17 and chordin are required for formation of Kupffer's vesicle and left-right asymmetry determination in zebrafish. *Dev. Dyn.* **239**, 2980–2988 (2010).
34. Yu, X., Ng, C. P., Habacher, H. & Roy, S. Foxj1 transcription factors are master regulators of the motile ciliogenic program. *Nat. Genet.* **40**, 1445–1453 (2008).
35. Stubbs, J. L., Oishi, I., Izpisúa Belmonte, J. C. & Kintner, C. The forkhead protein Foxj1 specifies node-like cilia in Xenopus and zebrafish embryos. *Nat. Genet.* **40**, 1454–1460 (2008).
36. Choksi, S. P., Babu, D., Lau, D., Yu, X. & Roy, S. Systematic discovery of novel ciliary genes through functional genomics in the zebrafish. *Development* **141**, 3410–3419

(2014).

37. Maisonneuve, C. *et al.* Bicaudal C, a novel regulator of Dvl signaling abutting RNA-processing bodies, controls cilia orientation and leftward flow. *Development* **136**, 3019–3030 (2009).
38. Beckers, A., Alten, L., Viebahn, C., Andre, P. & Gossler, A. The mouse homeobox gene Noto regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15765–15770 (2007).
39. Huang, C.-J., Tu, C.-T., Hsiao, C.-D., Hsieh, F.-J. & Tsai, H.-J. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228**, 30–40 (2003).
40. Szenker-Ravi, E. *et al.* RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of LGR4/5/6. *Nature* **557**, 564–569 (2018).
41. Blum, M. *et al.* Xenopus, an ideal model system to study vertebrate left-right asymmetry. *Dev. Dyn.* **238**, 1215–1225 (2009).
42. Hamada, H. Left-Right Asymmetry. *Mouse Development* 55–73 (2002).
43. Logan, M., Pagán-Westphal, S. M., Smith, D. M., Paganessi, L. & Tabin, C. J. The transcription factor Pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. *Cell* **94**, 307–317 (1998).
44. Yoshioka, H. *et al.* Pitx2, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry. *Cell* **94**, 299–305 (1998).
45. Ryan, A. K. *et al.* Pitx2 determines left-right asymmetry of internal organs in vertebrates. *Nature* **394**, 545–551 (1998).
46. Campione, M. *et al.* The homeobox gene Pitx2: mediator of asymmetric left-right signaling in vertebrate heart and gut looping. *Development* **126**, 1225–1234 (1999).
47. Schweickert, A. *et al.* The nodal inhibitor Coco is a critical target of leftward flow in Xenopus. *Curr. Biol.* **20**, 738–743 (2010).
48. Nakamura, T. *et al.* Fluid flow and interlinked feedback loops establish left-right asymmetric decay of Cerl2 mRNA. *Nat. Commun.* **3**, 1322 (2012).

49. Skoog, T. *et al.* MMP-21 is expressed by macrophages and fibroblasts in vivo and in culture. *Exp. Dermatol.* **15**, 775–783 (2006).
50. Ahokas, K. *et al.* Matrix metalloproteinase-21, the human orthologue for XMMP, is expressed during fetal development and in cancer. *Gene* **301**, 31–41 (2002).
51. Ahokas, K. *et al.* Matrix metalloproteinase-21 is expressed epithelially during development and in cancer and is up-regulated by transforming growth factor-beta1 in keratinocytes. *Lab. Invest.* **83**, 1887–1899 (2003).
52. Skoog, T. *et al.* Matrix metalloproteinase-21 expression is associated with keratinocyte differentiation and upregulated by retinoic acid in HaCaT cells. *J. Invest. Dermatol.* **129**, 119–130 (2009).
53. Wu, T. *et al.* Increased MMP-21 expression is associated with poor overall survival of patients with gastric cancer. *Med. Oncol.* **30**, 323 (2013).
54. Zhang, J. *et al.* Overexpression of MMP21 and MMP28 is associated with gastric cancer progression and poor prognosis. *Oncol. Lett.* **15**, 7776–7782 (2018).
55. Van Wart, H. E. & Birkedal-Hansen, H. The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 5578–5582 (1990).
56. Field, S. *et al.* Pkd1l1 establishes left-right asymmetry and physically interacts with Pkd2. *Development* **138**, 1131–1142 (2011).
57. Bell, E. Cell fate specification and competence by Coco, a maternal BMP, TGFbeta and Wnt inhibitor. *Development* **130**, 1381–1389 (2003).
58. Piccolo, S. *et al.* The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature* **397**, 707–710 (1999).
59. Lee, H. X., Ambrosio, A. L., Reversade, B. & De Robertis, E. M. Embryonic dorsal-ventral signaling: secreted frizzled-related proteins as inhibitors of tolloid proteinases. *Cell* **124**, 147–159 (2006).
60. Reversade, B. & De Robertis, E. M. Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field. *Cell* **123**, 1147–1160

(2005).

61. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
62. Niehrs, C. & Pollet, N. Synexpression groups in eukaryotes. *Nature* **402**, 483–487 (1999).
63. Danilchik, M. V., Brown, E. E. & Riegert, K. Intrinsic chiral properties of the Xenopus egg cortex: an early indicator of left-right asymmetry? *Development* **133**, 4517–4526 (2006).
64. Blum, M., Schweickert, A., Vick, P., Wright, C. V. E. & Danilchik, M. V. Symmetry breakage in the vertebrate embryo: when does it happen and how does it work? *Dev. Biol.* **393**, 109–123 (2014).
65. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
66. Sirota, F. L. *et al.* Beware of moving targets: reference proteome content fluctuates substantially over the years. *J. Bioinform. Comput. Biol.* **10**, 1250020 (2012).
67. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, (2016).
68. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
69. Abdelkhalek, H. B. *et al.* The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation. *Genes & Development* **18**, 1725–1736 (2004).
70. Stauber, M. *et al.* Identification of FOXJ1 effectors during ciliogenesis in the foetal respiratory epithelium and embryonic left-right organiser of the mouse. *Dev. Biol.* **423**, 170–188 (2017).
71. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Brachydanio Rerio)*. (University of Oregon Press, 1989).
72. Jao, L.-E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13904–13909

(2013).

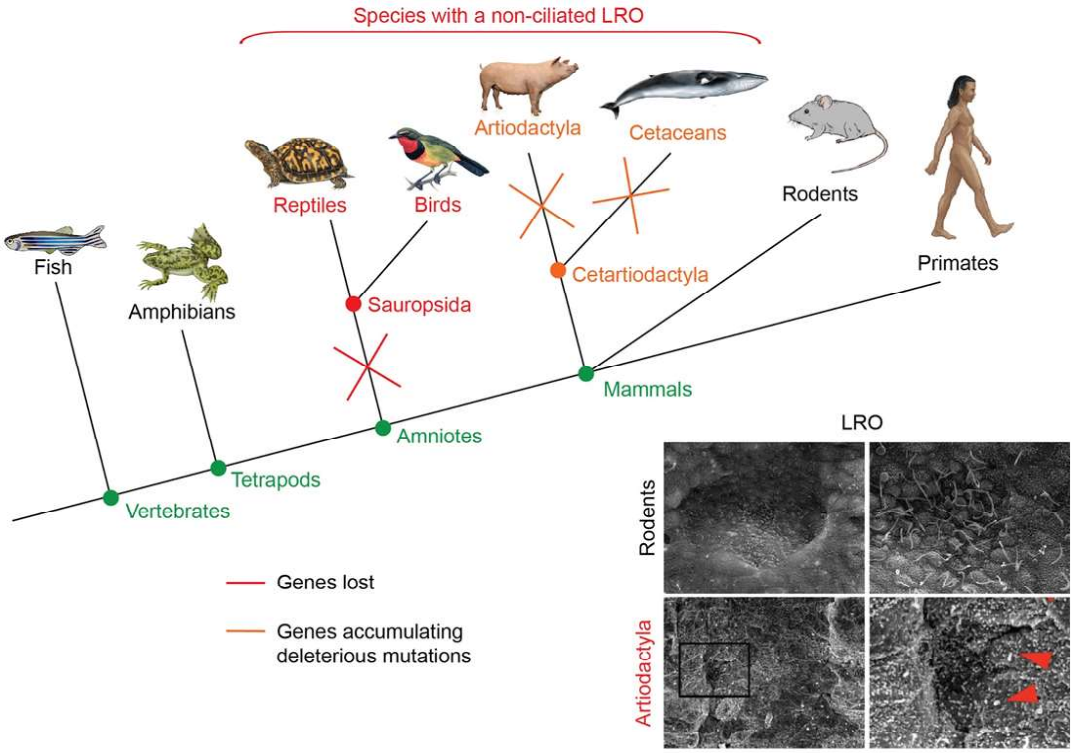
73. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008).
74. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–988 (2015).
75. Nakayama, T. *et al.* Cas9-based genome editing in Xenopus tropicalis. *Methods Enzymol.* **546**, 355–375 (2014).
76. Hsiau, T. *et al.* Inference of CRISPR Edits from Sanger Trace Data. doi:10.1101/251082.
77. Belo, J. A. *et al.* Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mech. Dev.* **68**, 45–57 (1997).
78. Schweickert, A. *et al.* Cilia-driven leftward flow determines laterality in Xenopus. *Curr. Biol.* **17**, 60–66 (2007).
79. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).
80. Sbalzarini, I. F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **151**, 182–195 (2005).
81. R Core Team. *R: A language and Environment for Statistical Computing*. https://www.R-project.org/ (2017).
82. Bahlo, M. & Bromhead, C. J. Generating linkage mapping files from Affymetrix SNP chip data. *Bioinformatics* **25**, 1961–1962 (2009).
83. Smith, K. R. *et al.* Reducing the exome search space for mendelian diseases using genetic linkage analysis of exome genotypes. *Genome Biol.* **12**, R85 (2011).
84. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. GRR: graphical representation of relationship errors. *Bioinformatics* **17**, 742–743 (2001).
85. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
86. Thiele, H. & Nürnberg, P. HaploPainter: a tool for drawing pedigrees with complex

haplotypes. *Bioinformatics* **21**, 1730–1732 (2005).

87. Lindner, T. H. & Hoffmann, K. easyLINKAGE: a PERL script for easy and automated two-/multi-point linkage analyses. *Bioinformatics* **21**, 405–407 (2005).
88. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
89. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
90. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).
91. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
92. Follit, J. A. *et al.* Arf4 is required for Mammalian development but dispensable for ciliary assembly. *PLoS Genet.* **10**, e1004170 (2014).

Figures and legends

a



b

Gene	Protein	Signal peptide	Present in species with		Role in left-right asymmetry		OMIM
			a ciliated node	an unciliated node	in mouse	in human	
TDT	Tout-de-travers	+	+	-	n.d.	this manuscript	n.d.
ALED	Alendroit	+	+	-	Cecilia Lo laboratory, MGI	(Petrovski S *et al.*, 2019)	n.d.
MMP21	Matrix metalloprotease 21	+	+	-	(Guimier A *et al.*, 2015)	(Guimier A *et al.*, 2015; Perles *et al.*, 2015; Akawi *et al.*, 2015)	HTX7, 616749
PKD1L1	Polycystin-1, like 1	+	+	-	(Vogel P *et al.*, 2010 ; Field S *et al.*, 2011)	(Vetrini F *et al.*, 2016)	HTX8, 617205
DAND5	DAN domain family member 5	+	+	-	(Marques S *et al.*, 2004)	(Cristo *et al.*, 2017)	n.d.
NODAL	Nodal	+	+	+	(Brennan J *et al.*, 2002)	(Mohaptra B *et al.*, 2009)	HTX5, 601265
GDF1	Embryonic growth/differentiation factor 1	+	+	+	(Rankin CT *et al.*, 2000)	(Karkera JD *et al.*, 2007; Kaasinen E *et al.*, 2010)	CHTD6, 613854; RAI, 208530
ACVR2B	Activin receptor type 2B	+	+	+	(Oh SP *et al.*, 1997)	(Kosaki R *et al.*, 1999)	HTX4, 602730
CFC1	Cryptic	+	+	+	(Yan YT *et al.*, 1999; Gaio U *et al.*, 1999)	(Bamford RN *et al.*, 2000)	HTX2, 605194
LEFTY2	Left-right determination factor 2	+	+	+	(Meno C *et al.*, 2001)	(Kosaki K *et al.*, 1999)	n.d.

c

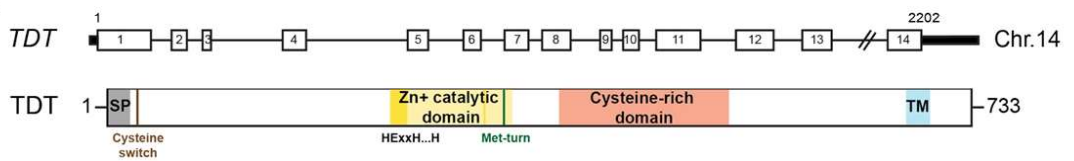
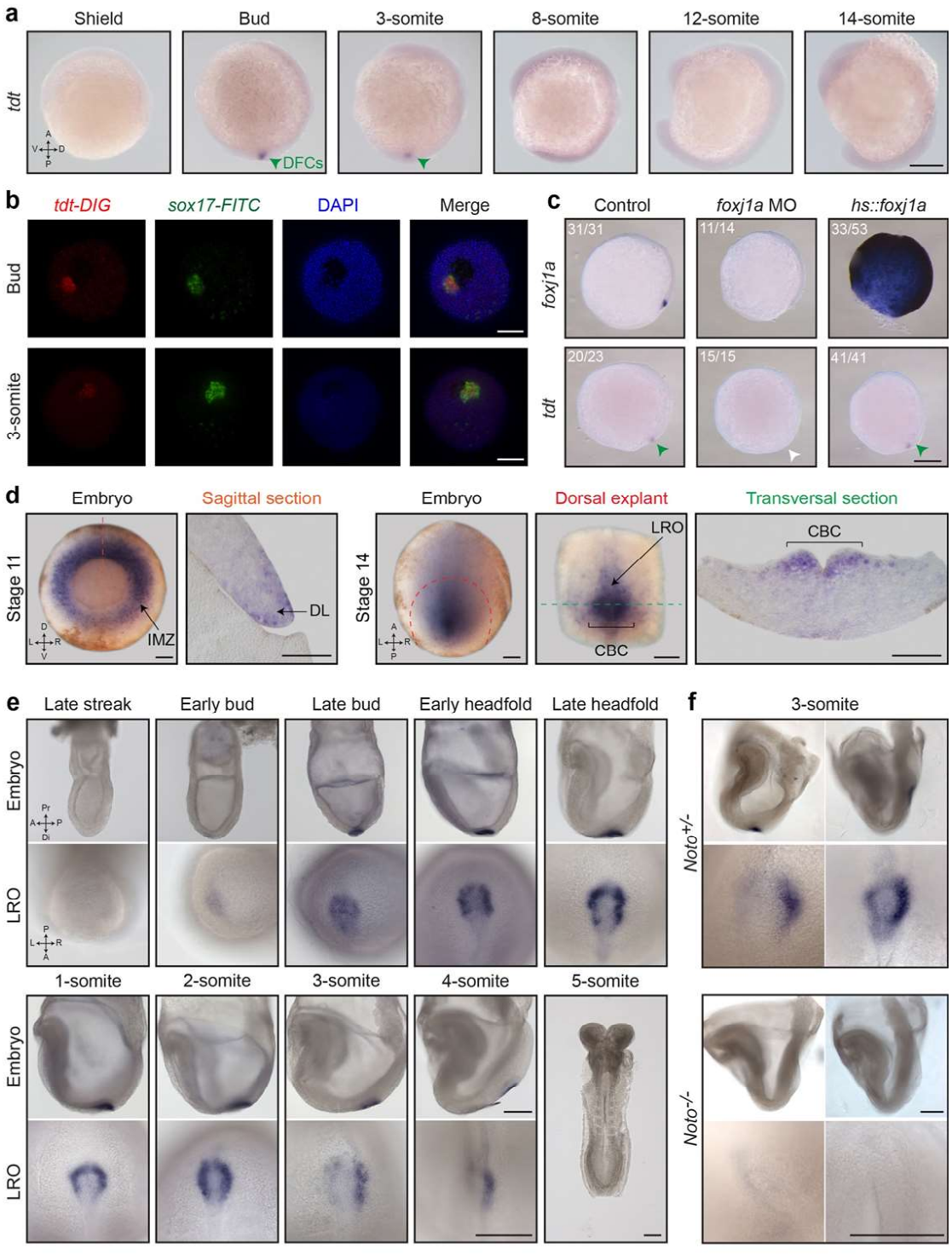


Figure 1 **E. Szenker-Ravi *et al.*, (2020)**

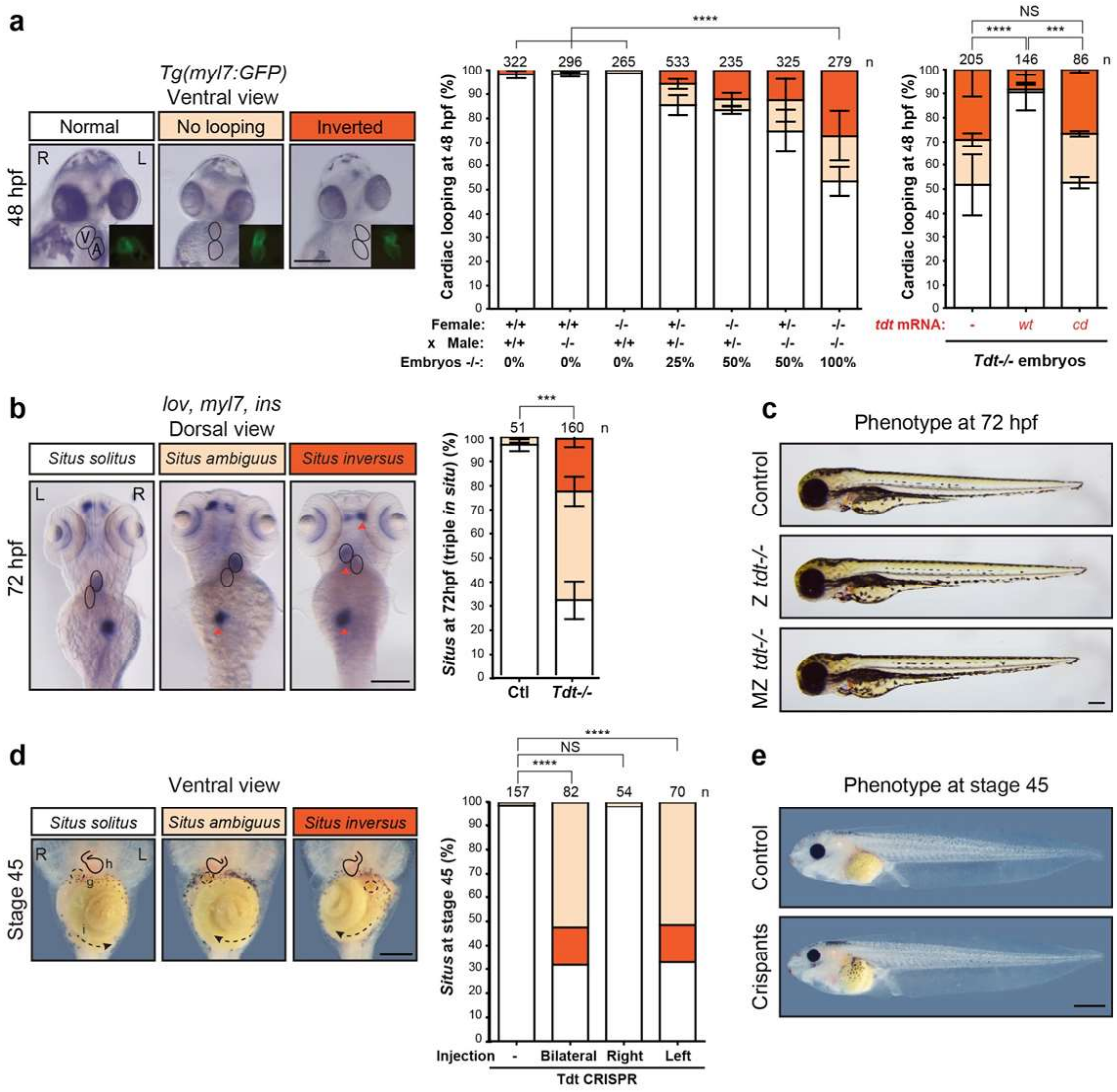
Figure 1: Identification of 5 genes that are mutated or lost in species with a non-ciliated LRO. (a) Simplified evolutionary tree of vertebrate species. Groups of species with a ciliated LRO (black) or a non-ciliated LRO (red and orange) are outlined. Red crosses indicate where genes specific to the function of ciliated LROs have been lost and orange crosses where genes have accumulated deleterious mutations during evolution. Representative electron microscopy images of a ciliated LRO in a rodent (mouse)[92] and a non-ciliated LRO in a cetartiodactyla (pig)[29] are shown. **(b)** Table indicating genes coding for proteins with a signal peptide and either present or absent in species with a non-ciliated LRO. The references for a role in LR asymmetry and OMIM numbers, when available, are indicated. n.d.: not determined. **(c)** Genomic and protein structures of human TDT. Protein domains are highlighted: signal peptide (SP, gray), cysteine switch (brown), Zn^{2+} catalytic domain (yellow) with a HExxH...H catalytic site (dark yellow) and a met-turn methionine (green), cysteine-rich domain (red), and transmembrane domain (TM, blue). Yellow line: catalytic site, green line: met-turn. Chr.: chromosome.



1153 **Figure 2 E. Szenker-Ravi *et al.*, (2020)**

1154

Figure 2: *Tdt* is expressed in the LRO of zebrafish, *Xenopus* and mouse embryos. (a-c) Whole mount *in situ* hybridization in zebrafish embryos at indicated stages. **(a)** *tdt* expression throughout development. The green arrowheads point to expression in the dorsal forerunner cells (DFCs). A: anterior, P: posterior, V: ventral, D: dorsal. Scale bar: 0.1 mm. **(b)** Double fluorescent *in situ* hybridization for *tdt* (DIG) and *sox17* (FITC). Scale bars: 0.1 mm. **(c)** *tdt* and *foxj1a* expression in control, *foxj1a* morphants (MO) and transgenic *hs::foxj1a* heat-shocked zebrafish embryos at 90% epiboly. Arrowheads point to the DFCs with (green) or without (white) *tdt* expression. The numbers of analyzed embryos are indicated. Scale bar: 0.1 mm. **(d)** Whole mount *in situ* hybridization in *Xenopus laevis* embryos at indicated stages. Colored dotted lines indicate the respective sections. D: dorsal, V: ventral, L: left, R: right, A: anterior, P: posterior, IMZ: involuting marginal zone, DL: dorsal lip of the blastopore, CBC: circumblastoporal collar, LRO: left-right organizer. Scale bar: 0.2 mm. **(e-f)** Whole mount *in situ* hybridization in mouse embryos at indicated stages. **(e)** *Tdt* expression in wildtype embryos throughout development. Whole embryos and zoomed-in pictures of the LRO are shown. Pr: proximal, Di: distal, A: anterior, P: posterior. Scale bars: 0.2 mm. **(f)** *Tdt* expression in *Noto*$^{+/-}$ and *Noto*$^{-/-}$ mouse embryos. Two representative embryos and LROs for each genotype are shown. Scale bars: 0.2 mm.

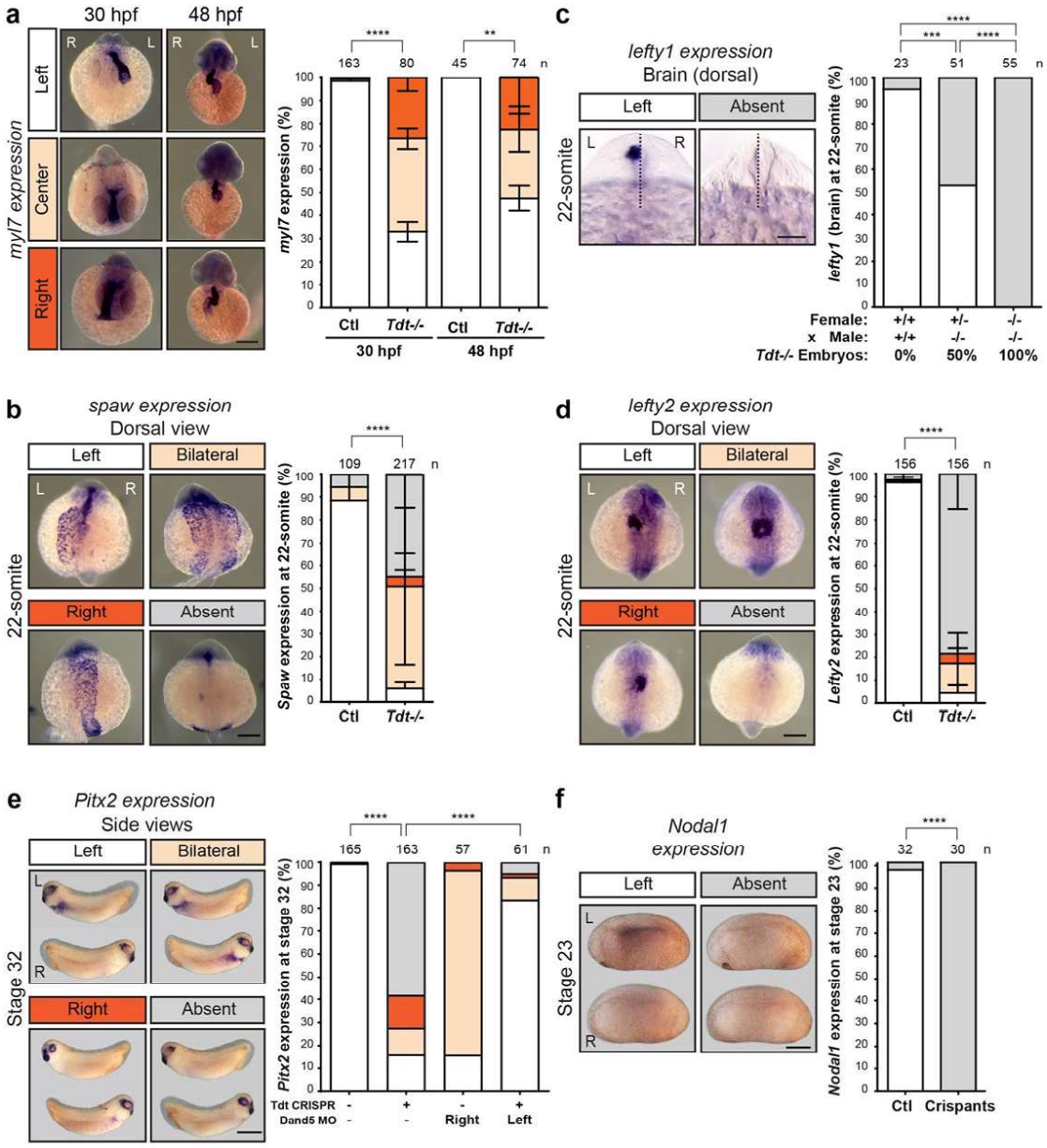


1173 **Figure 3 E. Szenker-Ravi *et al.*, (2020)**

1174

Figure 3: Zebrafish *tdt*$^{-/-}$ mutants and *Xenopus tdt* crispants present with LR asymmetry defects. (a-c) Zebrafish *tdt*$^{-/-}$ mutant embryos present with heterotaxy. **(a)** Left: Cardiac looping of embryos can be visualized at 48 hpf by eye (delineated with a black line) or with the *Tg(myl7:GFP)* transgenic line (inserts). Scale bar: 0.1 mm. R: right, L: left, V: ventricle, A: atrium. Normal, no looping or inverted looping are classified in white, light orange and dark orange, respectively. Middle: The genotype of the parents (female and male) and the resulting percentage of *tdt*$^{-/-}$ embryos are indicated for each type of cross under the graph. Cardiac looping of embryos at 48 hpf (percentage) is indicated. Right: *tdt*$^{-/-}$ embryos were injected with wildtype (*wt)*, or catalytic dead (*cd*, HExxH→AAxxA mutation) *tdt* mRNA at the one-cell stage, and cardiac looping was scored at 48 hpf. n: number of embryos analyzed from at least 3 independent experiments. NS: not significant, ****P<0.0001, ***P<0.001, Two-way ANOVA with Tukey test for multiple comparisons for the normal looping condition. **(b)** Triple whole mount *in situ* hybridization for *lov* (brain), *myl7* (heart) and *ins* (pancreas) in 72 hpf embryos. Orange arrowheads indicate signals in the wrong position (heterotaxia). Representative pictures are shown. *Situs ambiguus* (light orange) indicates any intermediate state between *situs solitus* (white) and *situs inversus* (dark orange) with at least one organ in the wrong position. n: number of embryos analyzed from at least 3 independent experiments. ***P<0.001, Two-way ANOVA with Sidak test for multiple comparisons for the *situs solitus* condition. L: left, R: right, scale bar: 0.1 mm. **(c)** Absence of external phenotype in 72 hpf embryos. Z: zygotic, MZ: maternal zygotic, scale bar: 0.1 mm. **(d-e)** *tdt* is specifically required on the left side of the LRO for proper LR asymmetry in *Xenopus laevis*. **(d)** Left: Representative images of the positioning of internal organs of stage 45 *Xenopus* larvae. The heart (h) and gallbladder (g) are delineated with black and dotted lines, respectively. The rotation of the intestine (i) is indicated with a dotted arrow. *Situs ambiguus* (light orange) indicates any intermediate state between *situs solitus* (white) and *situs inversus* (dark orange) with at least one organ in the wrong position. Scale bar: 0.5 mm. Right: *tdt* CRISPR/Cas9 was injected bilaterally, on the right or on the left at the

4-cell stage and *situs* was analyzed at stage 45. n: number of embryos analyzed. NS: not significant, ****P<0.0001, Chi-square test. **(e)** Absence of external phenotype at stage 45. Scale bar: 1 mm.



1206 **Figure 4 E. Szenker-Ravi *et al.*, (2020)**

1207

Figure 4: The Tdt requirement on the left side of the LRO can be bypassed by Dand5 depletion. (a-d) Whole mount *in situ* hybridization in zebrafish embryos at indicated stages and quantification. **(a)** *myl7* expression in the heart with normal (left), no (center) or inverted (right) jogging / looping are classified in white, light orange and dark orange, respectively. ****P<0.0001, **P<0.01 Two-way ANOVA with Sidak test for multiple comparisons for the normal condition. *spaw* **(b)** and *lefty2* **(d)** expression on the left, bilateral, on the right, or absent are classified in white, light orange, dark orange and gray, respectively. ****P<0.0001, Two-way ANOVA with Sidak test for multiple comparisons for the expression on the left. **(c)** *lefty1* expression on the left or no expression are classified in white and gray, respectively. The genotype of the parents (female and male) and the resulting percentage of $tdt^{-/-}$ embryos are indicated for each type of crosses under the graph. ****P<0.0001, ***P<0.001, Chi-square test. **(e-f)** Whole mount *in situ* hybridization in *Xenopus laevis* embryos at indicated stages and quantification. **(e)** Embryos were injected bilaterally at the 1- and 4-cell stage with *tdt* CRISPR/Cas9 and *dand5* morpholino (MO) on the right or on the left, respectively, as indicated. *pitx2* expression on the left, bilateral, right or absent are classified in white, light orange, dark orange, and gray, respectively. ****P<0.0001, Fisher's exact test. **(f)** Embryos were injected at the 1-cell stage with *tdt* CRISPR/Cas9. *nodal1* expression on the left or no expression are classified in white and gray, respectively. ****P<0.0001, Fisher's exact test. **(a-f)** L: left, R: right, scale bar: 0.1 mm. n: number of embryos analyzed from at least 3 independent experiments.

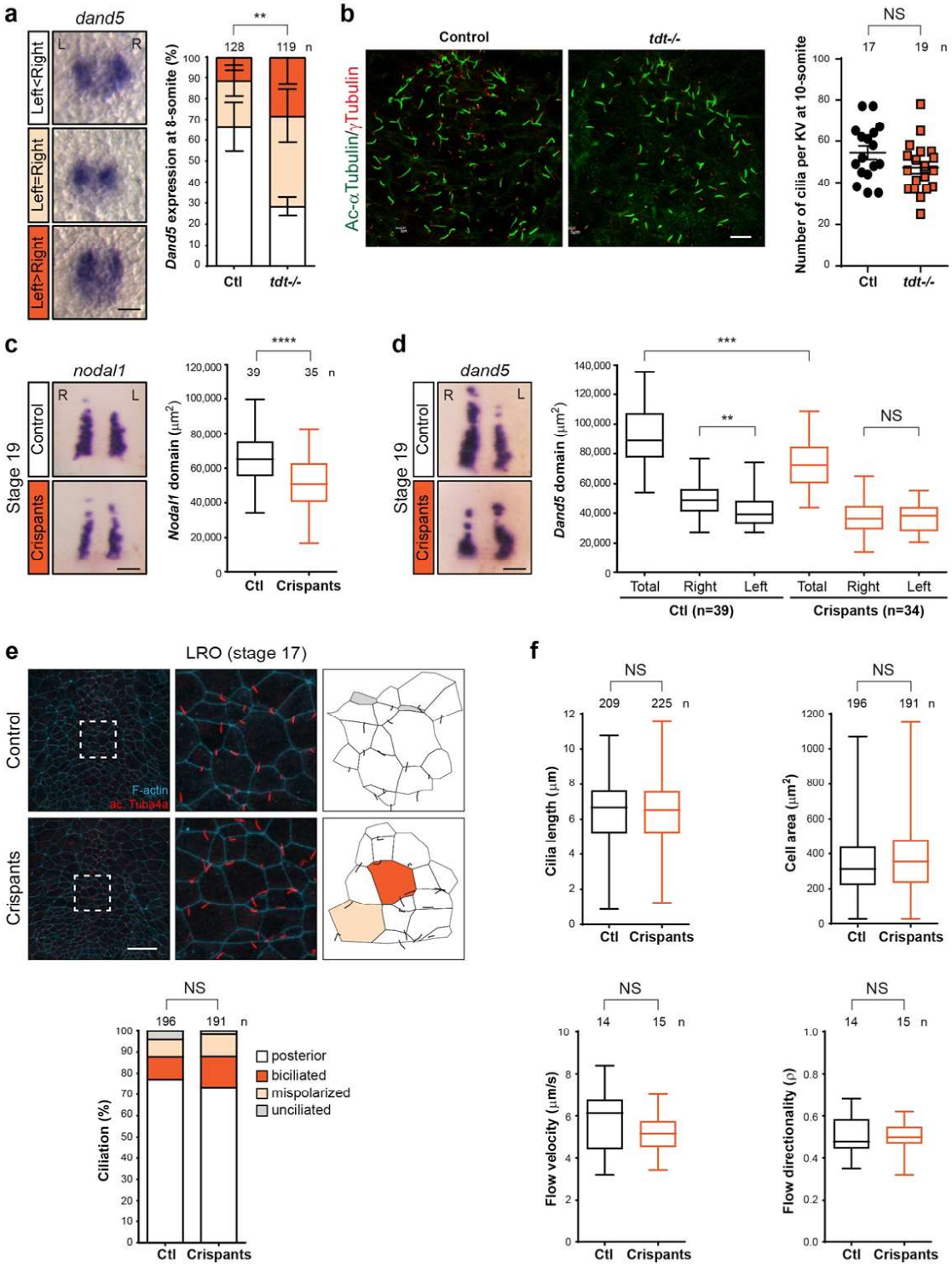


Figure 5 E. Szenker-Ravi *et al.*, (2020)

Figure 5: Tdt is required on the left side of the LRO for proper *dand5* downregulation downstream of leftward flow. (a-c) Analysis of *dand5* expression and ciliogenesis in control (ctl) and *tdt*$^{-/-}$ mutant zebrafish embryos. **(a)** *dand5* expression at the 8-somite stage. Stronger signal on the right is shown in white (normal condition), same intensity on both sides in light orange and stronger signal on the left in dark orange. **P<0.01, Two-way ANOVA with Sidak test for multiple comparisons for the normal condition. n= number of embryos. L= left, R: right, scale bar: 0.05 mm. **(b)** Representative images of an immunofluorescence for acetylated-αTubulin (green) and ɣTubulin (red) marking the axonemes and basal bodies of cilia, respectively, on KV at the 10-somite stage. Numbers of cilia per KV are plotted. NS: non-significant, unpaired t-test. n= number of embryos. KV: Kupffer's vesicle, scale bar: 10 µm. **(c-f)** Analyses in *Xenopus laevis* control (ctl) and *tdt* crispants. **(c-d)** Expression of *nodal1* **(c)** and *dand5* **(d)** in control (white) and *tdt* crispants (dark orange) at stage 19. NS: non-significant, ****P<0.0001, ***P<0.001, **P<0.01, unpaired t-test. n= number of embryos. R: right, L=left, scale bar: 0.1 mm. **(e)** Immunofluorescence on *Xenopus* LRO for acetylated-α-tubulin4a (red) and F-actin (cyan) marking the cilia and cell membrane, respectively. Cells with a single posterior-oriented cilium are shown in white, and cells that are bi-ciliated, unciliated or with a mispolorized cilium are represented in dark orange, gray and light orange, respectively. n= number of cells analyzed. NS: non-significant, Chi-square test. LRO: left right organizer, scale bar: 50 µm. **(f)** Measurements of cilia length (n= number of cilia), cell area (n= number of cells), and flow velocity and directionality (n= number of embryos) of control (ctl) and *tdt* crispants LRO. NS: non-significant, unpaired t-test.

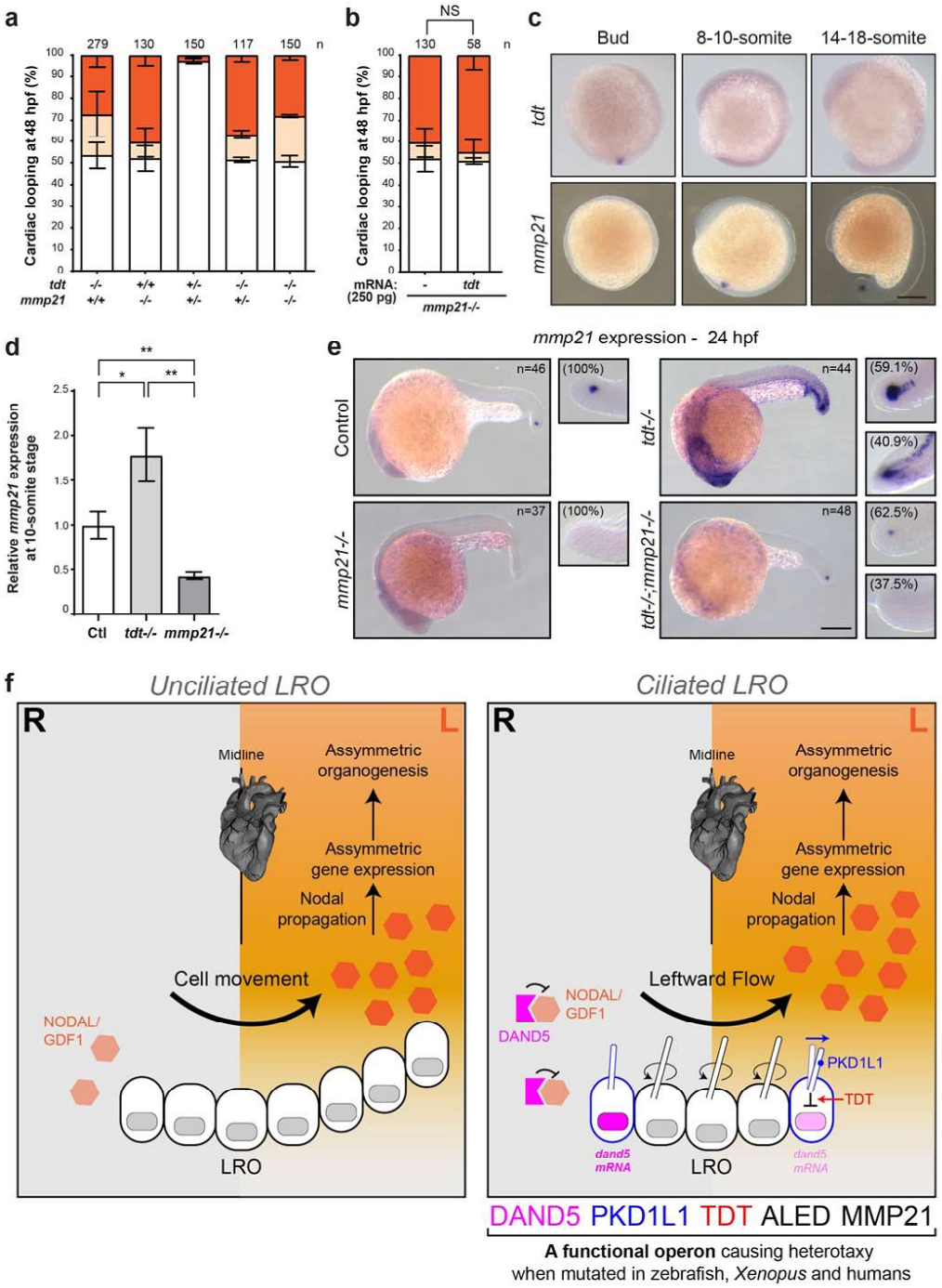


Figure 6 E. Szenker-Ravi *et al.*, (2020)

Figure 6: TDT acts upstream of MMP21. (a-b) Cardiac looping of embryos visualized at 48 hpf is classified as normal, no looping or inverted looping in white, light orange and dark orange, respectively. The genotype of the embryos is indicated under the graph. n: number of embryos analyzed from at least 3 independent experiments. **(a)** Embryos with indicated genotype for *tdt* and *mmp21.* **(b)** *mmp21*$^{-/-}$ embryos injected without or with 250 pg of *tdt* mRNA. **(c)** Whole mount *in situ* hybridization for *tdt* and *mmp21* in zebrafish embryos at indicated stages. DFCs: dorsal forerunner cells, KV: Kupffer's vesicle. Scale bar: 0.1 mm. **(d)** Real-time qPCR for *mmp21* relative to *actin* in zebrafish embryos of indicated genotype at the 10-somite stage. *P<0.05, **P<0.01, Two-way ANOVA with Tukey test for multiple comparisons. **(e)** Whole mount *in situ* hybridization for *mmp21* in zebrafish embryos of indicated genotype at the 24 hpf stage. The number of analyzed embryos is indicated and a zoom with expression in the tail region with percentages of embryos is shown. Scale bar: 0.1 mm. **(f)** Our working model proposes that TDT works in concert with at least 4 partner proteins DAND5, PKD1L1, ALED and MMP21 to form an extracellular module needed to sense and propagate the NODAL/GDF1 signalling cascade on the left. R: right, L: left, LRO: left-right organizer, central (black) and lateral (blue) LRO cells are shown with their motile and immotile cilia, respectively.

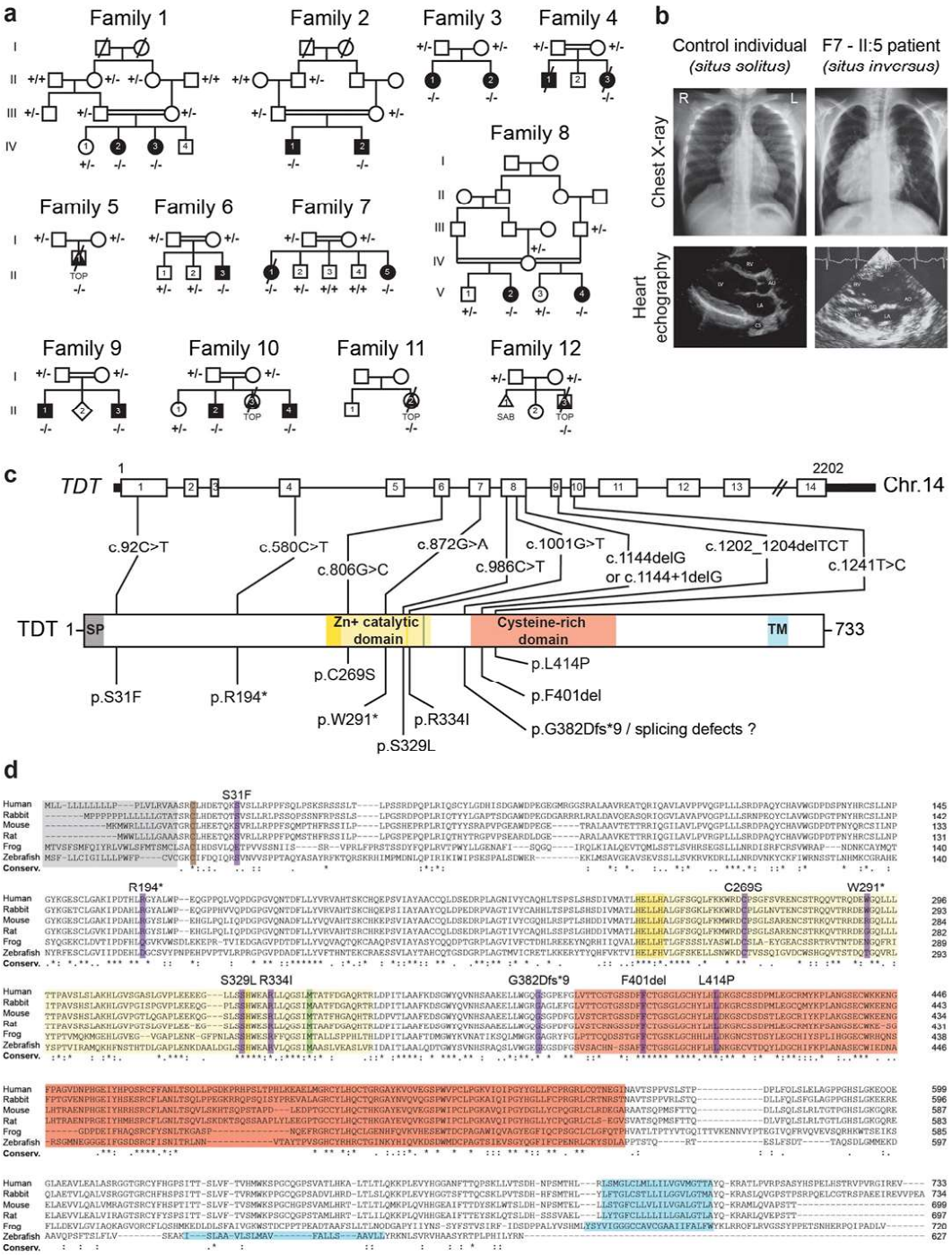


Figure 7 E. Szenker-Ravi *et al.*, (2020)

Figure 7: Identification of recessive *TDT* mutations in patients with heterotaxy associated with congenital heart defects. (a) Pedigrees of twelve families with individuals presenting with heterotaxy. The *TDT* genotype for available individuals are indicated. Squares, circles, diamonds and triangles denote males, females, unknown gender individuals and fetuses, respectively. Open and filled symbols are used for unaffected and affected family members, respectively, and deceased individuals are indicated by a diagonal slash through the symbol. +: wildtype allele, -: mutant allele, TOP: termination of pregnancy, SAB: spontaneous abortion. **(b)** Chest X-ray and heart echography of a control individual and the F7-II:5 patient presenting with *situs inversus* associated with a double outlet right ventricle of the heart. R: right, L: left. **(c)** Genomic and protein structures of human TDT. The position and nature of the identified mutations are indicated. TDT protein domains are highlighted: signal peptide (SP, gray), Zn+ catalytic domain (yellow), cysteine-rich domain (red), and transmembrane domain (TM, blue). Yellow line: catalytic site, green line: met-turn. Chr.: chromosome. **(d)** TDT proteins alignment highlighting the conservation of the affected amino acids in patients with heterotaxy (purple). TDT protein domains are highlighted: signal peptide (gray), Zn+ catalytic domain (yellow), cysteine-rich domain (red), and transmembrane domain (blue). Brown C: cysteine switch, dark yellow amino acids: catalytic site (HExxH...H), green M: met-turn, purple: mutated amino acids. Conserv.: conservation. Software used: Clustal Omega.

Table 1: Clinical characteristics of patients with biallelic *TDT* mutations. Abbreviations: -: absence of phenotype, AS: aortic valvar stenosis; ASD: atrial septal defect; ASD-OP: atrial septal defect ostium primum type; CA: common atrium; CAVC: complete atrioventricular canal; CoA: Aortic coarctation; DIRV: double inlet right ventricle; DORV: double outlet right ventricle; IAA: interrupted aortic arch; IVC: inferior vena cava; n.d.: not determined; PA: pulmonary atresia; PAPVC: partial anomalous pulmonary venous connection; PDA: patent ductus arteriosus; PFO: patent foramen ovale; PS: pulmonary stenosis; SVC: superior vena cava; TA: tricuspid valve atresia; TGA: transposition of the great arteries; TOP: termination of pregnancy; VSD: ventricular septal defect.

Supplementary figures and legends

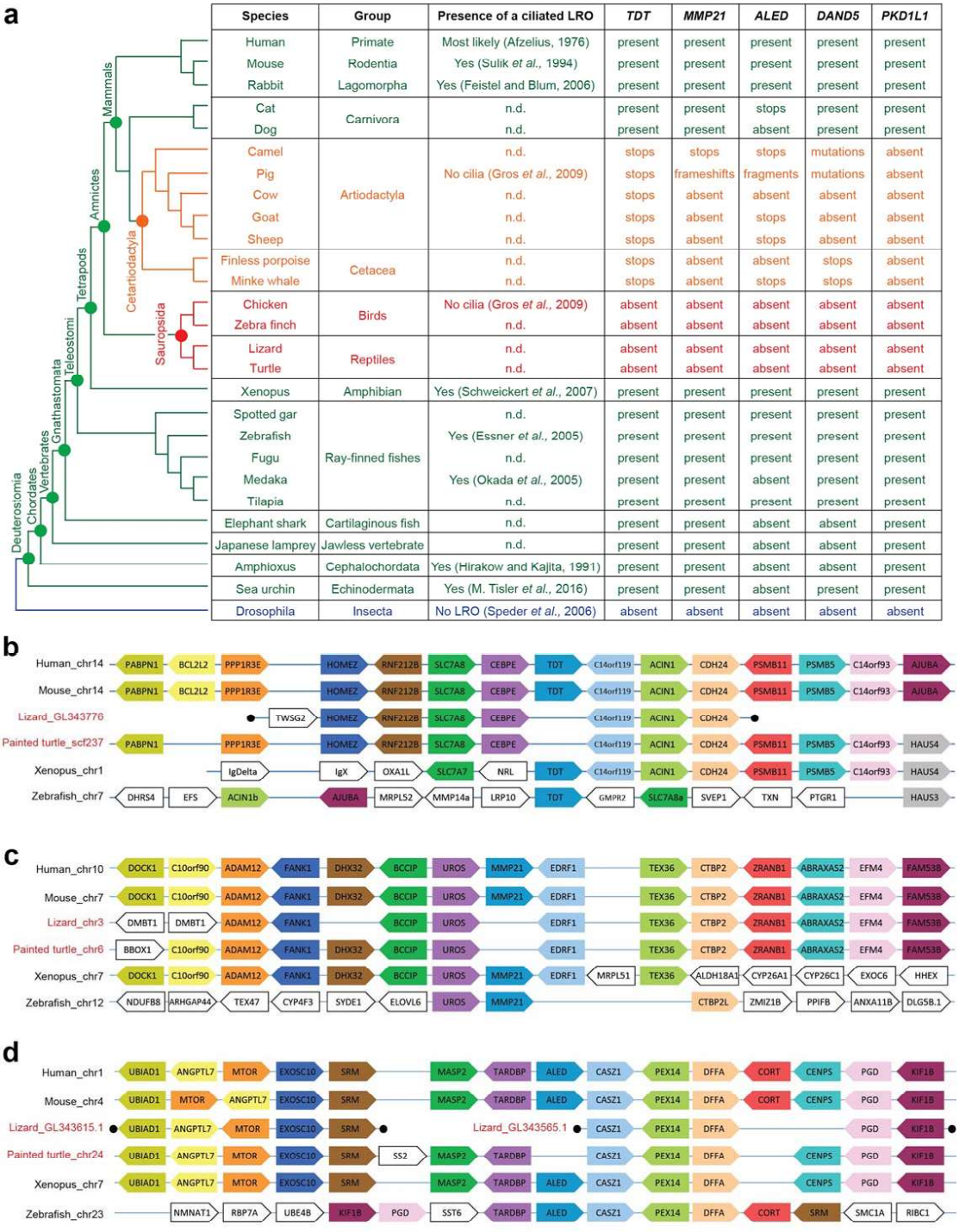


Figure S1 E. Szenker-Ravi *et al.*, (2020)

Figure S1: *TDT, MMP21, ALED*, *DAND5* and *PKD1L1* are specifically lost in species with a non-ciliated LRO. (a) Table indicating the presence of the *TDT, MMP21, ALED*, *DAND5* and *PKD1L1* genes in metazoan species in relation to the presence or absence of a ciliated LRO. It is believed that a ciliated LRO evolved in deuterostomia (green), but cilia in the LRO disappeared in cetartiodactyla (orange) and sauropsida (red). LRO: Left-right organizer, n.d.: not determined. **(b-d)** Representation of the *TDT* **(b)**, *MMP21* **(c)** and *ALED* **(d)** loci in representative jawed-vertebrates whose genomes have been sequenced, showing the loss of the 3 genes in lizards and turtles (reptilia). Block arrows represent genes with the direction of arrows denoting transcriptional orientation. Orthologous genes are shown with the same colour. Black circles represent ends of scaffolds.

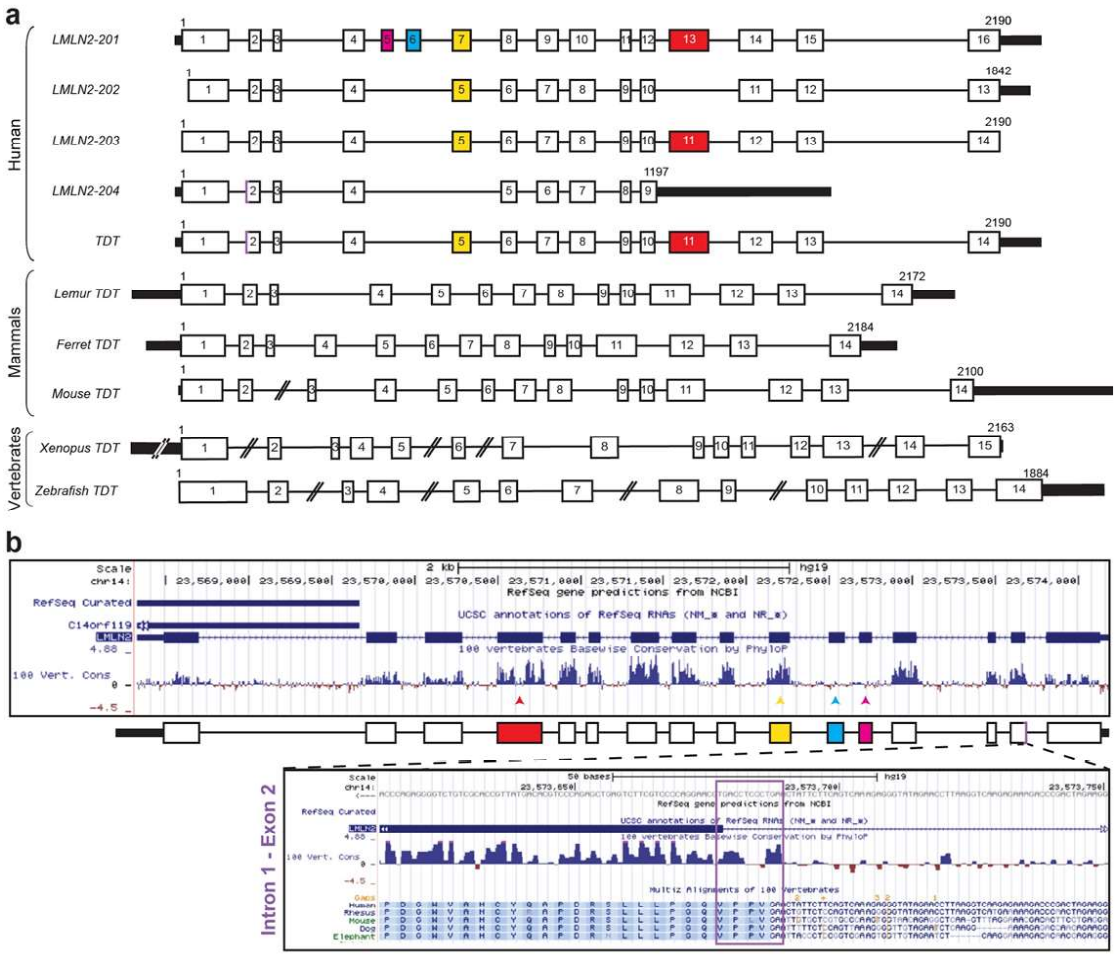


Figure S2 E. Szenker-Ravi *et al.*, (2020)

Figure S2: Genomic structure of *TDT.* (a) Genomic exon-intron structure of four distinct variants of human TDT (*LMLN2)* from Ensembl, and the predicted full length *TDT* as compared to that of other primates and vertebrates as indicated. The alternative exons are color coded as follows: pink and blue exons are only present in *LMLN2-201*, yellow exon is missing in *LMLN2-204* and red exon is missing in *LMLN2-202*. Full length human *TDT* corresponds to *LMLN2-203* with three extra amino acids in the beginning of exon 2 (purple) that are present in *LMLN2-204* (see **b**). The UTRs are highlighted in dark gray. **(b)** Genomic region on the UCSC browser containing *TDT* with the conservation in vertebrates, showing that the pink and blue exons are not conserved while the yellow and red ones are highly conserved. A zoom in the intron1-exon2 junction region confirms the conservation of the 3 additional amino acids (purple box) present in *LMLN2-204.*

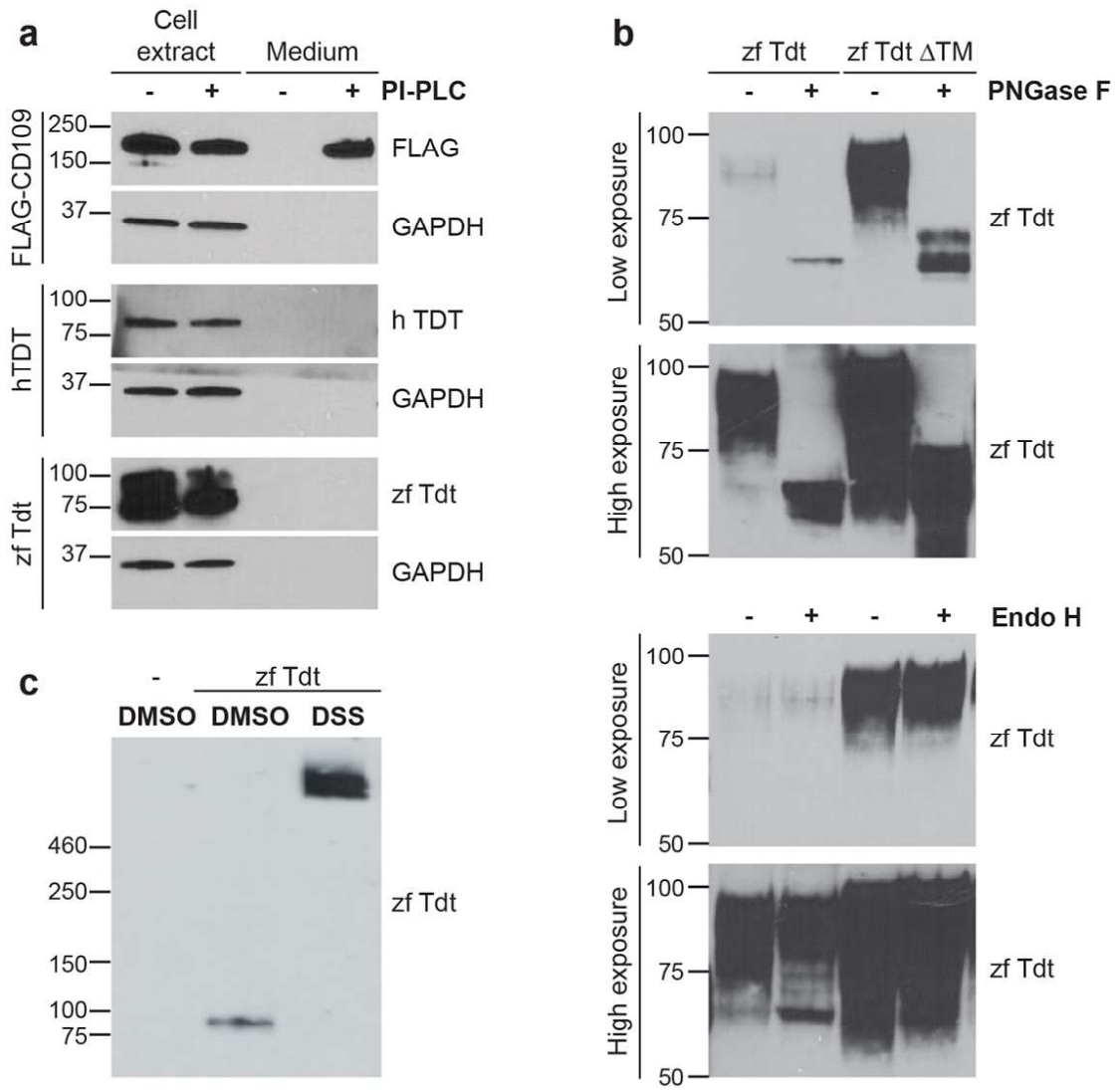
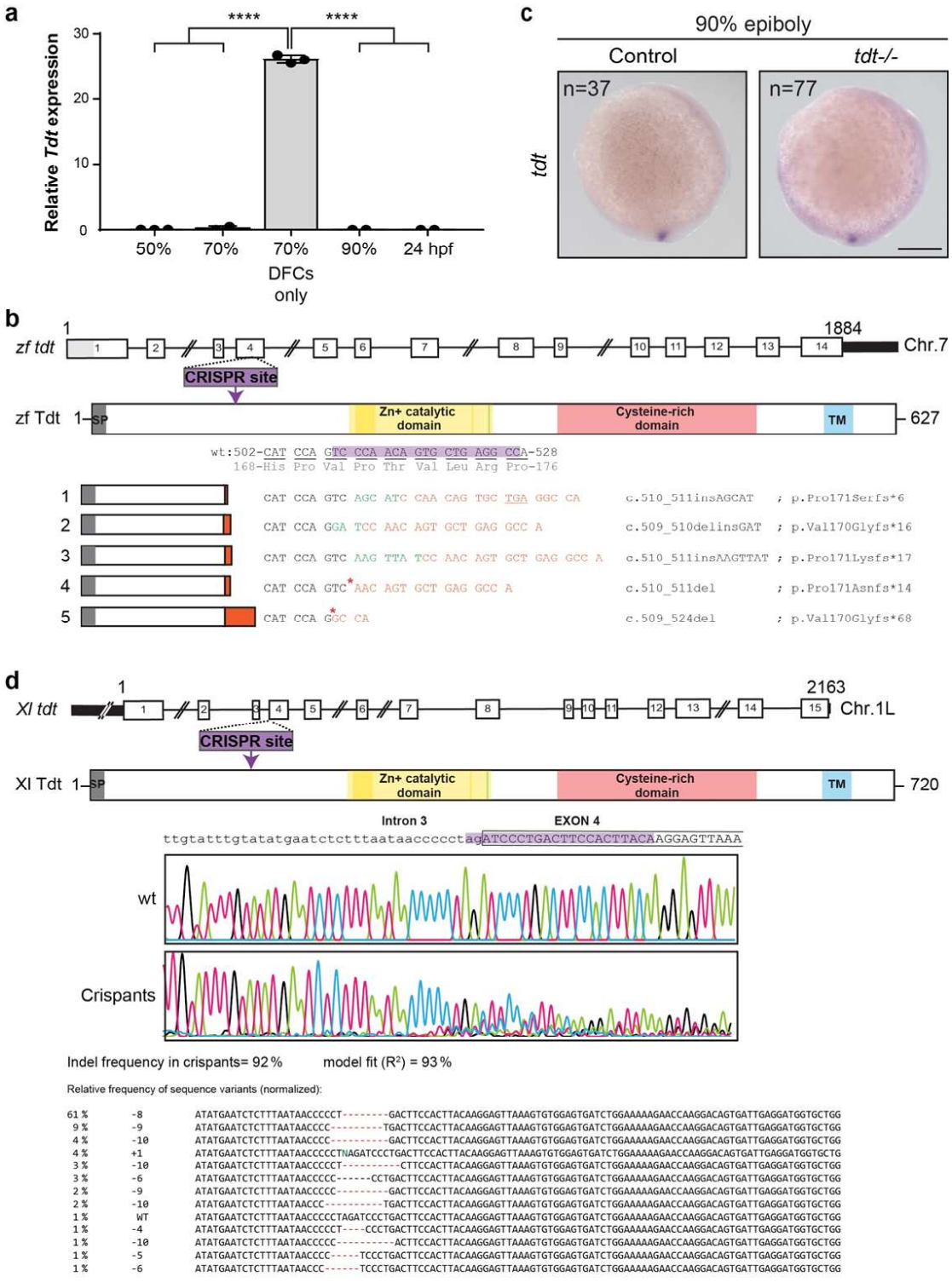

a
Cell extract
Medium
- + - +
PI-PLC
FLAG-CD109
250
150
37
FLAG
GAPDH
hTDT
100
75
37
h TDT
GAPDH
zf Tdt
100
75
37
zf Tdt
GAPDH
b
zf Tdt
zf Tdt ΔTM
- + - +
PNGase F
Low exposure
High exposure
100
75
50
zf Tdt
- + - +
Endo H
c
-
zf Tdt
DMSO DMSO DSS
460
250
150
100
75
zf Tdt

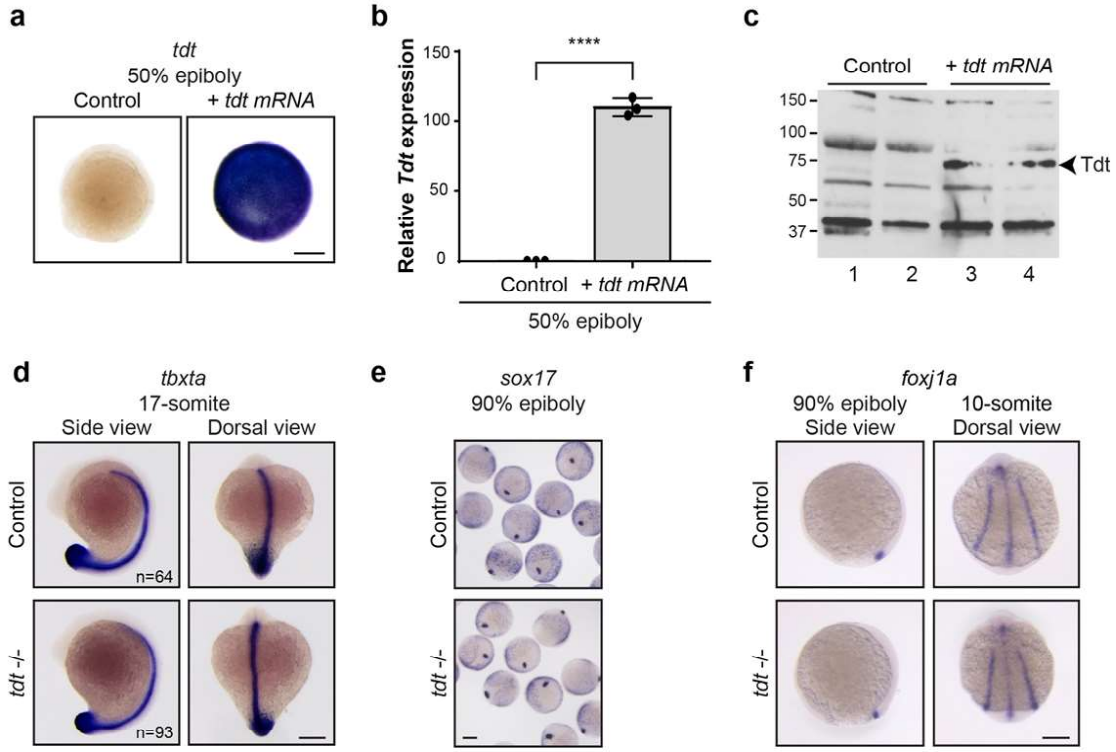
Figure S3: TDT is post-translationally modified but not GPI-anchored. (a) Western blot using cell extracts and conditioned media of HEK293T cells transfected with human *TDT*, zebrafish *Tdt* or *FLAG-CD109* and treated with or without PI-PLC. FLAG-CD109 serves as a positive control of a GPI-anchored protein. **(b)** Western blot using conditioned media from HEK293T cells transfected with zebrafish *Tdt* or zebrafish *Tdt* encoding a variant lacking the transmembrane domain (ΔTM) treated with or without PNGase F and Endo H, as indicated. **(c)** Western blot using conditioned media from HEK293T cells transfected with zebrafish Tdt and crosslinked in solution with DSS **(a-c)**. Primary antibodies used are indicated, sizes are in kDa.



1348 **Figure S4 E. Szenker-Ravi *et al.*, (2020)**

1349

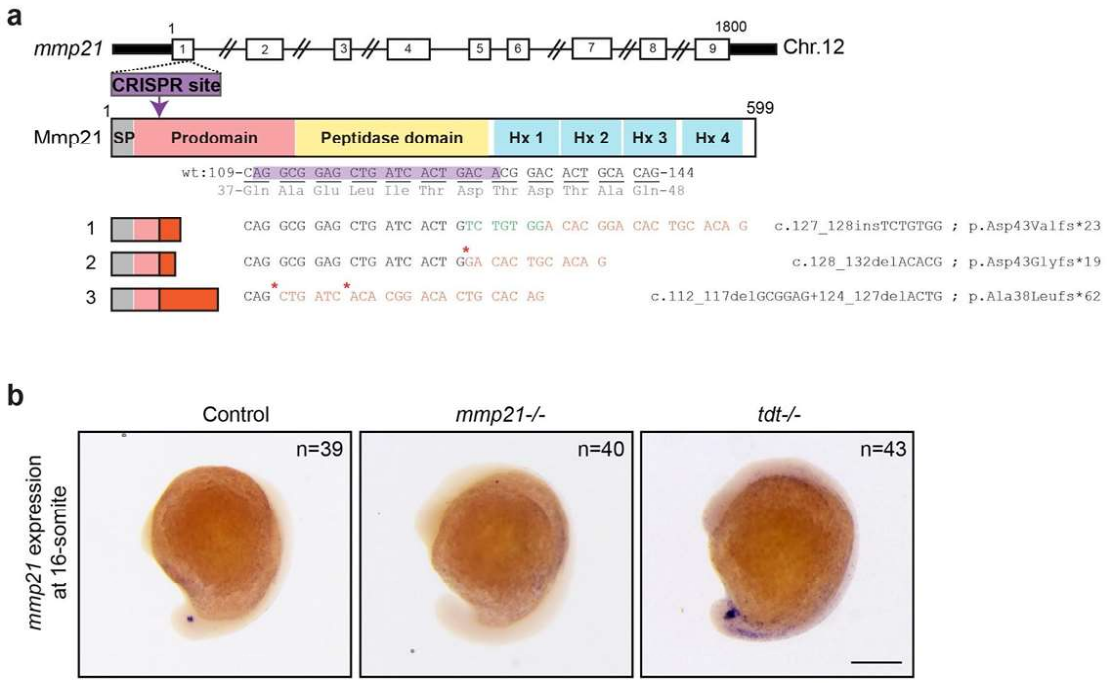
Figure S4: CRISPR-Cas9 knockout of Tdt in zebrafish and *Xenopus laevis*. (a) Real-time qPCR for *tdt* relative to *actin* in zebrafish embryos at indicated stages. *tdt* expression is only detectable using RNA extracted from tissue at the dorsal posterior side containing the DFCs. ****P<0.0001, Two-way ANOVA with Tukey test for multiple comparisons. DFC: dorsal forerunner cells. n= number of embryos. **(b)** Depiction of genomic and protein structures of zebrafish Tdt. The sequence used is XM_002662823 using a START codon located 168 nucleotides upstream of the original one (light gray). Tdt protein domains are highlighted: the signal peptide (gray), the Zn+ catalytic domain (yellow), the cysteine-rich domain (red), and the transmembrane domain (blue). Yellow line: catalytic domain, green line: met-turn. The site and targeted sequence by the CRISPR gRNA is indicated in purple. Five different alleles were obtained with indicated mutations (green: insertion, red star: deletion), all leading to a frameshift (orange) with an early stop codon (underlined for mutation 1). While all mutations lead to the same LR phenotype in the homozygous state, line 1 was used for further investigation. **(c)** Whole mount *in situ* hybridization for *tdt* in 90% epiboly control and *tdt*$^{-/-}$ zebrafish embryos. Scale bar: 0.1 mm. **(d)** Depiction of genomic and protein structures of *Xenopus laevis* Tdt. A CRISPR gRNA was designed in the beginning of exon 4 (purple), leading to mutations with 92% efficiency. Chr.: chromosome, SP: signal peptide, TM: transmembrane domain, wt: wildtype, Xl: *Xenopus laevis*.



1368 **Figure S5 E. Szenker-Ravi *et al.*, (2020)**

1369

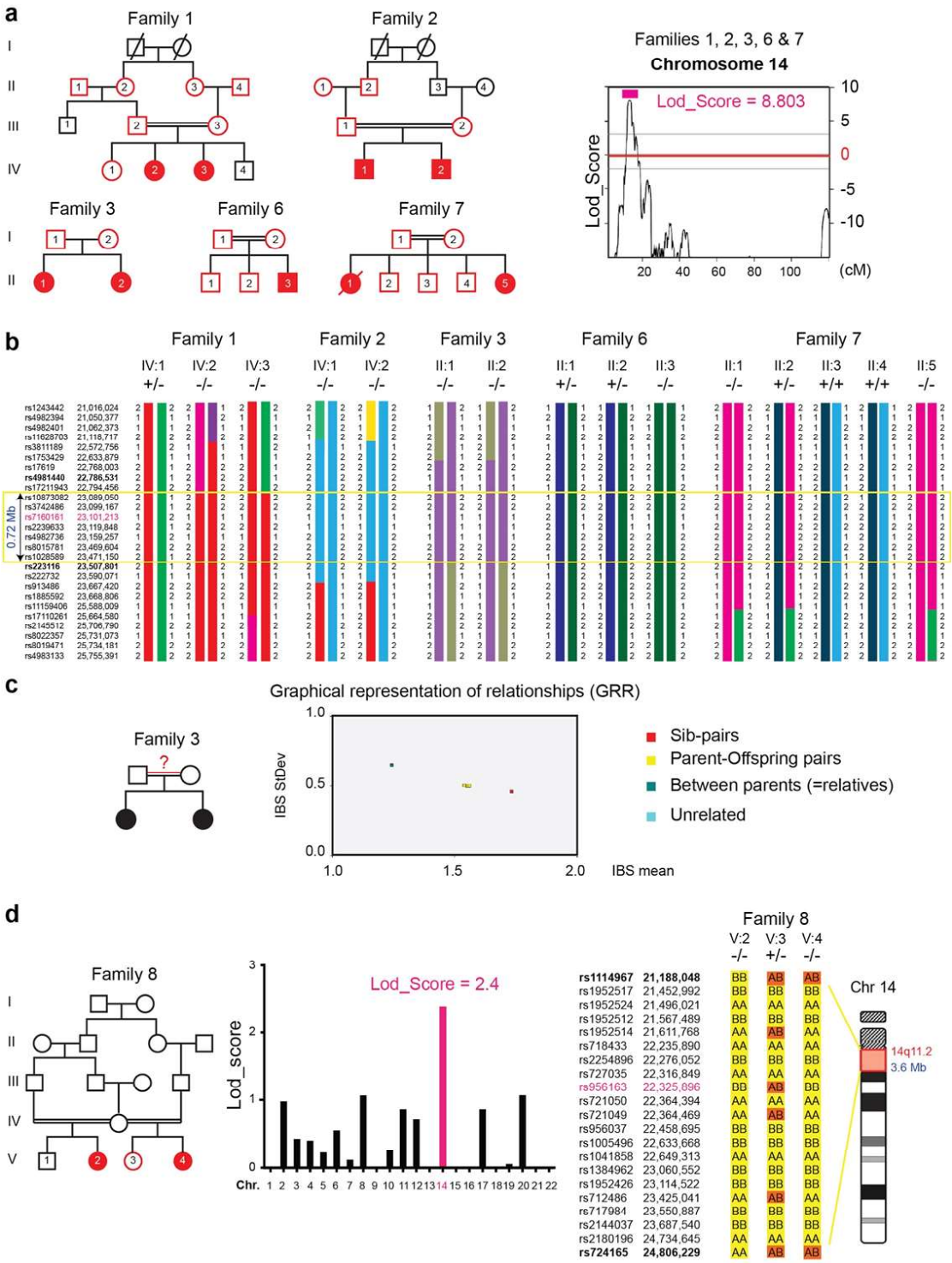
Figure S5: The absence of Tdt does not affect the DFCs or midline patterning. (a-c) Analysis of zebrafish embryos injected or not with *tdt* mRNA by whole mount *in situ* hybridization **(a)**, real-time qPCR **(b)** and western blotting **(c)**. Scale bar: 0.1 mm, ****P<0.0001, unpaired t-test. **(d-f)** Whole mount *in situ* hybridization in control and *tdt*$^{-/-}$ zebrafish embryos for *tbxta* **(d)**, *foxj1a* **(e)**, and *sox17* **(f)** at indicated stages. Scale bars: 0.1 mm.



1376 **Figure S6 E. Szenker-Ravi *et al.,* (2020)**

1377

Figure S6: Generation of *Mmp21* mutant zebrafish. (a) Depiction of genomic and protein structures of zebrafish Mmp21. MMP21 protein domains are highlighted: signal peptide (gray), prodomain (pink), peptidase domain (yellow) and hemopexin (hx) repeats (blue). The site and targeted sequence by the CRISPR gRNA is indicated in purple. F1 fish were a gift from Katsanis laboratory. Three different alleles were obtained with indicated mutations (green: insertion, red star: deletion), all leading to a frameshift (orange) with an early stop codon. While all mutations lead to the same LR phenotype in the homozygous state, line 3 was used for further investigation. **(b)** Whole mount *in situ* hybridization for *mmp21* in zebrafish embryos of indicated genotypes at the 16-somite stage. The numbers of analyzed embryos are indicated. Scale bar: 0.1 mm.



1388 **Figure S7 E. Szenker-Ravi *et al.*, (2020)**

1389

Figure S7: Mapping analysis of 6 families with patients presenting with heterotaxy reveal a region of homozygosity on chr. 14. (a-b) Individuals of Families 1, 2, 3, 6 and 7 with red symbols in pedigrees were SNP genotyped. Genotypes were used for linkage analysis which revealed a common region of homozygosity for affected individuals on chromosome 14, including the genomic region of *TDT*, with a total Lod_Score of 8.803. Graphical representation obtained using Merlin Autosome. **(b)** Haplotypes obtained with Merlin and presented with Haplopainter revealed a common region of homozygosity on chromosome 14q11.2 delimited by rs17211943 and rs223116 that is 0.72 Mb long. **(c)** Graphical representation of relationships (GRR) showing that parents of Family 3 are actually related. **(d)** Family 8 SNP genotyping of individuals V:2, V:3, V:4 (red in pedigree) reveal a Lod_Score max on chromosome 14q11.2 with SNP rs956163 (pink) for affected individuals. The region of homozygosity shared by V:2 and V:4 is 3.6 Mb long, delimited by rs1114967 and rs724165.

Table S1: List of the cases in the original cohort screened for *TDT* mutations.

Table S2: List of genes with a *S* score above 1.

Movie S1: Cilia movement in KV of a WT zebrafish embryo.

Movie S2: Cilia movement in KV of a *tdt*$^{-/-}$ zebrafish embryo.

Movie S3: Movement of endogenous particles in KV of a WT zebrafish embryo.

Movie S4: Movement of endogenous particles in KV of a *tdt*$^{-/-}$ zebrafish embryo.

Movies S1-S4: 8-10 somite stage zebrafish embryos were dechorionated and mounted in 2% low melting agarose, and KV were imaged with a 60 x water lens from the dorsal posterior end.

Movie S5: Analysis of leftward flow in the LRO of *Xenopus* wildtype and *Tdt* crispants. Dorsal explants were prepared and fluorescent micro beads added. Time-lapse movies were captured at 2 fps and transformed to gradient-time-trails (GTTs). GTT movies play at about 10 x real time and display bead transport from the right to the left side of the LRO. Note that velocity and directionality of bead transport were identical in WT (left) and *Tdt* crispant (right) specimens.

	Family 1		Family 2		Family 3		Family 4		Family 5	Family 6	Family 7		Family 8		Family 9		Family 10			Family 11	Family 12
	IV:2	IV:3	IV:1	IV:2	II:1	II:2	II:1	II:3	II:1	II:3	II:1	II:5	V:2	V:4	II:1	II:3	II:2	II:3	II:4	II:2	II:3
Gender	female	female	male	male	female	female	male	female	male	male	female	female	female	female	male	male	male	female	male	female	male
Year of birth	1989	1991	1987	1992	1984	1989	1985	1990	2014	1991	1983	1992	1989	2006	1997	2000	2005	2007	2009	2015	2015
Age at death	alive	alive	n.d.	n.d.	alive	alive	deceased (2 months)	deceased (10 years)	TOP	n.d.	deceased (9 years)	alive	alive	alive	n.d.	n.d.	alive	TOP	alive	TOP	TOP
Country of origin	Turkey		Turkey		Italy (South Tyrol)		Turkey		France	Turkey	Turkey		France		Lebanon		Lebanon			France	France
Consanguineous parents	yes		yes		not initially reported		yes		not reported	yes	yes		yes		yes		yes			not reported	no
***TDT* DNA**	c.986C>T		c.92C>T		c.872G>A		c.1001G>T		c.92C>T	c.1241T>C	c.1001G>T		c.809C>T		c.1202_1204delTCT		c.1144delG / c.1144+1delG			c.92C>T	c.92C>T (paternal) & c.580C>T (maternal)
TDT protein	p.Ser329Leu		p.Ser31Phe		p.Trp291*		p.Arg334Ile		p.Ser31Phe	p.Leu414Pro	p.Arg334Ile		p.Cys269Ser		p.Phe401del		p.Gly382Aspfs*8 / p.? (splicing defects?)			p.Ser31Phe	p.Ser31Phe & p.Arg194*
Mutation status	homozygous		homozygous		homozygous		no DNA available (parents heterozygous)	homozygous	homozygous	homozygous	homozygous	homozygous	homozygous		homozygous		homozygous	no DNA available (parents heterozygous)	homozygous	homozygous	compound heterozygous
Heterotaxy	***Situs inversus***	***Situs inversus***	***Abdominal situs ambiguus***	***Situs solitus***	***Situs solitus***	***Situs solitus***	***Situs solitus***	***Situs ambiguus***	***Left isomerism***	***Situs inversus***	***Situs solitus***	***Situs inversus***	***Situs inversus***	***Situs inversus***	***Situs solitus***	***Situs solitus***	***Situs inversus***	***Situs solitus***	***Situs inversus***	***Situs ambiguus***	***Situs ambiguus***
Position of organs	Complete reversal	Complete reversal	Midline liver	Isolated dextrocardia	Isolated dextrocardia	Isolated dextrocardia	Normal	Dextrocardia, Asplenia, midline liver (abdominal right isomerism)	Pulmonary and atrial left isomerism, Polysplenia, absent pancreatic tail	Complete reversal	Normal	Complete reversal	Complete reversal	Complete reversal	Isolated dextrocardia	Isolated dextrocardia	Complete reversal	Normal	Complete reversal	Intestinal malrotation, polysplenia, left pulmonary isomerism	Intestinal malrotation, polysplenia, annular pancreas, left pulmonary isomerism
Cardiac phenotypes	VSD, right SVC to the coronary sinus	Congenitally corrected TGA,VSD, PA	Left IVC with hemiazygous continuation, hepatic vein abnormality,CA, CAVC, DORV, subvalvular pulmonary stenosis	ASD, congenitally corrected TGA, PA, PDA	PFO, DIRV, PA, PDA, right aortic arch	ASD, discordant atrioventricular connection, subaortic VSD, vavular and subvavular PS, PDA	DORV, PDA, IAA after left subclavian artery	Congenitally corrected TGA, ASD+OP, TA, CoA, PDA	PAPVC, CAVC, left ventricle hypoplasia, AS, aortic arch hypoplasia	Common atrium, CAVC	Absent IVC with azygos continuation to right SVC, left SVC to coronary sinus, discordant atrio-ventricular connection	Right SVC to coronary sinus, left SVC to the right atrium, DORV, PS	no CHD	DORV	DORV, PS	no CHD	Left SVC, Common atrium, CAVC, Single Ventricle right type, PS, TGA	CAVC, Single Ventricle left type, hypoplastic pulmonary artery	CA, CAVC, Single ventricle right type, Pulmonary stenosis	Left SVC to the coronary sinus, CAVC, single coronary artery, hypoplastic aortic arch	Absent IVC with hemi-azygos continuation, PAPVC, CA with appendages left isomerism, single ventricle, double aortic arch
Others	n.a.	Pace maker for Atrioventricular block (1998)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	Microretrognathism, low-set ears, receding forehead, 11 rib pairs	n.a.	n.a.	n.a.	n.a.	Right kidney: dilated Left kidney: hyperechoic	n.a.	n.a.	n.a.	n.a.	n.a.	Dolichocephaly, hypertelorism	n.a.

Abbreviations: AS: aortic valve stenosis; ASD: atrial septal defect; ASD+OP: atrial septal defect ostium primum type; CA: common atrium; CAVC: complete atrioventricular canal; CHD: Congenital heart defect; CoA: Aortic coarctation; DIRV: double inlet right ventricle; DORV: double outlet right ventricle; IAA: interrupted aortic arch; IVC: inferior vena cava; n.a.: not applicable; n.d.: not determined; PA: pulmonary atresia; PAPVC: partial anomalous pulmonary venous connection; PDA: patent ductus arteriosus; PFO: patent foramen ovale; PS: pulmonary stenosis; SVC: superior vena cava; TA: tricuspid valve atresia; TGA: transposition of the great arteries; TOP: termination of pregnancy; VSD: ventricular septal defect

Table 1 Szenker-Ravi *et al.*, (2020)