Conserved enhancer logic controls the notochord expression of vertebrate *Brachyury*

Cassie L. Kemmler[1,†], Jana Smolikova[2,†], Hannah R. Moran[1,†], Brandon J. Mannion[3,4], Dunja Knapp[5], Fabian Lim[6,7,8], Anna Czarkwiani[5], Viviana Hermosilla Aguayo[9,10,11,12], Vincent Rapp[13], Olivia E. Fitch[14], Seraina Bötschi[15], Licia Selleri[9,10,11,12], Emma Farley[6,7], Ingo Braasch[14], Maximina Yun[5,16,17], Axel Visel[3,18,19], Marco Osterwalder[3,13,20], Christian Mosimann[1], Zbynek Kozmik[2,*], Alexa Burger[1,*]

[1] Section of Developmental Biology, Department of Pedi*atrics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA; [2] Institute of Molecular Genetics of the ASCR, v. v. i., Prague, Czech Republic; [3] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [4] Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA; [5] Technische Universität Dresden, CRTD/Center for Regenerative Therapies Dresden, Dresden, Germany; [6] Department of Medicine, Health Sciences, University of California San Diego, La Jolla, CA, USA; [7] Department of Molecular Biology, Biological Sciences, University of California San Diego, La Jolla, CA USA; [8] Biological Sciences Graduate Program, University of California San Diego, La Jolla, CA, USA; [9] Program in Craniofacial Biology, University of California San Francisco, San Francisco, CA, USA; [10] Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA; [11] Department of Orofacial Sciences, University of California San Francisco, San Francisco, CA, USA; [12] Department of Anatomy, University of California San Francisco, San Francisco, CA, USA; [13] Department for BioMedical Research (DBMR), University of Bern, Bern, Switzerland; [14] Department of Integrative Biology and Ecology, Evolution and Behavior Program, Michigan State University, East Lansing, MI, USA; [15] Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland; [16] Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany; [17] Cluster of Excellence Physics of Life, Technische Universität Dresden, Dresden, Germany; [18] US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [19] School of Natural Sciences, University of California Merced, Merced, CA, USA; [20] Department of Cardiology, Berne University Hospital, Berne, Switzerland*

[†] Equal contribution

** Correspondence to: zbynek.kozmik@img.cas.cz, alexa.burger@cuanschutz.edu*

ABSTRACT

The cell type-specific expression of key transcription factors is central to development. Brachyury/T/TBXT is a major transcription factor for gastrulation, tailbud patterning, and notochord formation; however, how its expression is controlled in the mammalian notochord has remained elusive. Here, we identify the complement of notochord-specific enhancers in the mammalian *Brachyury/T/TBXT* gene. Using transgenic assays in zebrafish, axolotl, and mouse, we discover three *Brachyury*-controlling notochord enhancers *T3*, *C*, and *I* in human, mouse, and marsupial genomes. Acting as Brachyury-responsive, auto-regulatory shadow enhancers, deletion of all three enhancers in mouse abolishes Brachyury/T expression selectively in the notochord, causing specific trunk and neural tube defects without gastrulation or tailbud defects. Sequence and functional conservation of *Brachyury*-driving notochord enhancers with the *brachyury/tbxtb* loci from diverse lineages of fishes dates their origin to the last common ancestor of jawed vertebrates. Our data define the enhancers for *Brachyury/T/TBXTB* notochord expression as ancient mechanism in axis development.

INTRODUCTION

The defining feature of the chordate body plan is the notochord, a principal structure formed by the axial or chorda mesoderm that provides stability and rigidity along the body axis (Corallo et al., 2015; Stemple, 2005). As mammals form an ossified spine, their notochord progressively regresses and its remnants form the nucleus pulposus within the intervertebral discs (Bagnat and Gray, 2020; Peck et al., 2017; Risbud et al., 2010; Stosiek et al., 1988; Wang et al., 2018). Notochord precursors emerge from the initial organizer and form in a stereotypical rostral-to-caudal trajectory as gastrulation proceeds, manifesting among the earliest visible structures in chordate embryos (Satoh et al., 2012; Stemple, 2005). The deeply conserved T-box transcription factor gene *Brachyury* (also called *T* or *TBXT*) is a key regulator of notochord formation. Originally identified as dominant mutation *T* that caused short tails in mice, *Brachyury* expression and function has been linked to notochord emergence across chordates (Corbo et al., 1997; Dobrovolskaia-Zavadskaia, 1927; Halpern et al., 1993; Herrmann, 1995; Holland et al., 1995; Schulte-Merker et al., 1994; Smith et al., 1991). In addition to its central role in notochord fate specification, the function of vertebrate *Brachyury* is required for proper primitive streak formation, tailbud specification, and subsequent neuromesodermal progenitor control (Henrique et al., 2015; Martin and Kimelman, 2008; Rivera-Pérez and Magnuson, 2005). However, how the expression of this central developmental transcription factor is selectively regulated to achieve its notochord activity in mammals remains unresolved.

The central contribution of the notochord and the tailbud to different morphological adaptions and locomotion strategies shows in the diversification of axial structures across vertebrates (Schwaner et al., 2021). Gain and loss of gene copies and of their associated gene-regulatory elements are major drivers of evolutionary innovation, and the *Brachyury* gene family itself is a prime example of this process. *Brachyury* predates the origin of, and was present as, a single copy gene in the chordate ancestor (Inoue et al., 2017; Sebé-Pedrós et al., 2013). Following

two whole genome duplications in early vertebrates and the subsequent loss of one of four *Brachyury* paralogs, three gene paralogs were present in the jawed vertebrate ancestor: *Tbxta*, *Tbxtb*, and *Tbx19* (Inoue et al., 2017). *Tbxta* became subsequently lost within the tetrapod lineage, resulting in mammals and birds ultimately only retaining *Tbxtb* (commonly called *Brachyury*/*T*/*TBXT* in tetrapods including humans) (Amemiya et al., 2013). In contrast, ray-finned fishes retained both *tbxta*/*ntla* and *tbxtb*/*ntlb*, the latter being the ortholog of the remaining human *Brachyury/T/TBXT* (*de facto TBXTB*) gene (Martin and Kimelman, 2008).

Curiously, *tbxta*/*ntla* has become the predominant functional *Brachyury*/*T*/*TBXT* gene in zebrafish, as documented in classic mutants for *ntla* (*no tail a*) that fail to form a tail and notochord (Halpern et al., 1993; Schulte-Merker et al., 1994). While no mutant for zebrafish *tbxtb*/*ntlb* has been reported to date, morpholino-based knockdown studies indicate that *tbxtb* function adds minimally to the dominant role of zebrafish *tbxta* (Martin and Kimelman, 2008). This variable copy number of *Brachyury* genes across vertebrates came along with selection and divergence of regulatory elements controlling *Brachyury* gene expression during distinct developmental timepoints and cell types. Promoter-proximal regions of *Brachyury* in the *Ciona Brachyury* gene and in the zebrafish *tbxta* gene drive early organizer and notochord activity (Corbo et al., 1997; Harvey et al., 2010). In contrast, the promoter-proximal region called *Tstreak* of *T*/*Tbxtb* in mouse, human, and *Xenopus* has previously been found to drive primitive streak expression in response to canonical Wnt/beta-catenin signaling, yet lacks any notochord-driving activity (Arnold et al., 2000; Clements et al., 1996; Latinkić et al., 1997).

Notably, recent work documented that deleting a large 37 kb-spanning region upstream of mouse *Brachyury*/*T*/*Tbxtb* leads to mutant phenotypes consistent with a selective loss of *Brachyury* notochord expression (Schifferl et al., 2021). A small element termed *TNE* in the 37 kb interval was sufficient to drive specific notochord expression in mouse reporter assays, yet its deletion showed mild to no phenotypic consequences (Schifferl et al., 2021). These pioneering data show that additional regulatory element(s) in addition to *TNE* contribute to *Brachyury/Tbxtb* expression specifically in the notochord. Uncovering the *cis*-regulatory logic of vertebrate *Brachyury* notochord enhancer(s) will expand our understanding of the evolutionary history of this key developmental regulator and of the mechanisms leading to notochord formation. In particular, comparison to the *Ciona* Brachyury locus containing two upstream shadow enhancers with well-defined regulatory grammar (Farley et al., 2016; Song et al., 2023) may inform *cis*-regulatory adaptations at the onset of vertebrate emergence.

Uncovering the regulatory elements responsible for its notochord expression also promises to shed light onto the role of *Brachyury* in adult human spine health and in chordoma tumors, a rare sarcoma of the spine that is hypothesized to arise from notochord remnants (Nibu et al., 2013; Vujovic et al., 2006; Yakkioui et al., 2014). Native Brachyury-expressing cells in the nucleus pulposus decrease in number with age along with a concomitant increase in cartilage-like cells (Nakamichi and Asahara, 2020; Richardson et al., 2017; Risbud et al., 2010; Tang et al., 2012). What role these long-lasting Brachyury-positive cells play in the adult spine, if they progressively differentiate into cartilage, and how *Brachyury* gene activity is sustained, remain unknown. Detection of Brachyury protein is a diagnostic marker for chordoma (Vujovic et al., 2006), yet the functional contribution of its re-activated or persistent expression in the tumor remains uncertain (D'Agati et al., 2019; Hu et al., 2014; Presneau et al., 2011; Zhu et al., 2016). Several familial chordomas harbor duplications or further complex amplifications of the *Brachyury/T/TBXTB* locus that possibly convey chordoma susceptibility to carriers (Bhadra and Casey, 2006; Hsu et al., 2011; Yang et al., 2009). These findings suggest that chordoma-associated *Brachyury/T/TBXTB* locus amplifications contain, or hijack the action of, *cis*-regulatory elements possibly driving *Brachyury/T/TBXTB* expression in chordoma, potentially with Brachyury controlling its own expression as suggested by ChIP-seq findings (Nelson et al., 2012).

Here, we identify the complement of three auto-regulated shadow enhancers *T3*, *C*, and *I* in the *Brachyury/T/Tbxtb* locus that convey notochord activity. We combined i) genomic data from human chordoma tumor cell lines, human embryonic stem cells, and mouse embryonic stem cells; ii) non-coding element conservation across mammals (human, mouse, Monodelphis) and all vertebrates; iii) transgenic reporter assays in zebrafish, mouse, Axolotl, and *Ciona*; iv) and enhancer knockouts in mice. In triple-enhancer-knockout mice, we document the selective absence of Brachyury protein in the notochord and subsequent neural tube and trunk defects as linked to notochord perturbations. Using comparative genomics, we uncover the conservation of the enhancers *T3*, *C*, and *I* in the *brachyury*/*tbxtb* loci across jawed vertebrates. Our data uncover a deep conservation of shadow enhancers regulating *Brachyury* expression in the notochord, one of the most prominent developmental structures of the vertebrate body and involved in spine and neural tube defects.

RESULTS

Defining a minimal regulatory region for human *Brachyury* notochord expression

To identify enhancer elements with notochord activity in the human *Brachyury*/*T*/*TBXTB* locus, we analyzed the *Brachyury/T/TBXTB* locus to narrow down a minimally required genomic region around the *Brachyury* gene body. Familial and sporadic chordoma feature duplications and/or complex amplifications of *Brachyury* (Bhadra and Casey, 2006; Hsu et al., 2011; Tarpey et al., 2017; Yang et al., 2009), suggesting that essential *cis*-regulatory elements for notochord expression lie within the commonly amplified region. Available genomic patient data outlined a minimally amplified region of approximately 50 kb surrounding the human *Brachyury* gene body, with individual tumors extending their amplifications proximal or distal of this minimal region (Tarpey et al., 2017; Yang et al., 2009) (**Fig. 1A**). Within this minimal interval and its vicinity, we uncovered several regions that have been charted as open chromatin in the chordoma cell lines U-CH2 and MUGCHOR using ATAC-seq (Nelson et al., 2012; Sharifnia et al., 2019), indicating potential regulatory elements in accessible

chromatin (**Fig. 1A**). These regions include a super-enhancer region previously proposed to be active in chordoma (Sharifnia et al., 2019) (**Fig. 1A**). Further, mammalian *Brachyury* has been postulated to control its own notochord expression, as underlined by the *TNE* enhancer that features putative T-box motifs (Beisaw et al., 2018; Schifferl et al., 2021). Using *Brachyury*/*T* ChIP-seq data from the human chordoma tumor cell line U-CH1 and human ES-derived mesoderm progenitor cells (Faial et al., 2015; Nelson et al., 2012), we found discrete Brachyury binding events within the minimal amplification interval and its vicinity (**Fig. 1A**). Genome alignment of human versus other mammalian species indicated candidate enhancer regions (conserved non-coding elements; CNEs) through non-coding sequence conservation in mouse and the more distant marsupial *Monodelphis domestica* (Mikkelsen et al., 2007) (**Fig. 1A**).

From our combined locus analysis, we identified the six initial candidates *T3* (the human ortholog of mouse *TNE)* (Schifferl et al., 2021), *K*, *J*, *C*, *I*, and *L* as putative notochord enhancer elements in the vicinity of the human *Brachyury* gene (**Fig. 1A, Supplemental Table 1**; all Supplemental Data is included in the Supplemental Data Files archive). While *K* and *J* represent conserved sequence to other mammalian genomes, candidates *I* and *L* notably lie in the annotated chordoma super-enhancer region (Sharifnia et al., 2019). From this combined analysis, we hypothesized that individual or combined elements among the six enhancer candidates could convey notochord activity to the human *Brachyury* gene.

Distal enhancers in the human *Brachyury* locus have autonomous notochord activity

Given the evolutionarily conserved notochord expression of vertebrate Brachyury genes, we hypothesized that the human enhancers may be correctly interpreted in a model vertebrate. We initially tested all six enhancer candidates in zebrafish that allows for highly efficient reporter gene activity screening in developing embryos. To test their activity within a broad evolutionary framework, we cloned the human enhancer element candidates *T3*, *K*, *J, C*, *I*, and *L* into reporter vectors coupled with the mouse *betaE-globin* minimal promoter to express the blue fluorophore *mCerulean* for enhancer testing in zebrafish embryos (Kemmler et al., 2023). Upon co-injection into one cell-stage zebrafish embryos together with *ubi:mCherry* as injection control, the human *hs_T3*, *hs_C*, and *hs_I* elements resulted in *mCerulean* expression in the developing zebrafish notochord during early somitogenesis, followed by strong, selective notochord activity in injected embryos at 24 hours post-fertilization (hpf) (n=32/61, n=155/227, n=76/117; mCerulean-positive notochord/total mCherry-positive embryos) (**Fig. 1B-D, Supplemental Table 2**). Zebrafish embryos injected with *hs_T3*, *hs_C*, and *hs_I* reporters maintained notochord-specific *mCerulean* expression throughout our observations until 5 days post-fertilization (dpf). In contrast, we did not observe any specific *mCerulean* reporter expression at any timepoint with elements *hs_K*, *hs_J,* and *hs_L* (n= 0/68, n=0/63, n=0/254) (**Supplemental Table 2**). Notably, *hs_C* was still active when further trimming the sequence 5' and 3' (*hs_Cshort*, n=55/103) (**Supplemental Fig. 1A-C, Supplemental Table 2**). Germline-transmitted, stable transgenic integrations for *mCerulean* reporters based on *hs_T3*, *hs_C*, and *hs_I* recapitulated the transient reporter results and consistently showed selective notochord expression, with minimal variability across independent transgenic insertions for each enhancer reporter (followed to at least F3 generation) (**Fig. 1E-G**). Together, these data indicate that the three enhancer elements *hs_T3*, *hs_C*, and *hs_I* within the human *Brachyury/T/TBXTB* locus convey notochord activity when tested in zebrafish.

Next, we tested the activity of *hs_T3*, *hs_C*, and *hs_I* in axolotl (*Ambystoma mexicanum*) as a representative amphibian species (Nowoshilow et al., 2018; Prummel et al., 2019). Upon microinjection, reporters based on *hs_T3*, *hs_C,* and *hs_I* enhancer elements showed consistent reporter expression in the notochord of axolotl embryos (n=23/47, n=14/16, n=3/3) throughout tailbud stages (st. 30-41) and beyond (**Fig. 1H-J**, **Supplemental Fig. 1D-M, Supplemental Table 2**). Notably, 50% of *hs_T3*-positive F0 animals had additional expression in other mesodermal tissues such as trunk muscles. In contrast, 80% and 100% of positive *hs_C* and *hs_I* F0 animals, respectively, showed expression exclusively in the notochord. In addition, *hs_C* and *hs_I* reporter expression was distributed along the entire rostral-caudal axis in all observed embryos, while *hs_T3* reporter expression was frequently restricted to more caudal portions of the notochord. Combined, these results indicate that the human enhancers *hs_T3*, *hs_C*, and *hs_I* also integrate regulatory input for driving notochord activity in amphibians.

We next tested if human enhancers *hs_T3*, *hs_C* and *hs_I* also drive notochord-specific reporter activity in mouse embryos. For increased specificity and reproducibility, we used a site-directed transgenic integration strategy at the *H11* locus (enSERT) (Kvon et al., 2020) to generate mouse embryos harboring *enhancer-LacZ* reporter transgenes. As observed in zebrafish and axolotl, *hs_T3*, *hs_C* and *hs_I* elements exhibited specific and selective notochord expression in mouse embryos at E9.5 (n=3/3, n=2/2 and n=5/5) (**Fig. 1 K,M,O, Supplemental Table 2**). Of note, *hs_T3* reporter activity appeared predominantly confined to the posterior notochord compared to *hs_C* or *hs_I,* which showed reporter activity in the entire mouse notochord. Histological analysis of Nuclear Fast Red-stained transversal sections from transgenic mouse embryos further confirmed reproducible, notochord-specific activity for human notochord enhancer elements *hs_T3*, *hs_C*, and *hs_I* (**Fig. 1 L, N, P**).

Taken together, we identified three enhancer candidates in the human *Brachyury/T/TBXTB* locus, including *hs_T3* as ortholog of the mouse *TNE* element (Schifferl et al., 2021), that all display notochord enhancer activity as transgenic reporters when tested in teleost fish, amphibian, and rodent embryos, suggesting pan-bony vertebrate activity and function.

Dependence of human *Brachyury* enhancers on T-box motifs

Published ChIP-seq data indicated Brachyury binding at *hs_T3*, *hs_C*, and *hs_I* (**Fig 1A**), suggesting that notochord expression of the *Brachyury*/*T*/*Tbxtb* gene might be auto-regulated by Brachyury itself (Beisaw et al.,

2018; Schifferl et al., 2021). We investigated if the three human notochord enhancer elements contained a TBXT binding motif (short T-box, **Fig. 2A**) using FIMO (Grant et al., 2011): we found that enhancer element *hs_T3* contained two low p-value T-box motifs, *hs_I* contained one low p-value T-box motif, while enhancer element *hs_C* contained a possibly degenerate T-box that we only identified when increasing the p-value (**Fig. 2B**). We then generated the reporter constructs *hs_T3ΔTbox:mApple*, *hs_CshortΔTbox:mApple*, and *hs_IΔTbox:mApple*, in which we deleted the respective T-boxes in the enhancer elements, as well as constructs containing the wildtype enhancer elements in an identical backbone (**Fig. 2C**). The reporter constructs further harbor the transgenesis marker *exorh:EGFP* (expression in the pineal gland, **Fig. 2D-I**) for precise quantification of reporter activity (Kemmler et al., 2023). After injection into zebrafish embryos and in line with the enhancer element activity at 24 hpf (**Fig. 1B-D**), we observed continued and reproducible notochord expression at 48 hpf with all three wildtype enhancer element reporters *T3, C*, and *I* (n=42/58, n=39/57 and n=62/79 (mCerulean-positive notochord/total EGFP pineal gland-positive embryos)) (**Fig. 2D,F,H, Supplemental Table 2**). However, we observed a complete loss of specific notochord reporter activity in zebrafish embryos injected with the deletion constructs *hs_T3ΔTbox:mApple* (n=6/113) and *hs_IΔTbox:mApple* (n=1/41), with positive embryos containing few labelled notochord cells (**Fig. 2E,I**). In contrast, *hs_Cshort ΔTbox* retained reporter activity in the notochord in injected zebrafish embryos with mosaic mApple expression throughout the notochord despite deletion of the T-box motif (n=28/50) (**Fig. 2G, Supplemental Table 2**).

Together, we conclude that the T-box elements in the notochord enhancers *hs_T3* and *hs_I* are critical to the activity of these regulatory elements in reporter assays. In contrast, our mutagenesis of the most-obvious T-box motif in enhancer *hs_C* did not interfere with its notochord activity, suggesting either additional T-box motifs to be important for its reporter activity (**Supplemental Fig. 2** new; need to edit following Supplemental Figs.) or that *hs_*C responds to other upstream input than *Brachyury/T/TBXTB* (Song et al., 2023). Nevertheless, these data support the model in which *Brachyury/T/TBXTB* auto-regulates its own expression in the notochord through select regulatory elements (Beisaw et al., 2018; Schifferl et al., 2021).

The human *Brachyury* enhancers are conserved across mammals

We next sought to determine if other mammalian genomes harbor orthologous *T3*, *C*, and *I* enhancer regions in their *Brachyury/T/Tbxtb* loci. Here, we focused on the orthologous *T3*, *C*, and *I* enhancer candidate regions from mouse (**Fig. 3A**). As in the human *Brachyury/T/TBXT*B locus, we found open chromatin and Brachyury protein binding events at the mouse orthologs of the putative enhancer elements *T3*, *C*, and *I*, as well as the well-characterized murine *Brachyury/T/Tbxtb* promoter *Tstreak* (**Fig. 3A**).

Mouse *TNE* has been established to act as autonomous notochord enhancer when tested in mouse embryos and gastruloid cultures (Schifferl et al., 2021). When tested in zebrafish, both mouse enhancer *mm_T3*/*TNE* and *mm_I* showed reporter activity emerging at the forming shield (**Supplemental Fig. 3A-D**) before expression in the developing notochord (n=46/67, n=61/66) at 24 hpf (**Fig. 3B,D, Supplemental Table 2**). In contrast, mouse enhancer *mm_C* failed to drive any reporter expression in the zebrafish notochord (n=0/88) (**Fig. 3C, Supplemental Table 2**). Imaging of transgenic zebrafish carrying mouse *mm_I* as stable reporter documented robust notochord expression, again with little variability across independent transgenic insertions (**Supplemental Fig. 3E**). The murine *Brachyury/T/Tbxtb* promoter region *Tstreak* (Arnold et al., 2000; Clements et al., 1996; Latinkić et al., 1997) showed transient, variable reporter expression in the zebrafish shield (around 6 hpf), with no reporter activity upon somitogenesis (n=79/102) (**Supplemental Table 2**). We further tested the mouse ortholog of enhancer candidate *mm_J*, as well as the two lesser conserved elements *mm_T1* and *mm_T5*, none of which showed reporter activity in zebrafish embryos up to 5 dpf (n=0/98, n=0/98, n=0/79) (**Supplemental Table 2**).

Tested with site-directed reporter transgenesis at *H11*, *mm_T3*/*TNE* and *mm_I* conveyed specific notochord activity in mouse embryos at E9.5 (n=2/2, n=4/4) (**Fig. 3E,G, Supplemental Table 2**). In contrast, and consistent with our observations in zebrafish reporter assays, *mm_C* did not show any detectable reporter activity in the notochord in mouse embryos at E9.5 (n=0/2) (**Fig. 3F, Supplemental Table 2**).

While humans and mice diverged approximately 90 million years ago, marsupials split from eutherians (placental mammals) approximately 160 million years ago (Goodstadt et al., 2007; Kumar et al., 2017; Mikkelsen et al., 2007; Nei et al., 2001; Pippucci et al., 2011). The opossum *Monodelphis domestica* is a representative marsupial species and provides a more distant comparative species to human and mouse (**Supplemental Fig. 4A**). Detailed sequence alignments documented dispersed conserved regions along the entire sequences for all three enhancer candidates in *Monodelphis* (**Fig. 4A**). When injected into zebrafish embryos as *mCerulean* reporters, the Monodelphis-derived *md_T3*, *md_C*, and *md_I* enhancer element candidates all conveyed specific notochord activity at 24 hpf (n=47/62, n=142/184, n=74/97) (**Fig. 4B-D, Supplemental Table 2**), with *md_T3* transiently expressing at 80% epiboly (**Supplemental Fig. 4B,C**). In addition to the notochord activity, *md_C* reporter-injected zebrafish embryos showed transient reporter expression in the heart whereas *md_I* reporter-injected embryos showed transient expression in the brain and spinal cord neurons (**Fig. 4C,D).**

Given the mammalian sequence conservation and differential responses in reporter assays, we next tested the notochord enhancer element candidates in the tunicate *Ciona intestinalis* as non-vertebrate outgroup. As a chordate, *Ciona* forms a bona fide notochord (Passamaneck et al., 2009). Testing *T3*/*TNE*, *C*, and *I* of human, mouse, and *Monodelphis* by reporter gene assays in *Ciona*, we found that only *Monodelphis*-derived *md_C* showed specific and robust reporter activity in the notochord (n=119/150) compared to all other eight elements (n=0/150) and minimal promoter only control (n=0/150) (**Fig. 4E,F, Supplemental Table 2**).

Taken together, and extending previous work on the

mouse *TNE* element (Schifferl et al., 2021), our data indicate that three distant elements in the mammalian *Brachyury/T/Tbxtb* locus with differential activity converge on providing notochord-specific activity in reporter assays across chordates.

Enhancer deletions cause notochord-selective loss of *Brachyury/T/Tbxtb* expression in mice

While especially enhancer element *C* seems to have diverged in activity (or is sensitive to the specific *trans* environment it was tested in), all three elements *T3*, *C*, and *I* remain conserved and detectable at the sequence level throughout the mammalian clade. In mice, homozygous *Brachyury/T/Tbxtb* mutations in the gene body cause preimplantation defects leading to embryonic lethality between E9.5 and E10.5 (Gluecksohn-Schoenheimer, 1938; Gluecksohn-Schoenheimer, 1944; Yanagisawa, 1990). Previous work established that deletion of mouse enhancer *T3*/*TNE* does not cause a fully penetrant loss of *Brachyury/T/Tbxtb* expression in the developing notochord, indicating the presence of additional shadow elements interacting with, or compensating for, *T3*/*TNE* (Schifferl et al., 2021). To functionally test if the three enhancer elements are involved in *Brachyury/T/Tbxtb* expression in the mouse notochord, we generated a series of knockout alleles targeting the three mouse enhancer elements *T3*, *C*, and *I* (**Fig. 5A**).

We employed CRISPR-Cas9 genome editing using target sites flanking the enhancers and established heterozygous and homozygous mice carrying individual and combined enhancer deletions (**Fig. 5A, Supplemental Fig. 5A**). Compared to E9.5 wildtype control embryos (**Fig. 5B**) (n=14/14), neither homozygous deletion of mouse *C* ($T^{\Delta C/\Delta C}$) (n=7/7) or *I* ($T^{\Delta I/\Delta I}$) (n=7/7) alone, nor heterozygous ($T^{+/\Delta C,I}$) (n=12/12) (**Supplemental Fig. 5B-D**) or homozygous deletion of both *C* and *I* ($T^{\Delta C,I/\Delta C,I}$) (n=9/9) (**Fig. 5C**) altered *Brachyury/T/Tbxtb* expression in the notochord as determined by Brachyury/T antibody staining.

In contrast, we observed reduced Brachyury/T/Tbxtb expression in the notochord of E9.5 embryos in a dose-dependent manner when we combined *ΔT3*/*TNE* with *ΔC,I* deletions. E9.5 embryos heterozygous for the triple knockout chromosome carrying *ΔT3,C,I* ($T^{+/\Delta T3,C,I}$) *in cis* appeared normal (n=14/14) (**Supplemental Fig. 5E**). In contrast, we documented a partial loss of Brachyury/T/Tbxtb protein in trans-heterozygous E9.5 embryos carrying *ΔC,I* and *ΔT3,C,I* alleles ($T^{\Delta C,I/\Delta T3,C,I}$) (n=14/18) (**Fig. 5D**). Comparably, Brachyury/T /Tbxtb protein levels were even further reduced or lost in the notochord of embryos trans-heterozygous for *ΔT3* and *ΔT3,C,I* alleles ($T^{\Delta T3/\Delta T3,C,I}$) (n=10/10) (**Fig. 5E**). These data are consistent with, and expand upon, previous observations that the severity of *Brachyury/T/Tbxtb* phenotypes correlate with gene dosage (Yanagisawa, 1990). Importantly, the $T^{\Delta T3/\Delta T3,C,I}$ genotype with severely reduced Brachyury/T/Tbxtb protein levels is consistent with the loss of Brachyury/T/Tbxtb protein in the notochord in mice trans-heterozygous for *T3*/*TNE* deletion and a large, locus-spanning *Brachyury/T/Tbxtb* deletion that includes elements *C* and *I* (Schifferl et al., 2021), revealing the actual relevant enhancer regions (**Fig. 1,3,4**) and motifs (**Fig. 2**). Finally, E9.5 homozygous triple knockout *ΔT3,C,I* embryos ($T^{\Delta T3,C,I/\Delta T3,C,I}$) showed a complete absence of Brachyury/T/Tbxtb protein in the prospective notochord region (n=0/5) yet all embryos retained Brachyury/T/Tbxtb protein in the tailbud (n=5/5) (**Fig. 5F**). Taken together, our data establish the notochord-specific *Brachyury/T/Tbxtb* loss-of-function mutant in mice by means of deleting three conserved enhancer elements *in cis*.

Next, we examined phenotypic defects resulting from perturbed *Brachyury/T/Tbxtb* expression using various allele combinations involving *ΔC,I and ΔT3,C,I*. Consistent with the phenotypes at E9.5 (**Fig. 5B-F**), we observed a gradual increase of phenotype severity with deletion of the three different enhancer elements at E12.5 (**Fig. 5G-Z**). Wildtype control (n=25/25) (**Fig. 5G,L**), homozygous $T^{\Delta C,I/\Delta C,I}$ embryos (n=24/24) (**Fig. 5H,M**), heterozygous $T^{+/\Delta C,I}$ (n=5/5) and $T^{+/\Delta T3,C,I}$ embryos (n=23/23) (**Supplemental Fig. 5F,G**) appeared grossly normal. In contrast, we observed rudimentary/short/small tails with additional enhancer deletions. Rudimentary/short/small tails appeared in trans-heterozygous $T^{\Delta C,I/\Delta T3,C,I}$ embryos in 4.7 % (n=2/43) (**Fig. 5I,N**) and was fully penetrant in trans-heterozygous $T^{\Delta T3/\Delta T3,C,I}$ embryos (n=10/10) (**Fig. 5J,O**), as well as in triple homozygous $T^{\Delta T3,C,I/\Delta T3,C,I}$ embryos (n=16/16) (**Fig. 5K,P**). In addition, trans-heterozygous $T^{\Delta T3/\Delta T3,C,I}$ embryos displayed caudal spina bifida (n=16/16) (**Fig. 5O**), as previously reported for homozygous *TNE* embryos (Schifferl et al., 2021). Finally, triple-homozygous $T^{\Delta T3,C,I/\Delta T3,C,I}$ embryos lacking all three enhancers displayed spina bifida along 3/4 of the spine (n=16/16) (**Fig. 5P**), reminiscent of previous observations using *Brachyury/T/Tbxtb*-targeting RNAi in mouse embryos (Pennimpede et al., 2012; Zhu et al., 2016). These results provide compelling phenotypic evidence of the impact of cumulative enhancer deletions on *Brachyury/T/Tbxtb* expression in the notochord.

We further validated these phenotypes with immunohistochemistry and histology. We visualized Brachyury/T/Tbxtb protein in transversal sections of E12.5 embryos together with the neural plate marker Sox2: compared to wildtype (**Fig. 5Q**), heterozygous $T^{+/\Delta C,I}$, $T^{+/\Delta T3,C,I}$ (**Supplemental Fig. 5H,I**) as well as homozygous $T^{\Delta C,I/\Delta C,I}$ (**Fig. 5R**) embryos that were all grossly normal, we found decreased Brachyury protein in the notochord of $T^{\Delta C,I/\Delta T3,C,I}$ embryos (**Fig. 5S**). Strikingly, we observed a complete absence of Brachyury protein in $T^{\Delta T3/\Delta T3,C,I}$ embryos (**Fig. 5T**) and $T^{\Delta T3,C,I/\Delta T3,C,I}$ embryos (**Fig. 5U**). In contrast, Sox2 expression was comparable in all embryos (**Fig. 5Q-U, Supplemental Fig, 5H,I**), even in $T^{\Delta T3,C,I/\Delta T3,C,I}$ embryos that clearly displayed spina bifida along the entire spine (**Fig. 5U**). Compared to wildtype embryos (**Fig. 5V**), additional histology assessed by H&E staining confirmed wildtype-looking notochords in $T^{+/\Delta C,I}$, $T^{+/\Delta T3,C,I}$, and homozygous $T^{\Delta C,I/\Delta C,I}$ embryos (**Supplemental Fig. 5J,K, Fig. 5W**), smaller/rudimentary notochords in $T^{\Delta C,I/\Delta T3,C,I}$ embryos (**Fig. 5X**), and absent notochords in $T^{\Delta T3/\Delta T3,C,I}$ and $T^{\Delta T3,C,I/\Delta T3,C,I}$ embryos (**Fig. 5Y,Z**).

We found that the two most severe enhancer mutants are not viable since we did not recover adult triple-homozygous $T^{\Delta T3,C,I/\Delta T3,C,I}$ (n=0/59) or trans-heterozygote $T^{\Delta T3/\Delta T3,C,I}$ (n=0/31) animals at term (**Supplemental Fig. 5L**), indicating lethality prior to or shortly after birth. In contrast, $T^{\Delta C,I/\Delta T3,C,I}$ (n=46) trans-heterozygotes and

homozygous $T^{\Delta C,I/\Delta C,I}$ (n=100) animals survived to adulthood. Notably, a variable percentage of $T^{\Delta C,I/\Delta C,I}$, $T^{\Delta C,I/\Delta T3,C,I}$, and $T^{+/\Delta T3}$ animals presented with kinked tails (**Supplemental Fig. 5M**), with two $T^{\Delta C,I/\Delta T3,C,I}$ animals displaying a short/small tail (**Supplemental Fig. 5N**), reminiscent of hypomorphic *Brachyury*/*T*/*Tbxtb* mutants and *in vivo Brachyury/T/Tbxtb* knockdown by siRNA (Dobrovolskaia-Zavadskaia, 1927; Pennimpede et al., 2012; Schifferl et al., 2021; Zhu et al., 2016). Taken together, our data are consistent with the correlation of *Brachyury/T/Tbxtb*-mutant phenotypes and gene dosage controlled by enhancer activity, as revealed by increasing phenotype severity with an increasing number of combined enhancer deletions in *Brachyury/T/Tbxt*.

In summary, our data establishes that the combined activity of the enhancers *T3/TNE*, *C*, and *I* in the mouse *Brachyury/T/Tbxtb* locus are necessary to convey notochord expression of *Brachyury/T/Tbxtb*. Upon combined loss of these enhancers, the notochord is lost.

T3, *C* and *I* are conserved in sequence and function among jawed vertebrates

The evolutionary trajectory of chordate *Brachyury* control in the notochord remains unresolved. The notochord-regulatory elements driving *Brachyury* expression in *Ciona* are promoter-proximal (Corbo et al., 1997; Nibu et al., 2013; Satoh et al., 2012). Zebrafish *tbxta/ntla* harbors a -2.1 kb upstream notochord element containing the two smaller elements *E1* and *E2* (Harvey et al., 2010). In contrast, zebrafish *tbxtb* descended from the same ancestral *Brachyury* gene as the single mammalian *Tbxtb* gene. Further, while zebrafish *tbxtb* remains expressed in the notochord (Inoue et al., 2017; Martin and Kimelman, 2010), its regulatory elements have not been reported. Using direct sequence comparisons of mammalian *T3*, *C*, and *I* to the zebrafish genome, we did not find any sequences of significant sequence similarity.

Identifying non-coding sequence conservation across vertebrate lineages, whether from human or other tetrapods to the fast-evolving teleost fishes like zebrafish, remains notoriously challenging. Species with slow rates of molecular evolution can help as "genomic bridges" to provide sequence connectivity across distant vertebrate groups (Braasch et al., 2016; Thompson et al., 2021). The spotted gar (*Lepisosteus oculatus*) is a slowly evolving ray-finned fish that has diverged from zebrafish and other teleosts before a teleost-specific whole-genome duplication, providing a bridge species for genomic comparisons between tetrapods and teleosts (Braasch et al., 2016). Using BLAST searches, we found sequence similarity between human *T3*/*TNE*, *C*, and *I* and regions of the spotted gar *tbxtb* locus with equivalent positions relative to the gar *tbxtb* gene body compared to mammals (**Fig. 6A**). Next, we used these spotted gar *T3*/*TNE*, *C*, and *I* regions as BLAST queries to bridge to the genomes of zebrafish and other fish lineages (**Supplemental Table 4**). This approach uncovered candidate regions for *T3*/*TNE* and *I*, but not *C*, within the zebrafish *tbxtb* locus (**Fig. 6A**).

Analogous to our tests with mammalian enhancer candidates, we cloned reporter transgenes coupled with the *betaE-globin:mCerulean* cassette using the *T3*/*TNE*, *C*, and *I* enhancer elements from the spotted gar *tbxtb* locus. Upon injection into zebrafish embryos, both spotted gar *lo_T3* and *lo_I* displayed specific and reproducible notochord reporter activity (n=39/54, n=82/122) (**Fig. 6B,D, Supplemental Table 2**). In contrast, and akin to the mouse *mm_C* enhancer element, spotted gar element *lo_C* did not convey any notochord reporter activity in zebrafish embryos (n=0/92) (**Fig. 6C, Supplemental Table 2**). The zebrafish-derived *dr_T3* and *dr_I* also showed selective notochord activity when tested in zebrafish F) transgenic reporter assays (n=122/160, n=81/117) (**Fig. 6E,F, Supplemental Table 2**). Further confirming our results, we found robust reporter activity in the notochord of stable transgenic zebrafish lines based on *dr_T3* and *dr_I* (**Fig. 6G,H**).

Using the three gar elements as queries, in addition to clupeocephalan teleosts (e.g. zebrafish), we found *T3* and *I* also in the other two major teleost lineages elopomorphs (e.g. eel) and osteoglossomorphs (e.g. arowana). However, we did not detect any equivalent sequence for *C* in any teleosts, indicating that this element has been lost or diverged beyond recognition in the teleost lineage (**Fig. 6I**). However, we detected orthologs of all three elements, including *C*, at expected locations around the *tbxtb* genes in additional non-teleost ray-finned fishes (e.g. bowfin, sturgeon, reedfish) as well as in the more basally diverging cartilaginous fishes (e.g. sharks, skate) (**Supplemental Table 4**); in contrast, we only detected *T3* and *I* in the lobe-finned coelacanth (**Fig. 6I**). To explore the presence of the three enhancer elements among tetrapods, we used the painted turtle, characterized by a slow genomic evolutionary rate (Bradley Shaffer et al., 2013; Takezaki, 2018), as an additional bridge species within tetrapods. We found all three elements in the turtle *Brachyury/T/Tbxtb* locus and through use of the painted turtle as reference also in other reptiles and birds, as well as in amphibians (e.g. axolotl) (**Fig. 6I**, **Supplemental Table 4**), but did not detect any of the three elements in the jawless cyclostome (e.g. lamprey, hagfish) genomes. Furthermore, we found that the human T-box motifs, which we identified using FIMO (**Fig. 2**) in our enhancers, are conserved across tetrapods and fishes as distantly related as ghost shark based on sequence alignments (**Supplemental Fig. 6A-C**) as well as multi-species FIMO analyses (**Supplemental Data File**). Cross-species sequence conservation centers at the T-box motifs (**Supplemental Fig. 6A-C**) which supports both their functional importance as well as their evolutionary ancestry since at least the last common ancestor of jawed vertebrates.

Taken together, our observations provide strong evidence that notochord enhancers *T3*, *I*, and *C* are deeply conserved *cis*-regulatory elements of the *Brachyury/T/Tbxtb* gene that were already present in the last common ancestor of jawed vertebrates over 430 million years ago.

DISCUSSION

How the *Brachyury/T/Tbxtb* gene is controlled during notochord development is fundamental to our understanding of how basic concepts of body plan formation remain conserved or have diverged across species. Shadow enhancers, seemingly redundant transcriptional *cis*-regulatory elements that regulate the

same gene and drive overlapping expression patterns, are a pervasive feature of developmental gene regulation (Kvon et al., 2021). The concept of enhancer redundancy through one or more shadow enhancers acting on the same gene in addition to primary enhancer has been established for numerous loci (Antosova et al., 2016; Cannavò et al., 2015; Hong et al., 2008; Kvon et al., 2021; Letelier et al., 2018; Osterwalder et al., 2018). Shadow enhancers are thought to provide robustness to gene expression and buffer against genetic and environmental variations (Kvon et al., 2021), a hypothesis validated in mammals (Antosova et al., 2016; Osterwalder et al., 2018).

Here, we discovered a deeply conserved set of three notochord-specific shadow enhancers within the human TBXT locus as ancient *cis*-regulatory logic. Cross-species enhancer testing reveals that the *cis*-regulatory grammar of all three human enhancers (*T3*, *I*, and *C*, respectively) is correctly interpreted in other vertebrates including mice, salamanders, and zebrafish, but not in the invertebrate chordate *Ciona*. The three notochord enhancers described here are not the only conserved non-coding elements across mammalian *Brachyury/T/Tbxtb* loci (**Fig. 1A, 3A, 4A**): while our reporter gene assays in zebrafish did not reveal any notochord activity of six tested human enhancer elements besides *T3*, *C* and *I*, we cannot rule out synergistic or interdependent notochord activity conveyed by additional elements. Reporter gene assays indicate that not all three *Brachyury/T/Tbxtb* notochord enhancers have equal potency. Enhancer element *C* shows variable activity and remains unrecognized in teleost fishes and Coelacanth. Compared to human *C* with reproducible notochord activity in all tested models (**Fig. 1C,F,I,M**) and Monodelphis *C* that is active in zebrafish and uniquely in Ciona (**Fig. 4C,E**), mouse *C* showed no discernible activity in any assay including in mouse embryos (**Fig. 3C,G**) despite significant sequence conservation. We speculate that while mouse *C* is not active in isolation, it may contribute together with *T3*/*TNE* and *I* to *Brachyury* activity in the notochord. This model is consistent with the impact of *TNE* deletions when combined with larger deletions that include *TNE* and *C* in mouse trans-heterozygotes (Schifferl et al., 2021) (**Fig. 5**). The potential auto-regulation of *Brachyury/T/Tbxtb* by its protein product via in part conserved T-box motifs in enhancers *T3/TNE* and *I* might contribute to the enhancer redundancy and divergent activity of element *C* when tested in isolation (**Fig. 2**). Our data propose that enhancer *C* is an auxiliary element to *T3/TNE* and might contribute to duration, expression levels, or other features that differ among *Brachyury/T/Tbxtb* notochord expression across vertebrates. Our combined data proposes a model in which notochord expression of vertebrate *Brachyury/T/Tbxtb* is cumulatively or cooperatively driven by enhancers *T3*, *C*, and *I*. In this model, sequence variants of *T3*, *C*, and *I* that modulate their individual potency became selected for modulating Brachyury/T levels to species-specific requirements.

The conservation of gene order (micro-synteny) between species can be indicative of the presence of *cis*-regulatory elements, which are crucial for controlling expression of the physically linked genes (Irimia et al., 2012). The finding of functionally relevant distant enhancers 5' and 3' of the *Brachyury/T/Tbxtb* gene body is further supported by the conserved gene linkage *Sftd2-(Prr18)-Brachyury/T/Tbxtb-Pde10a* across the entire jawed vertebrate phylogeny. In agreement with a distinct gene linkage surrounding *Brachyury/T/Tbxtb* in agnathans (**Fig. 6I**), we were unable to identify any of the three distant enhancers in two species representing this clade. Likewise, a distinct gene linkage associates with *Tbxta*, the second *Tbxtb* paralog in fish, which apparently lacks any of the three notochord enhancers described here. *tbxta/ntla* expression is instead controlled by two mesoderm/notochord enhancers located close to the gene promoter (Harvey et al., 2010), a possible example of evolutionary novelty following ancestral gene duplication. In contrast, the functionally less impactful zebrafish *tbxtb/ntlb* gene retained the *cis*-regulatory logic of the *Tbxtb* gene from the jawed vertebrate ancestor (**Fig. 6**). We did not find any evidence for sequence conservation of the *Tbxtb T3*, *I*, or *C* regions within vertebrate *Tbxta* loci or any other genomic regions. Future detailed studies across vertebrate *Tbxt* paralogs are needed to evaluate whether or not the three *Tbxtb* regulatory elements identified here were already part of the single *Tbxt* gene in a vertebrate ancestor. Notably, zebrafish mutants of *tbxta/ntla* have been widely studied as model for *Brachyury* function in notochord formation (Amacher et al., 2002; Halpern et al., 1993; Schulte-Merker et al., 1994), while the seemingly less impactful *tbxtb* has retained ancestral regulation. Why zebrafish, and possibly other fish lineages, use *tbxta* as their main functional Brachyury paralog, and how the regulatory balance between *T3*, *C*, and *I* plays out across individual vertebrate lineages, warrants future efforts.

We found that *Brachyury/T/Tbxtb* notochord enhancers *T3* and *I*, and possibly further supported by enhancer *C*, represent a shadow enhancer combination that contributes to the robust *Brachyury/T/Tbxt* expression in mammals. In mice, neither deletion of enhancer *T3*/*TNE* (Schifferl et al., 2021), nor deletion of enhancer *C*, *I,* or C and I, resulted in a discernable notochord phenotype (**Fig. 5**). Nonetheless, by combining deletions of all three notochord enhancer elements, we showed a dose response for Brachyury/T expression in the notochord. In particular, in embryos where *ΔT3* is combined with a chromosome harboring *ΔT3,C,I* as trans-heterozygotes ($T^{\Delta T3/\Delta T3,C,I}$) or in triple homozygous knock-out embryos ($T^{\Delta T3,C,I/\Delta T3,C,I}$), we observed loss of Brachyury/T protein in the notochord as well as notochord-specific phenotypes, such as spina bifida (**Fig. 5**). The neural tube closure defects are similar to phenotypes observed in *Brachyury/T/Tbxtb* knockdown embryos (Pennimpede et al., 2012; Zhu et al., 2016) or hypomorphic *Brachyury/T/Tbxtb* mutants (Dobrovolskaia-Zavadskaia, 1927). These results assign an essential, combinatorial role to the enhancer elements *T3/TNE*, *C* and *I* in regulating *Brachyury/T/Tbxtb* in the notochord. Notably, previous work (Rennebeck et al., 1995; Rennebeck et al., 1998) has described the *T2* mutant caused by a large viral integration 5' of the mouse *Brachyury*/*Tbxt* locus that i) is recessive lethal with phenotypes reminiscent of *Brachyury* loss, and ii) does complement loss-of-function alleles for Brachyury. *T2* has been hypothesized to encode a short protein off a long mRNA (Rennebeck et al., 1995; Rennebeck et al., 1998). The described genomic position of the viral integration in *T2* places it in the vicinity and

upstream of enhancer element *C*. We note that various vertebrate *Brachyury/tbxtb* loci feature annotated long non-coding RNAs upstream of the main gene body that are reminiscent of enhancer RNAs (**Fig. 3A, 6A**). We therefore hypothesize that the *T2* mutation is caused by a disruption of the gene-regulatory landscape of the mouse *Brachyury/Tbxt* gene by the viral integration, changing the interaction of distant enhancer elements with the promoter. Inspection of the chromatin landscape of the *Brachyury/Tbxt* locus, also in *T2* mutants, could shed light on the architecture of the locus during notochord development.

The significance of *Brachyury/T/Tbxtb* regulation in the notochord translates to chordoma tumors that feature expression of this T-box transcription factor as key diagnostic readout (Sangoi et al., 2011; Takei and Powell, 2010; Vujovic et al., 2006). Both sporadic and familial chordoma are hypothesized to derive from notochord remnants in the spine that do not convert to nucleus pulposus tissue (Choi et al., 2008; Heaton and Turner, 1985; Vujovic et al., 2006). However, the mechanisms leading to retained or reinitiated *Brachyury/T/TBXTB* expression in chordoma remain unknown. Our analysis of reported familial and sporadic chordoma amplifications indicate that amplifications invariantly retain the notochord enhancer *I* together with the gene body including the promoter (Tarpey et al., 2017; Yang et al., 2009). Enhancer *I* lies within a super-enhancer region identified in chordoma cell lines (Sharifnia et al., 2019), further implicating its transcriptional engagement in chordoma. Amplifications occurring in tandem with the original locus propose a scenario where the retained enhancer *I* could synergize with *C* and *T3* from the original locus on the newly amplified gene copies, potentially resulting in increased *Brachyury/T/TBXTB* expression (**Fig. 1A**). Beyond chordoma, changes in enhancer sequence or relative distance to the *Brachyury/T/TBXTB* gene body could also impact spine formation and health by altering the robustness of Brachyury expression in the notochord and subsequent nucleus pulposus.

Tremendous progress in *in vitro* differentiation regimens have resulted in stem cell-derived models for body segmentation and different organ structures (Brink et al., 2020; Lópezlópez-Anguita et al., 2022; Moris et al., 2020; Veenvliet et al., 2020). However, notochord formation has only been reported in more complex systems that recapitulate major hallmarks of embryo patterning (Rito et al., 2023; Xu et al., 2014; Xu et al., 2021). Reporters based on our isolated enhancers potentially provide potent readouts to screen for differentiation regimens that result in notochord fates. Together, our uncovered set of shadow enhancers in *Brachyury/T/TBXTB* advance our concepts of how this key contributor to notochord formation is regulated and de-regulated in development and disease.

MATERIALS AND METHODS

Brachyury locus annotations

The UCSC genome browser was used to identify and visualize enhancer elements in the human, mouse, and Monodelphis *Brachyury* locus. *.bed files were generated with the approximate genomic location of human *Brachyury* amplifications in chordoma tumors from different patients (Tarpey et al., 2017; Yang et al., 2009). Previously published ATAC sequencing data of U-CH2 cells and MUGCHOR cells (Sharifnia et al., 2019), as well as Brachyury/T ChIP sequencing data of human embryonic stem cells (hESCs) (Faial et al., 2015) and U-CH1 cells (Nelson et al., 2012) were added. Further, the repeat masker track and the conservation track for mouse and Monodelphis were added. Ultimately, using this strategy, the human enhancer element candidates *T3*, *K*, *J*, C, *I*, and *L* were identified. For detailed information, see **Supplemental Table 1 and 3**.

The same strategy was applied to find the corresponding mouse enhancer elements. Published ATAC-seq data of mouse ESCs (Tosic et al., 2019) and Brachyury/T-positive fluorescence-activated cell sorted cells from the caudal ends of wild-type mouse embryos (TS12/8 dpc and TS13/8.5 dpc) (Koch et al., 2017), as well as Brachyury/T ChIP sequencing data of mouse ESCs (Beisaw et al., 2018; Koch et al., 2017) were used. Again, the repeat masker track, the ENCODE Candidate Cis-Regulatory Elements (cCREs, combined from all cell types) track, and the Vertebrate Multiz Alignment & Conservation track to check for conservation in human, Monodelphis, and zebrafish, were added. This approach identified the mouse enhancer element candidates *T1*, *T2*, *T3*/*TNE*, *J*, *C2*, *C*, *Tstreak*, *I*, *T4*, *T5*, and *T6*, of which *T1*, *T3*/*TNE*, *J*, *C*, *Tstreak*, *I*, *and T5* were pursued and tested (**Supplemental Table 3 and 5**).

To find the corresponding Monodelphis elements, the repeat masker and 9-Way Multiz Alignment & Conservation track were included to identify *T3*, *C*, and *I* (**Supplemental Table 3 and 5**).

Cloning of the enhancer element reporter plasmids

Each *Brachyury* enhancer element candidate was amplified from either human, mouse, Monodelphis, spotted gar, or zebrafish genomic DNA using the Expand Hi-Fidelity PCR System (11732641001, Roche). Exact coordinates are listed in **Supplemental Table 3**. Each enhancer candidate was TOPO-cloned into the *pENTR5'-TOPO* plasmid (K59120, Invitrogen) using halt-volume reactions according to the manufacturer's instructions (half-volume reactions). Subsequent Multisite Gateway cloning (half-volume reactions) were performed using LR Clonase II Plus (12538120, Invitrogen) according to the manufacturer's instructions (half-volume reactions) and recommended reaction calculations (Mosimann, 2022). *5'* entry plasmids containing the different enhancer elements were assembled into reporter expression plasmids together with the middle entry plasmid (*pME*) containing the mouse *betaE-globin* minimal promoter expressing mCerulean (*pSN001*) as well as mApple (*pCK068*), the 3'plasmid *#302* (*p3E_SV40polyA*), and the destination plasmid *pDESTTol2A2* containing *crybb1:mKate2* (*pCB59*) and *pDESTexorh:EGFP* containing EGFP expression in the pineal gland (*pCK017*) as transgenesis markers (Kemmler et al., 2023). Assembled vectors were verified using restriction digest and Sanger sequencing using standard sequencing primers for Multisite Gateway assemblies (Kemmler et al., 2023; Mosimann, 2022).

Zebrafish husbandry, transgenic reporter assays and stable transgenic lines

Zebrafish animal care and procedures were carried out in accordance with the IACUC of the University of Colorado Anschutz Medical Campus (protocol # 00979), Aurora, Colorado. Zebrafish embryos were raised in E3 embryo medium at 28 °C.

To test the transient activity of our putative enhancer elements, 25 ng/µL *Tol2* mRNA, 12.5 ng/µL reporter expression plasmid DNA, and 12.5 ng/µL *ubi:mCherry* plasmid (Mosimann et al., 2011) as injection control were co-injected into one-cell stage wild type zebrafish embryos (Prummel et al., 2019). At 24 hpf, embryos were anesthetized with 0.016% Tricaine-S (MS-222, Pentair Aquatic Ecosystems Inc.) in E3 embryo medium and embedded in E3 with 1% low-melting-point agarose (A9045, Sigma Aldrich). Embryos were mounted laterally on glass bottom culture dishes (627861, Greiner Bio-One) and confocal imaging was performed with a Zeiss LSM880 using a ×10/0.8 air-objective lens. Fluorescence channels were acquired sequentially with maximum speed in bidirectional mode in 3 µM slices. The range of detection for each channel was adapted to avoid any crosstalk between the channels. Images of acquired Z-stacks were reconstructed with ImageJ/Fiji as a maximum intensity projections.

To generate stable transgenic lines, 25 ng/µL *Tol2* mRNA were co-injected with 25 ng/µL reporter expression plasmid DNA (Felker and Mosimann, 2016; Kwan et al., 2007). Multiple F0 founders were screened for specific *mCerulean* and *mKate2* expression, raised to adulthood, and screened for germline transmission. Resulting F1 single-insertion transgenic strains were established and verified through screening for a 50% germline transmission rate outcrosses in the subsequent generations as per our previously outlined procedures (Felker and Mosimann, 2016).

Axolotl husbandry, transgenic reporter assays and immunostaining

Axolotl husbandry and experiments (non-free feeding stages) were performed at the CRTD/Center for Regenerative Therapies Dresden, Dresden, Germany. Transgenic animals were generated using *Tol2* transposase following standard protocols (Khattak et al., 2014). For live imaging, the animals were anaesthetized by bathing in 0.01% benzocaine and imaged on an Olympus SZX16 fluorescence stereomicroscope. Embryos were staged as described previously (Armstrong and Malacinski, 1989).

For immunostaining, axolotl embryos were fixed in MEMFA at 4 °C overnight, washed in PBS, embedded in 2% low-melting temperature agarose, and sectioned by vibratome into 200 µm-thick sections. Fibronectin was detected using mouse anti-Fibronectin antibody (ab6328, Abcam) at 5 µg/mL. Confocal images were acquired on a Zeiss LSM780-FCS inverted microscope.

Transgenic mouse reporter assays

Research was conducted at the E.O. Lawrence Berkeley National Laboratory (LBNL) and performed under U.S. Department of Energy Contract DE-AC02-05CH11231, University of California (UC). Transgenic mouse assays were performed in *Mus musculus* FVB strain mice (animal protocol number 290003; reviewed and approved by the Animal Welfare and Research Committee at Lawrence Berkeley National Laboratory). Animals of both sexes were used in these analyses. Sample size selection and randomization strategies were conducted as follows: sample sizes were selected empirically based on previous experience in transgenic mouse assays for >3,000 total putative enhancers (VISTA Enhancer Browser: https://enhancer.lbl.gov/). Mouse embryos were excluded from further analysis if they did not encode the reporter transgene or if the developmental stage was not correct. All transgenic mice were treated with identical experimental conditions. Randomization and experimenter blinding were unnecessary and not performed. For comprehensive analysis of species-specific *T3*, *C* and *I*, enSERT enhancer analysis was used, allowing for site-directed insertion of transgenic constructs at the *H11* safe-harbor locus (Hippenmeyer et al., 2010; Tasic et al., 2011). EnSERT is based on co-injection of Cas9 protein and *H11*-targeted sgRNA in the pronucleus of FVB single cell-stage mouse embryos (E0.5) with the targeting vector encoding a candidate enhancer element upstream of the *Shh*-promoter-*LacZ* reporter cassette (Kvon et al., 2020). Enhancer elements were PCR-amplified from human, mouse and Monodelphis genomic DNA and cloned into the respective *LacZ* expression vector (Osterwalder et al., 2022). Embryos were excluded from further analysis if they did not contain a reporter transgene in tandem. CD-1 females served as pseudo-pregnant recipients for embryo transfer to produce transgenic embryos which were collected at E9.5 and stained with X-gal using standard methods (Osterwalder et al., 2022).

Histological analysis of Nuclear Fast Red-stained sections from transgenic mouse embryos

After LacZ staining, E9.5 transgenic mouse embryos were dehydrated in serial alcohols (1x 70%, 1x 80%, 1x 90%, 2x 96%, 2x 100% ethanol, followed by 1x 100% isopropanol for 20 minutes each) and cleared twice for 30 minutes with Histo-Clear II (HS-202, National Diagnostics) for paraffin wax embedding. 10 µm-thick transverse sections were obtained with a Leica Biosystems RM2245 Semi-Automated Rotary Microtome. Sections were de-waxed, rehydrated, and stained with Nuclear Fast Red (R5463200, Ricca Chemical) for two minutes. After staining, sections were dehydrated and mounted with Omnimount (HS-110, National Diagnostics). Images were obtained using a Leica M205 FA stereo microscope.

Ciona reporter assays

Ciona experiments were performed at UCSD as described previously (Farley et al., 2015; Song et al., 2023). Adult *Ciona intestinalis* type A aka *Ciona robusta* (obtained from M-Rep) were maintained under constant illumination in seawater (obtained from Reliant Aquariums) at 18 °C. Briefly, human, mouse and Monodelphis enhancer elements *T3*, *C* and *I* were subcloned into appropriate plasmids suited for expression in *Ciona*, upstream of a basal *Ciona* Forkhead promoter driving GFP (Farley et al., 2016; Harafuji et al., 2002). *Ciona* embryos were electroporated with 70 µg of each plasmid as previously described (Christiaen et al., 2009) and reporter expression was counted blind in 50 embryos per biological repeat, 3 repeats were performed. Images

were taken of representative embryos.

Deletion of mouse enhancer elements T3, C and I

All mouse experimental procedures and animal care were approved by the Animal Care Committee of the Institute of Molecular Genetics (IMG), Czech Academy of Sciences, Prague, Czech Republic, and covered under protocol permission number 357/2021. Experiments were performed in compliance with the European Communities Council Directive of November 24, 1986 (86/609/EEC), as well as national and institutional guidelines.

For this study, inbred C57BL/6N mice were used. Mice carrying deletions of enhancer elements *T3/TNE*, *C*, and *I* were generated using CRISPR-Cas9 technology. The cRNAs (purchased from Integrated DNA technologies) were designed to target the 5' and 3' ends of the mouse enhancer elements *T3/TNE*, *C* and *I* to delete the genomic regions in between. For genomic location and sequence of the selected target sites, as well as genomic coordinates of the deleted enhancer element sequences, see **Supplemental Table 5**.

A ribonucleoprotein (RNP) complex of crRNA/TRACR (1072532, Integrated DNA technologies) and SpCas9 protein (1081058, Integrated DNA technologies) was electroporated into fertilized zygotes isolated from C57BL/6N mice. Zygote electroporation and transfer into pseudo-pregnant foster females was performed as previously described (Jenickova et al., 2021). Founder animals from multiple embryo transfers were genotyped from tail biopsies using PCR and Sanger sequencing and the positive animals were backcrossed to C57BL/6N mice.

Independent knockout lines for enhancer element *C* (*ΔC*) and *I* (*ΔI*) were generated. Heterozygous *ΔC* and *ΔI* ($T^{+/\Delta C}$ and $T^{+/\Delta I}$) and homozygous *ΔC* and *ΔI* ($T^{\Delta C/\Delta C}$ and $T^{\Delta I/\Delta I}$) embryos were investigated for potential overall phenotypes, but appeared phenotypically normal. Pups were born normally and grew up into fertile adults.

To generate a double knockout *ΔC,I* strain, homozygous $T^{\Delta C/\Delta C}$ mice were used for electroporation of CRISPR-Cas9 RNP complexes deleting enhancer element *I*. Pups homozygous for *ΔC,I* ($T^{\Delta C,I/\Delta C,I}$) were born phenotypically normal and developed into fertile adults; however, around 20% of the animals had a kinked tail (**Supplemental Fig. 5M,N**).

To generate a triple knockout *ΔT3,C,I* mouse strain, heterozygous *ΔC,I* ($T^{+/\Delta C,I}$) mice were used for electroporation of CRISPR-Cas9 RNP complexes deleting enhancer element *T3* (*ΔT3*). Heterozygous $T^{+/\Delta T3,C,I}$ or trans-heterozygous $T^{\Delta T3/\Delta C,I}$ embryos were phenotypically normal and grew up into fertile adults. To establish a single knockout line for enhancer element *T3* (*ΔT3*), $T^{\Delta T3/\Delta C,I}$ animals were outcrossed to establish $T^{+/\Delta T3}$.

$T^{\Delta C,I/\Delta T3,C,I}$ animals were generated by mating *ΔC,I* ($T^{\Delta C,I/\Delta C,I}$) and *ΔT3,C,I* ($T^{+/\Delta T3,C,I}$) strains and $T^{\Delta T3/\Delta T3,C,I}$ by mating *ΔT3* ($T^{+/\Delta T3}$) and *ΔT3,C,I* ($T^{+/\Delta T3,C,I}$) strains, respectively.

Around 60% of $T^{\Delta C,I/\Delta T3,C,I}$ pups were born with a tail defect and adult animals displayed a kinked tail, with around 2% of the $T^{\Delta C,I/\Delta T3,C,I}$ pups displaying a small tail. In contrast, adult trans-heterozygous $T^{\Delta T3/\Delta T3,C,I}$ and homozygous $T^{\Delta T3,C,I/\Delta T3,C,I}$ animals were never recovered likely due to lethality at around birth or during early postnatal life.

The deletion breakpoints in the individual enhancer alleles were determined by Sanger sequencing. Mice were genotyped using PCR with dedicated primer sets (**Supplemental Table 5)**. Mouse embryos of the given stage were harvested from timed pregnant mice. The day of plug was counted as embryonic day 0.5 (E0.5).

E9.5 whole mount immunostaining and imaging

E9.5 mouse embryos were collected and whole mount immunostaining was done as previously described (Mašek et al., 2016). Brachyury/T/Tbxt expression in E9.5 embryos was visualized using anti-Brachyury (ab209665, Abcam). Images were obtained using a Zeiss AxioZoom V16 macroscope with Apotome with a Zeiss Axiocam 512 mono camera. A qualitative analysis of all investigated embryos can be found in **Supplemental Table 6**.

E12.5 embryo preparation, immunostaining and imaging

E12.5 mouse embryos were collected and fixed overnight in 4% paraformaldehyde. Whole embryo images were acquired using a Olympus SZX9 stereo microscope with a Olympus DP72 camera. Afterwards, embryos were embedded in paraffin, and 9 µm-thick transverse sections were obtained using a Microtome Leica RM2255. Sections were deparaffinized, rehydrated, and stained with hematoxylin & eosin (H-3502, Vectorlabs) for histology, or anti-Brachyury (ab209665, Abcam), anti-Sox2 Y-17 (sc-17320, Santa Cruz), and DAPI (10236276001, Roche Diagnostics) according to the manufacturer's instructions. After staining, sections were mounted with Mowiol (81381, Millipore Sigma). Images of sections were obtained using a Leica DM6000 widefield fluorescence microscope with a Leica DFC 9000 camera.

Gar and turtle bridge alignment

To establish genomic connectivity across distant vertebrate lineages, a "bridging approach" that leverages species with slowly evolving genomic sequences, such as spotted gar within ray-finned fishes (Braasch et al., 2016) and painted turtle within tetrapods (Bradley Shaffer et al., 2013), was used. Using human *T3*, *C*, and *I* as queries, BLASTN searches (search sensitivity: distant homologies) at Ensembl.org against the bridge species genomes were performed. Candidate BLAST hit regions were manually inspected for their location in relation to the *Tbxtb* gene locus for further consideration. Core regions based on the initial BLAST hits in both bridge species were expanded in both directions up to the next annotated repeat element. Once the three elements were established in the bridge species, their sequences were used for as queries for BLASTN searches with genomes representative species across all major vertebrate lineages as targets (see **Supplemental Table 4** for species list, genome assemblies, and enhancer element coordinates). Further BLASTN chaining through additional species was performed as needed (e.g., human->gar->goldfish->zebrafish for *T3* and *I*). All BLAST hits were manually inspected for proximity to the *Tbxtb* gene. Multi-species alignments of the three elements were generated with MAFFT version 1.5.0 (Katoh et al., 2009) and presence of T-box motifs in individual species was established with FIMO version 5.4.1 (Grant et al., 2011) using the human

TBXT motif *TBXT_MA0009.2.meme* obtained from JASPAR as input sequence.

Acknowledgements

We thank Christine Archer, Molly Waters, Nikki Tsuji, Ainsley Gilbard, and Olivia Gomez (CU Anschutz), as well as Vesna Barros, Lukas Obernosterer, and Yorgos Panayotu (UZH) for excellent zebrafish husbandry support; Jitka Lachova (IMG) for outstanding expert assistance with mouse embryo preparation and strain maintenance; Beate Gruhl and Anja Wagner (TU Dresden) for excellent axolotl husbandry support; Dr. Alexa Sadier and Dr. Karen Sears (UCLA) for generously providing us with Monodelphis genomic DNA; Dr. Jelena Kresoja and Elena Cabello for bioinformatics input; and all members of our individual laboratories for critical input and discussion on experiments, concepts, and the manuscript. Species silhouettes adapted from PhyloPic.

Financial Support

This work has been supported by an UZH URPP "Translational Cancer Research" seed grant and NIH/NIDDK grant 1R01DK129350-01A1 to A.B.; the Children's Hospital Colorado Foundation, National Science Foundation Grant 2203311, and Swiss National Science Foundation (SNSF) Sinergia grant CRSII5_180345 to C.M.; a project grant from the Swiss Cancer League, the SwissBridge Award 2016 from the SwissBridge Foundation, Additional Ventures SVRF2021-1048003 grant, and the University of Colorado Anschutz Medical Campus to C.M. and A.B.; NIH/NIGMS 1T32GM141742-01, 3T32GM121742-02S1, and NIH/NHLBI F31HL167580 to H.R.M.; Czech Science Foundation grant 23-07056S to Z.K. and Czech Center for Phenogenomics infrastructure support LM2018126, OP VaVpI CZ.1.05/2.1.00/19.0395, and CZ.1.05/1.1.00 /02.0109 to IMG; NIH grant 5R01DE024745-09 to L.S.; NSF EDGE Award #2029216 to I.B.; SNSF Eccellenza professorship PCEFP3_186993 to M.O.; an Alexander von Humboldt postdoctoral Fellowship to A.C.; Deutsche Forschungsgemeinschaft (DFG) Grants 22137416, 450807335 and 497658823, as well as TUD and CRTD funds to M.H.Y.; NIH grant DP2HG010013 to E.K.F and F.L.; NIH grants R01HG003988, R01DE028599, and R01HL162304 to A.V. and B.J.M.

Author contributions

C.L.K., H.R.M., S.B., C.M. and A.B. – Zebrafish experiment design and performance, UCSC browser data analysis, manuscript writing.

D.K., A.C. and M.H.Y. – Axolotl experiment design and performance, data analysis, manuscript writing.

B.M., V.R., A.V. and M.O. – Mouse enhancer testing experiment design and performance, data analysis, manuscript writing.

V.H.A. and L.S. – Design and performance of experiments to assess enhancer activity on mouse embryonic sections, histology, data analysis, manuscript writing.

F.L. and E.F. – *Ciona* experiment design and performance, data analysis, manuscript writing.

J.S. and Z.K. – Mouse CRIPSR-Cas9 knockout experiment design and performance, data analysis, manuscript writing.

O.E.F. and I.B. – Bridge species research design, data analysis, manuscript writing.

Competing financial interests

The authors declare no competing financial interests.

REFERENCES

Amacher, S. L., Draper, B. W., Summers, B. R. and Kimmel, C. B. (2002). The zebrafish T-box genes no tail and spadetail are required for development of trunk and tail mesoderm and medial floor plate. *Development* 3311–3323.

Amemiya, C. T., Alfoldi, J., Lee, A. P., Fan, S., Philippe, H., MacCallum, I., Braasch, I., Manousaki, T., Schneider, I., Rohner, N., et al. (2013). The African coelacanth genome provides insights into tetrapod evolution. *Nat. 2013 4967445* **496**, 311–316.

Antosova, B., Smolikova, J., Klimova, L., Lachova, J., Bendova, M., Kozmikova, I., Machon, O. and Kozmik, Z. (2016). The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLOS Genet.* **12**, e1006441.

Armstrong, J. B. and Malacinski, G. M. (1989). Developmental biology of the axolotl. 320.

Arnold, S. J., Stappert, J., Bauer, A., Kispert, A., Herrmann, B. G. and Kemler, R. (2000). Brachyury is a target gene of the Wnt/β-catenin signaling pathway. *Mech. Dev.* **91**, 249–258.

Bagnat, M. and Gray, R. S. (2020). Development of a straight vertebrate body axis. *Development* **147**,.

Beisaw, A., Tsaytler, P., Koch, F., Schmitz, S. U., Melissari, M., Senft, A. D., Wittler, L., Pennimpede, T., Macura, K., Herrmann, B. G., et al. (2018). BRACHYURY directs histone acetylation to target loci during mesoderm development. *EMBO Rep.* **19**, 118.

Bhadra, A. K. and Casey, A. T. (2006). Familial chordoma. A report of two cases. *J Bone Jt. Surg Br* **88**, 634–636.

Braasch, I., Gehrke, A. R., Smith, J. J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., et al. (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* **48**, 427–437.

Bradley Shaffer, H., Minx, P., Warren, D. E., Shedlock, A. M., Thomson, R. C., Valenzuela, N., Abramyan, J., Amemiya, C. T., Badenhorst, D., Biggar, K. K., et al. (2013). The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**,.

Brink, S. C. van den, Alemany, A., Batenburg, V. van, Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K. F., Arias, A.

M., et al. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nat. 2020* 1–5.

Cannavò, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korbel, J. O. and Furlong, E. E. M. (2015). Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol.*

Choi, K. S., Cohn, M. J. and Harfe, B. D. (2008). Identification of nucleus pulposus precursor cells and notochordal remnants in the mouse: implications for disk degeneration and chordoma formation. *Dev Dyn* **237**, 3953–3958.

Christiaen, L., Wagner, E., Shi, W. and Levine, M. (2009). Electroporation of Transgenic DNAs in the Sea Squirt Ciona. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5345.

Clements, D., Taylor, H. C., Herrmann, B. G. and Stott, D. (1996). Distinct regulatory control of the Brachyury gene in axial and non-axial mesoderm suggests separation of mesoderm lineages early in mouse gastrulation. *Mech. Dev.* **56**, 139–149.

Corallo, D., Trapani, V. and Bonaldo, P. (2015). The notochord: structure and functions. *Cell. Mol. Life Sci.* **72**, 2989–3008.

Corbo, J. C., Levine, M. and Zeller, R. W. (1997). Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. *Development* **124**, 589–602.

D'Agati, G., Cabello, E. M., Frontzek, K., Rushing, E. J., Klemm, R., Robinson, M. D., White, R. M., Mosimann, C. and Burger, A. (2019). Active receptor tyrosine kinases, but not Brachyury, are sufficient to trigger chordoma in zebrafish. *Dis. Model. Mech.* **12**,.

Dobrovolskaia-Zavadskaia, N. (1927). Sur la mortification spontanee de la chez la souris nouveau-nee et sur l'existence d'un caractere (facteur) hereditaire, non-viable. *Crit Rev Soc Biol* **97**, 114–116.

Faial, T., Bernardo, A. S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G. E., Mascetti, V. L., Trotter, M. W. B., Smith, J. C., Pedersen, R. A., et al. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* **142**, 2121.

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. and Levine, M. S. (2015). Suboptimization of developmental enhancers. *Science* **350**, 325–8.

Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. and Levine, M. S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6508–13.

Felker, A. and Mosimann, C. (2016). Contemporary zebrafish transgenesis with Tol2 and application for Cre/lox recombination experiments. *Methods Cell Biol.* **135**, 219–44.

Gluecksohn-Schoenheimer, S. (1938). THE DEVELOPMENT OF TWO TAILLESS MUTANTS IN THE HOUSE MOUSE. *Genetics* **23**, 573–584.

Gluecksohn-Schoenheimer, S. (1944). The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick. *Proc. Natl. Acad. Sci.* **30**, 134–140.

Goodstadt, L., Heger, A., Webber, C. and Ponting, C. P. (2007). An analysis of the gene complement of a marsupial, Monodelphis domestica: evolution of lineage-specific genes and giant chromosomes. *Genome Res.* **17**, 969–981.

Grant, C. E., Bailey, T. L. and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017.

Halpern, M. E., Ho, R. K., Walker, C. and Kimmel, C. B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* **75**, 99–111.

Harafuji, N., Keys, D. N. and Levine, M. (2002). Genome-wide identification of tissue-specific enhancers in the Ciona tadpole. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6802–6805.

Harvey, S. A. A., Tümpel, S., Dubrulle, J., Schier, A. F. F., Smith, J. C. C., Tumpel, S., Dubrulle, J., Schier, A. F. F. and Smith, J. C. C. (2010). no tail integrates two modes of mesoderm induction. *Development* **137**, 1127–1135.

Heaton, J. M. and Turner, D. R. (1985). Reflections on notochordal differentiation arising from a study of chordomas. *Histopathology* **9**, 543–550.

Henrique, D., Abranches, E., Verrier, L. and Storey, K. G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864–2875.

Herrmann, B. G. (1995). The mouse Brachyury (T) gene. *Semin. Dev. Biol.* **6**, 385–394.

Hippenmeyer, S., Youn, Y. H., Moon, H. M., Miyamichi, K., Zong, H., Wynshaw-Boris, A. and Luo, L. (2010). Genetic Mosaic Dissection of Lis1 and Ndel1 in Neuronal Migration. *Neuron* **68**, 695.

Holland, P. W., Koschorz, B., Holland, L. Z. and Herrmann, B. G. (1995). Conservation of Brachyury (T) genes in amphioxus and vertebrates: developmental and evolutionary implications. *Development* **121**, 4283–4291.

Hong, J. W., Hendrix, D. A. and Levine, M. S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science (80-.).* **321**, 1314.

Hsu, W., Mohyeldin, A., Shah, S. R., ap Rhys, C. M., Johnson, L. F., Sedora-Roman, N. I., Kosztowski, T. A., Awad, O. A., McCarthy, E. F., Loeb, D. M., et al. (2011). Generation of chordoma cell line JHC7 and the identification of Brachyury as a novel molecular target. *J Neurosurg* **115**, 760–769.

Hu, Y., Mintz, A., Shah, S. R. R., Quinones-Hinojosa, A. and Hsu, W. (2014). The FGFR/MEK/ERK/brachyury pathway is critical for chordoma cell growth and survival. *Carcinogenesis* **35**, 1491–1499.

Inoue, J., Yasuoka, Y., Takahashi, H. and Satoh, N. (2017). The chordate ancestor possessed a single

copy of the Brachyury gene for notochord acquisition. *Zool. Lett.* **3**, 1–7.

Irimia, M., Tena, J. J., Alexis, M. S., Fernandez-Miñan, A., Maeso, I., Bogdanović, O., De La Calle-Mustienes, E., Roy, S. W., Gómez-Skarmeta, J. L. and Fraser, H. B. (2012). Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* **22**, 2356–2367.

Jenickova, I., Kasparek, P., Petrezselyova, S., Elias, J., Prochazka, J., Kopkanova, J., Navratil, M., Barinka, C. and Sedlacek, R. (2021). Efficient allele conversion in mouse zygotes and primary cells based on electroporation of Cre protein. *Methods* **191**, 87–94.

Katoh, K., Asimenos, G. and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64.

Kemmler, C. L., Moran, H. R., Murray, B. F., Scoresby, A., Klem, J. R., Eckert, R. L., Lepovsky, E., Bertho, S., Nieuwenhuize, S., Burger, S., et al. (2023). Next-generation plasmids for transgenesis in zebrafish and beyond. *Development* **150**,.

Khattak, S., Murawala, P., Andreas, H., Kappert, V., Schuez, M., Sandoval-Guzmán, T., Crawford, K. and Tanaka, E. M. (2014). Optimized axolotl (Ambystoma mexicanum) husbandry, breeding, metamorphosis, transgenesis and tamoxifen-mediated recombination. *Nat. Protoc.* **9**, 529–540.

Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K. and Herrmann, B. G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev. Cell*.

Kumar, S., Stecher, G., Suleski, M. and Blair Hedges, S. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819.

Kvon, E. Z., Zhu, Y., Kelman, G., Novak, C. S., Plajzer-Frick, I., Kato, M., Garvin, T. H., Pham, Q., Harrington, A. N., Hunter, R. D., et al. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell* **180**, 1262.

Kvon, E. Z., Waymack, R., Elabd, M. G., Wunderlich, Z., Gad, M. and Wunderlich, Z. (2021). Enhancer redundancy in development and disease. **22**,.

Kwan, K. M., Fujimoto, E., Grabher, C., Mangum, B. D., Hardy, M. E., Campbell, D. S., Parant, J. M., Yost, H. J., Kanki, J. P. and Chien, C. Bin (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev Dyn* **236**, 3088–3099.

Latinkić, B. V., Umbhauer, M., Neal, K. A., Lerchner, W., Smith, J. C. and Cunliffe, V. (1997). The Xenopus Brachyury promoter is activated by FGF and low concentrations ofactivinandsuppressed by high concentrationsof activin and by paired-type homeodomain proteins. *Genes Dev.* **11**, 3265–3276.

Letelier, J., De La Calle-Mustienes, E., Pieretti, J., Naranjo, S., Maeso, I., Nakamura, T., Pascual-Anaya, J., Shubin, N. H., Schneider, I., Martinez-Morales, J. R., et al. (2018). A conserved Shh cis-regulatory module highlights a common developmental origin of unpaired and paired fins. *Nat. Genet.* **50**, 504–509.

Lópezlópez-Anguita, N., Gassaloglu, S. I., Stö Tzel, M., Bolondi, A., Conkar, D., Typou, M., Renébuschow, R., Veenvliet, J. V and Bulut-Karslioglu, A. (2022). Hypoxia induces an early primitive streak signature, enhancing spontaneous elongation and lineage representation in gastruloids. *Development* **149**,.

Martin, B. L. and Kimelman, D. (2008). Regulation of Canonical Wnt Signaling by Brachyury Is Essential for Posterior Mesoderm Formation. *Dev. Cell* **15**, 121–133.

Martin, B. L. and Kimelman, D. (2010). Brachyury establishes the embryonic mesodermal progenitor niche. *Genes Dev.* **24**, 2778–83.

Mašek, J., Machoň, O., Kořínek, V., Taketo, M. M. and Kozmik, Z. (2016). Tcf7l1 protects the anterior neural fold from adopting the neural crest fate. *Development* **143**, 2206–2216.

Mikkelsen, T. S., Wakefield, M. J., Aken, B., Amemiya, C. T., Chang, J. L., Duke, S., Garber, M., Gentles, A. J., Goodstadt, L., Heger, A., et al. (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447**, 167–77.

Moris, N., Anlas, K., van den Brink, S. C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A. and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. *Nature* 1–6.

Mosimann, C. (2022). Multisite Gateway Calculations: Excel spreadsheet. *protocols.io*.

Mosimann, C., Kaufman, C. K., Li, P., Pugach, E. K., Tamplin, O. J. and Zon, L. I. (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development* **138**, 169–177.

Nakamichi, R. and Asahara, H. (2020). The transcription factors regulating intervertebral disc development. *JOR Spine* **3**, e1081.

Nei, M., Xu, P. and Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci.* **98**, 2497–2502.

Nelson, A. C., Pillay, N., Henderson, S., Presneau, N. N., Tirabosco, R., Halai, D., Berisha, F., Flicek, P., Stemple, D. L., Stern, C. D., et al. (2012). An integrated functional genomics approach identifies the regulatory network directed by brachyury (T) in chordoma. *J. Pathol.* **228**, 274–285.

Nibu, Y., Jose-Edwards, D. S., Di Gregorio, A., Jos??-Edwards, D. S. and Di Gregorio, A. (2013). *From notochord formation to hereditary chordoma: The many roles of brachyury*.

Nowoshilow, S., Schloissnig, S., Fei, J. F., Dahl, A., Pang, A. W. C., Pippel, M., Winkler, S., Hastie, A. R., Young, G., Roscito, J. G., et al. (2018). The axolotl genome and the evolution of key tissue

formation regulators. *Nat. 2018 5547690* **554**, 50–55.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243.

Osterwalder, M., Tran, S., Hunter, R. D., Meky, E. M., von Maydell, K., Harrington, A. N., Godoy, J., Novak, C. S., Plajzer-Frick, I., Zhu, Y., et al. (2022). Characterization of Mammalian In Vivo Enhancers Using Mouse Transgenesis and CRISPR Genome Editing. *Methods Mol. Biol.* **2403**, 147–186.

Passamaneck, Y. J., Katikala, L., Perrone, L., Dunn, M. P., Oda-Ishii, I. and Di Gregorio, A. (2009). Direct activation of a notochord cis-regulatory module by Brachyury and FoxA in the ascidian Ciona intestinalis. *Development* **136**, 3679–3689.

Peck, S. H., McKee, K. K., Tobias, J. W., Malhotra, N. R., Harfe, B. D. and Smith, L. J. (2017). Whole Transcriptome Analysis of Notochord-Derived Cells during Embryonic Formation of the Nucleus Pulposus. *Sci. Reports 2017 71* **7**, 1–14.

Pennimpede, T., Proske, J., König, A., Vidigal, J. A., Morkel, M., Bramsen, J. B., Herrmann, B. G. and Wittler, L. (2012). In vivo knockdown of Brachyury results in skeletal defects and urorectal malformations resembling caudal regression syndrome. *Dev. Biol.* **372**, 55–67.

Pippucci, T., Savoia, A., Perrotta, S., Pujol-Moix, N., Noris, P., Castegnaro, G., Pecci, A., Gnan, C., Punzo, F., Marconi, C., et al. (2011). Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am. J. Hum. Genet.* **88 1**, 115–120.

Presneau, N., Shalaby, A., Ye, H., Pillay, N., Halai, D., Idowu, B., Tirabosco, R., Whitwell, D., Jacques, T. S., Kindblom, L.-G. G., et al. (2011). Role of the transcription factor T (brachyury) in the pathogenesis of sporadic chordoma: a genetic and functional-based study. *J Pathol* **223**, 327–335.

Prummel, K. D., Hess, C., Nieuwenhuize, S., Parker, H. J., Rogers, K. W., Kozmikova, I., Racioppi, C., Brombacher, E. C., Czarkwiani, A., Knapp, D., et al. (2019). A conserved regulatory program initiates lateral plate mesoderm emergence across chordates. *Nat. Commun.* **10**, 3857.

Rennebeck, G. M., Lader, E., Chen, Q., Bohm, R. a, Cai, Z. S., Faust, C., Magnuson, T., Pease, L. R. and Artzt, K. (1995). Is there a Brachyury the Second? Analysis of a transgenic mutation involved in notochord maintenance in mice. *Dev. Biol.* **172**, 206–217.

Rennebeck, G., Lader, E., Fujimoto, A., Lei, E. P. and Artzt, K. (1998). Mouse Brachyury the second (T2) is a gene next to classical T and a candidate gene for tct. *Genetics* **150**, 1125–1131.

Richardson, S. M., Ludwinski, F. E., Gnanalingham, K. K., Atkinson, R. A., Freemont, A. J. and Hoyland, J. A. (2017). Notochordal and nucleus pulposus marker expression is maintained by sub-populations of adult human nucleus pulposus cells through aging and degeneration. *Sci. Reports 2017 71* **7**, 1–11.

Risbud, M. V, Schaer, T. P. and Shapiro, I. M. (2010). Toward an understanding of the role of notochordal cells in the adult intervertebral disc: from discord to accord. *Dev. Dyn.* **239**, 2141–8.

Rito, T., Libby, A. R. G., Demuth, M. and Briscoe, J. (2023). Notochord and axial progenitor generation by timely BMP and NODAL inhibition during vertebrate trunk formation. *bioRxiv* 2023.02.27.530267.

Rivera-Pérez, J. A. and Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. *Dev. Biol.* **288**, 363–371.

Sangoi, A. R., Karamchandani, J., Lane, B., Higgins, J. P., Rouse, R. V, Brooks, J. D. and McKenney, J. K. (2011). Specificity of brachyury in the distinction of chordoma from clear cell renal cell carcinoma and germ cell tumors: a study of 305 cases. *Mod Pathol* **24**, 425–429.

Satoh, N., Tagawa, K. and Takahashi, H. (2012). How was the notochord born? **14**, 56–75.

Schifferl, D., Scholze-Wittler, M., Wittler, L., Veenvliet, J. V., Koch, F. and Herrmann, B. G. (2021). A 37 kb region upstream of Brachyury comprising a notochord enhancer is essential for notochord and tail development. *Development*.

Schulte-Merker, S., van Eeden, F. J., Halpern, M. E., Kimmel, C. B. and Nüsslein-Volhard, C. (1994). no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**, 1009–15.

Schwaner, M. J., Hsieh, S. T., Swalla, B. J. and McGowan, C. P. (2021). An Introduction to an Evolutionary Tail: EvoDevo, Structure, and Function of Post-Anal Appendages. *Integr. Comp. Biol.* **61**, 352–357.

Sebé-Pedrós, A., Ariza-Cosano, A., Weirauch, M. T., Leininger, S., Yang, A., Torruella, G., Adamski, M., Adamska, M., Hughes, T. R., Gómez-Skarmeta, J. L., et al. (2013). Early eèolution of the T-box transcription factor family. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16050–16055.

Sharifnia, T., Wawer, M. J., Chen, T., Huang, Q.-Y., Weir, B. A., Sizemore, A., Lawlor, M. A., Goodale, A., Cowley, G. S., Vazquez, F., et al. (2019). Small-molecule targeting of brachyury transcription factor addiction in chordoma. *Nat. Med.* 1.

Smith, J. C., Price, B. M. J. J., Green, J. B. A. A., Weigel, D. and Herrmann, B. G. (1991). Expression of a xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79–87.

Song, B. P., Ragsac, M. F., Tellez, K., Jindal, G. A., Grudzien, J. L., Le, S. H. and Farley, E. K. (2023). Diverse logics and grammar encode notochord enhancers. *Cell Rep.* **42**,.

Stemple, D. L. (2005). Structure and function of the notochord: an essential organ for chordate development. *Development* **132**, 2503–12.

Stosiek, P., Kasper, M. and Karsten, U. (1988).

Expression of cytokeratin and vimentin in nucleus pulposus cells. *Differentiation* **39**, 78–81.

Takei, H. and Powell, S. Z. (2010). Novel immunohistochemical markers in the diagnosis of nonglial tumors of nervous system. *Adv Anat Pathol* **17**, 150–153.

Takezaki, N. (2018). Global Rate Variation in Bony Vertebrates. *Genome Biol. Evol.* **10**, 1803–1815.

Tang, X., Jing, L. and Chen, J. (2012). Changes in the Molecular Phenotype of Nucleus Pulposus Cells with Intervertebral Disc Aging. *PLoS One* **7**, 52020.

Tarpey, P. S., Behjati, S., Young, M. D., Martincorena, I., Alexandrov, L. B., Farndon, S. J., Guzzo, C., Hardy, C., Latimer, C., Butler, A. P., et al. (2017). The driver landscape of sporadic chordoma. *Nat. Commun.* **8**, 1–6.

Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y. and Luo, L. (2011). Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7902–7.

Thompson, A. W., Hawkins, M. B., Parey, E., Wcisel, D. J., Ota, T., Kawasaki, K., Funk, E., Losilla, M., Fitch, O. E., Pan, Q., et al. (2021). The bowfin genome illuminates the developmental evolution of ray-finned fishes. *Nat. Genet. 2021* 1–12.

Tosic, J., Kim, G.-J., Pavlovic, M., Schröder, C. M., Mersiowsky, S.-L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L., et al. (2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nat. Cell Biol.* **21**, 1518–1531.

Veenvliet, J. V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Kumar, A. S., Pustet, M., et al. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science (80-.).* **370**,.

Vujovic, S., Henderson, S., Presneau, N., Odell, E., Jacques, T. S., Tirabosco, R., Boshoff, C. and Flanagan, A. M. (2006). Brachyury, a crucial regulator of notochordal development, is a novel biomarker for chordomas. *J Pathol* **209**, 157–165.

Wang, F., Zhang, C., Shi, R., Xie, Z.-Y., Chen, L., Wang, K., Wang, Y.-T., Xie, X.-H. and Wu, X.-T. (2018). The embryonic and evolutionary boundaries between notochord and cartilage: A new look at nucleus pulposus-specific markers. *Osteoarthr. Cartil.*

Xu, P.-F., Houssin, N., Ferri-Lagneau, K. F., Thisse, B. and Thisse, C. (2014). Construction of a vertebrate embryo from two opposing morphogen gradients. *Science* **344**, 87–9.

Xu, P. F., Borges, R. M., Fillatre, J., de Oliveira-Melo, M., Cheng, T., Thisse, B. and Thisse, C. (2021). Construction of a mammalian embryo model from stem cells organized by a morphogen signalling centre. *Nat. Commun.* **12**,.

Yakkioui, Y., van Overbeeke, J. J., Santegoeds, R., van Engeland, M. and Temel, Y. (2014). The origin of chordoma. *Biochim. Biophys. Acta*.

Yanagisawa, K. O. (1990). Does the T gene determine the anteroposterior axis of a mouse embryo? *Japanese J. Genet.* **65**, 287–297.

Yang, X. R., Ng, D., Alcorta, D. A., Liebsch, N. J., Sheridan, E., Li, S., Goldstein, A. M., Parry, D. M. and Kelley, M. J. (2009). T (brachyury) gene duplication confers major susceptibility to familial chordoma. *Nat Genet* **41**, 1176–1178.

Zhu, J., Kwan, K. M. and Mackem, S. (2016). Putative oncogene Brachyury (T) is essential to specify cell fate but dispensable for notochord progenitor proliferation and EMT. *Proc. Natl. Acad. Sci. U. S. A.* 1601252113-.

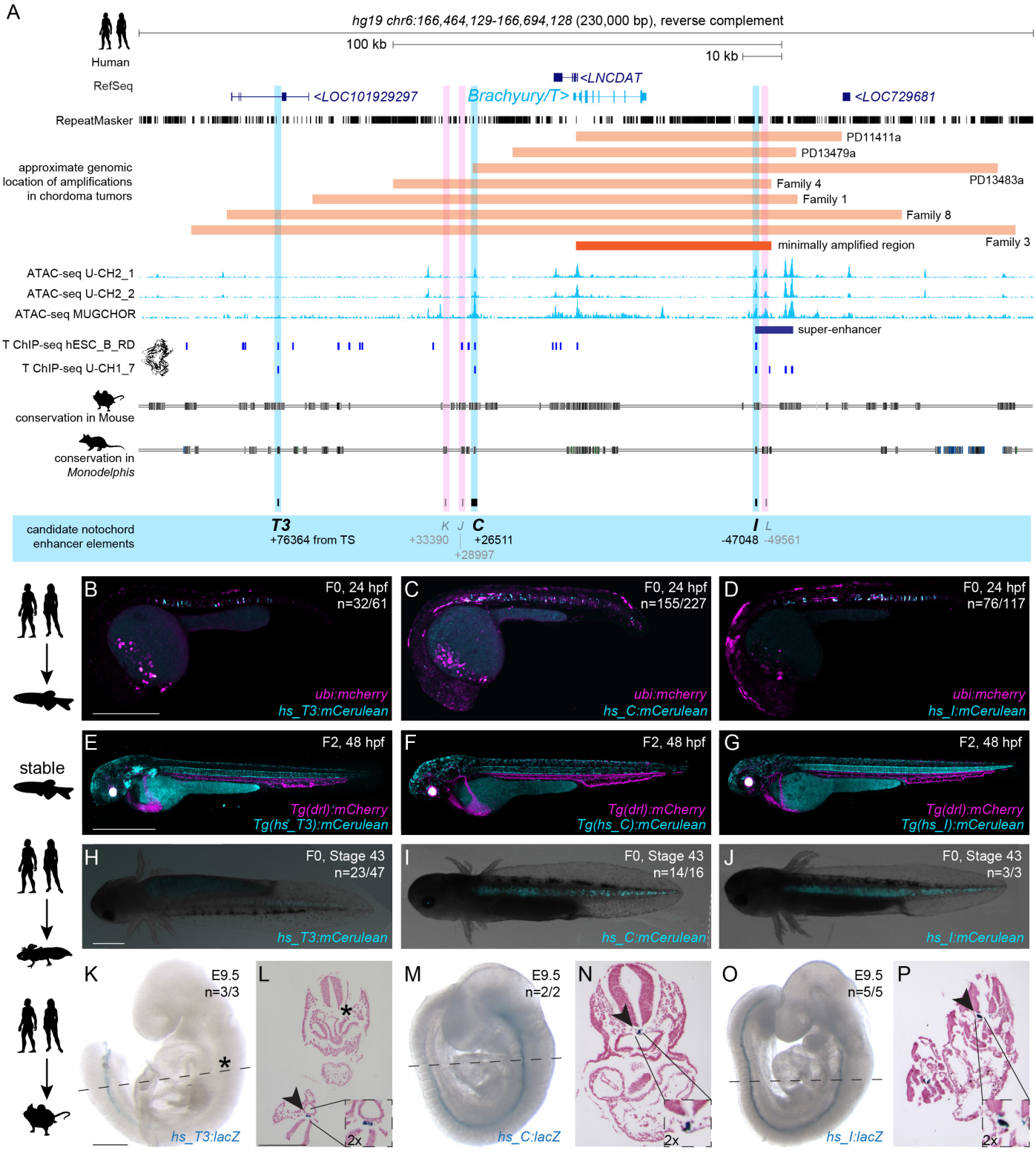


Figure 1: Human *Brachyury* enhancer elements *T3*, *C* and *I* are active in different species.
(**A**) Human *Brachyury/T/TBXTB* locus with surrounding gene loci adapted from UCSC genome browser. Repeats marked in black using the RepeatMasker track. Further annotated are approximate amplifications (light orange) and the minimally amplified region (dark orange) in chordoma tumors. ATAC-sequencing (light blue peaks) and T ChIP-sequencing (dark blue lines) suggest enhancer elements (light pink highlight, not active; light blue highlight, active) that are conserved in mouse and the marsupial *Monodelphis domestica*.
(**B,C,D**) Representative F0 zebrafish embryos injected with the human enhancer elements *hs_T3* (**B**), *hs_C* (**C**), and *hs_I* (**D**) showing mosaic *mCerulean* reporter expression in the notochord at 24 hpf and expression of *ubi:mCherry* as injection control. N represents the number of animals expressing mCerulean in the notochord relative to the total number of animals expressing mCherry. Scale bar in **B**: 0.5 mm, applies to **B-C**.
(**E,F,G**) Representative images of stable transgenic F2 embryos at 48 hpf for each of the human enhancer elements *hs_T3*, *hs_C*, and *hs_I* crossed to *Tg(drl:mCherry)* that labels lateral plate mesoderm and later cardiovascular lineages. Transgenic F2 embryos recapitulate the F0 expression pattern in the notochord, with *hs_T3* (**E**) additionally expressing

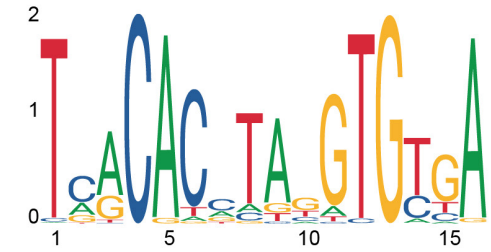
cerulean in the pharyngeal arches and fin, and *hs_I* (**G**) in the proximal kidney close to the anal pore. Enhancer element *hs_C* (**F**) stable transgenic lines have lower relative notochord reporter activity than *hs_T3* and *hs_I*. Scale bar in **E**: 0.5 mm, applies to **E-G**.

(**H,I,J**) Representative F0 axolotl embryos at peri-hatching stages expressing mCerulean from the human enhancers *hs_T3* (**G**), *hs_C* (**H**), *hs_I* (**I**). N represent the number of animals expressing mCerulean in the notochord relative to the total number of animals showing any mCerulean expression. Scale bar in **H**: 1 mm, applies to **H-J**.

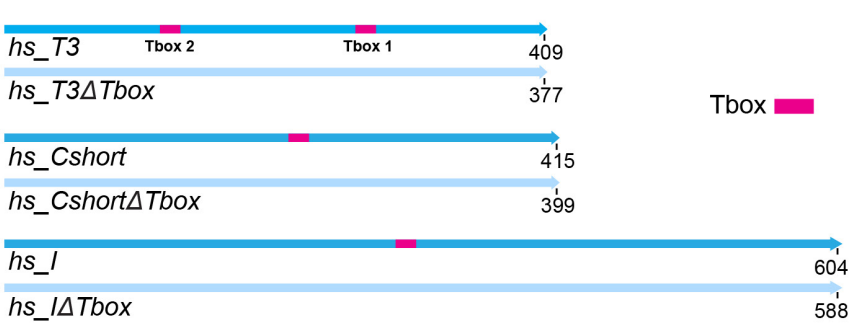
(**K,M,O**) Representative images of transgenic E9.5 mouse embryos expressing *lacZ* (encoding beta-galactosidase) under the human enhancers *hs_T3* (**K**), *hs_C* (**M**), and *hs_I* (**O**) visualized with X-gal whole-mount staining. While *hs_C* and *hs_I* express beta-galactosidase in the entire notochord, beta-galactosidase expression from *hs_T3* is restricted to the posterior notochord. Black asterisk marks absence of beta-galactosidase in the anterior notochord. N represent the number of animals expressing beta-galactosidase in the notochord relative to the total number of animals with tandem integrations at *H11*. Dotted lines represent the sectioning plane. Scale bar in **K**: 0.5 mm, applies to **K**,**M**,**O**.

(**L,N,P**) Representative images of Fast Red-stained cross sections from embryos shown on the left, *hs_T3* (**L**), *hs_C* (**N**), and *hs_I* (**P**). Black arrowheads point at notochord, and inserts show notochords at 2x higher magnification. Scale bar in **L**: 0.25 mm, applies to **L**,**N**,**P**.

A



C



B

Motif ID	Name	Input	Length	Start	Stop	Strand	Score	P value	Q value	Matched sequence
MA0009.2	TBXT	*hs_T3*	409 bp	130	145	+	7.80769	7.00E-05	0.0323	CCTCACCTTTTTGTGA
MA0009.2	TBXT	*hs_T3*	409 bp	130	145	-	8.38462	5.63E-05	0.0312	TCACAAAAAGGTGAGG
MA0009.2	TBXT	*hs_T3*	409 bp	277	292	+	9.03846	4.38E-05	0.0312	TCACAGTTTGCTGTGA
MA0009.2	TBXT	*hs_T3*	409 bp	277	292	-	8.96154	4.51E-05	0.0312	TCACAGCAAACTGTGA
MA0009.2	TBXT	*hs_C*	415 bp	184	199	+	-7.88462	0.00473	1	CCGCAGTTGGCTGGGG
MA0009.2	TBXT	*hs_C*	415 bp	184	199	-	-7.82692	0.00468	1	CCCCAGCCAACTGCGG
MA0009.2	TBXT	*hs_I*	604 bp	309	324	+	10.7692	2.16E-05	0.0312	TCCCACATAGGTGAGG
MA0009.2	TBXT	*hs_I*	604 bp	309	324	-	9.57692	3.53E-05	0.0312	CCTCACCTATGTGGGA

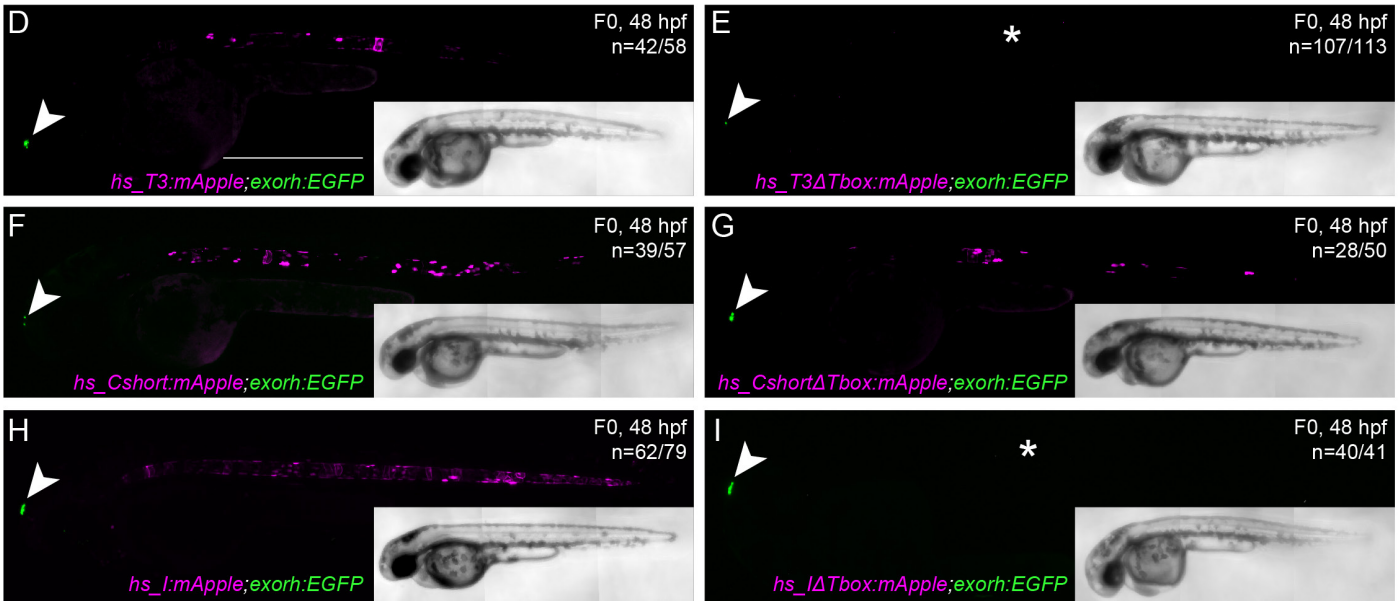


Figure 2: Identified TBXT binding sites in the enhancer elements are essential for reporter activity.
(**A**) Sequence of the human TBXT binding site (T-box) using JASPAR.
(**B**) FIMO output with location of the T-box, statistical significance, and matched sequence within the enhancer elements.
(**C**) Schematic depiction of the wildtype human enhancer elements with the TBXT binding site/T-box (pink box) and the enhancer elements without the respective T-box sites (*ΔTbox*). All human enhancer elements are depicted in the reverse complement direction.
(**D-I**) Injection of the wildtype enhancer elements *hs_T3* (**D**), *hs_C* (**F**), *and hs_I* (**H**) as reporter constructs results in mApple fluorophore expression in the notochord at 48 hpf, whereas injection of *hs_T3ΔTbox* (**E**) and *hs_IΔTbox* (**I**) show complete loss of notochord expression (asterisks in **E,I**) in comparison to *hs_CshortΔTbox* (**G**), which has residual reporter activity even without the T-box. Arrowheads (**D-I**) mark EGFP expression in the pineal gland from the transgenesis marker *exorh:EGFP*. Scale bar in **D**: 0.5 mm, applies to **D-I**.

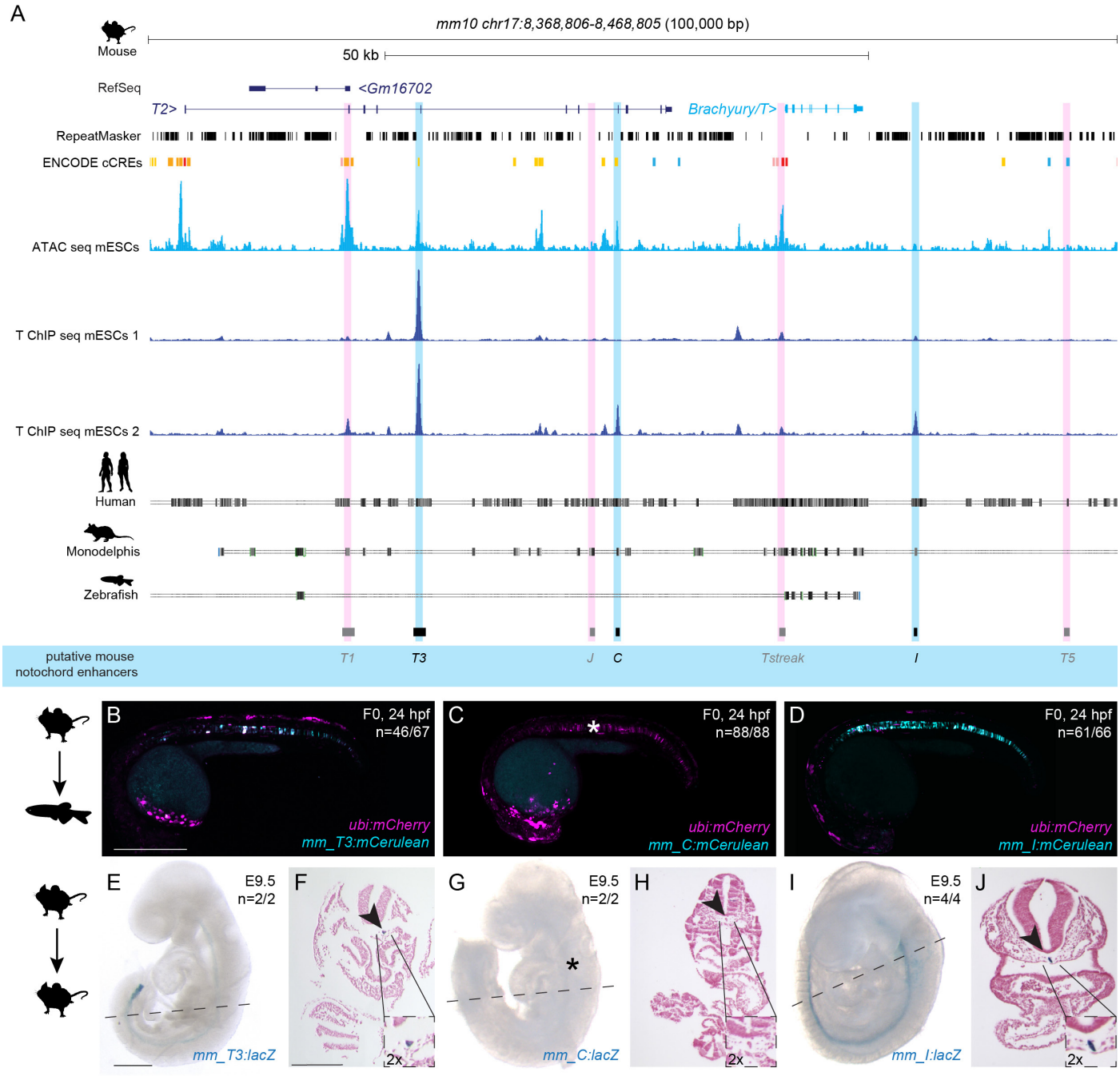


Figure 3: Mouse *Brachyury* enhancer elements are active in different species.
(**A**) Mouse *Brachyury/T/TBXTB* locus adapted from the UCSC genome browser. Repeats marked in black using the RepeatMasker track. Further annotated is the track with ENCODE conserved *cis*-regulatory elements (cCREs). ATAC-sequencing (light blue peaks) and T ChIP-sequencing (dark blue lines) indicate enhancer elements (light pink highlight, not active; light blue highlight, active) that are conserved in human and Monodelphis.
(**B,C,D**) Representative F0 zebrafish embryos injected with the mouse enhancer elements *mm_T3* (**B**), *mm_C* (**C**), and *mm_I* (**D**). *mm_T3* and *mm_I* show mosaic *mCerulean* reporter expression in the notochord at 24 hpf and mosaic *ubi:mCherry* expression as injection control. Mouse enhancer element *mm_C* is not active in the zebrafish notochord (asterisk in C). N represent the number of animals expressing mCerulean in the notochord relative to the total number of animals expressing mCherry. Scale bar in B: 0.5 mm, applies to **B-D**.
(**E,G,I**) Representative images of transgenic E9.5 mouse embryos expressing *lacZ* (encoding beta-galactosidase) under the mouse enhancer elements *mm_T3* (**E**), *mm_C* (**G**) and *mm_I* (**I**) visualized with X-gal whole mount staining. While *mm_T3* and *mm_I* express beta-galactosidase in the entire notochord, beta-galactosidase expression from mouse *mm_C* is absent (asterisk in **G**). N represent the number of animals expressing beta-galactosidase in the notochord relative to the total number of animals with tandem integrations at *H11*. Dotted lines represent the sectioning plane. Scale bar in **E**: 0.5 mm, applies to E,G,I.
(**F,H,J**) Representative images of Fast Red-stained cross sections from embryos shown on the left, *mm_T3* (**F**), *mm_C* (**H**), and *mm_I* (**J**). Black arrowheads point at notochord, and inserts show notochords at 2x higher magnification. Scale bar in **F**: 0.25 mm, applies to **F,H,J**.

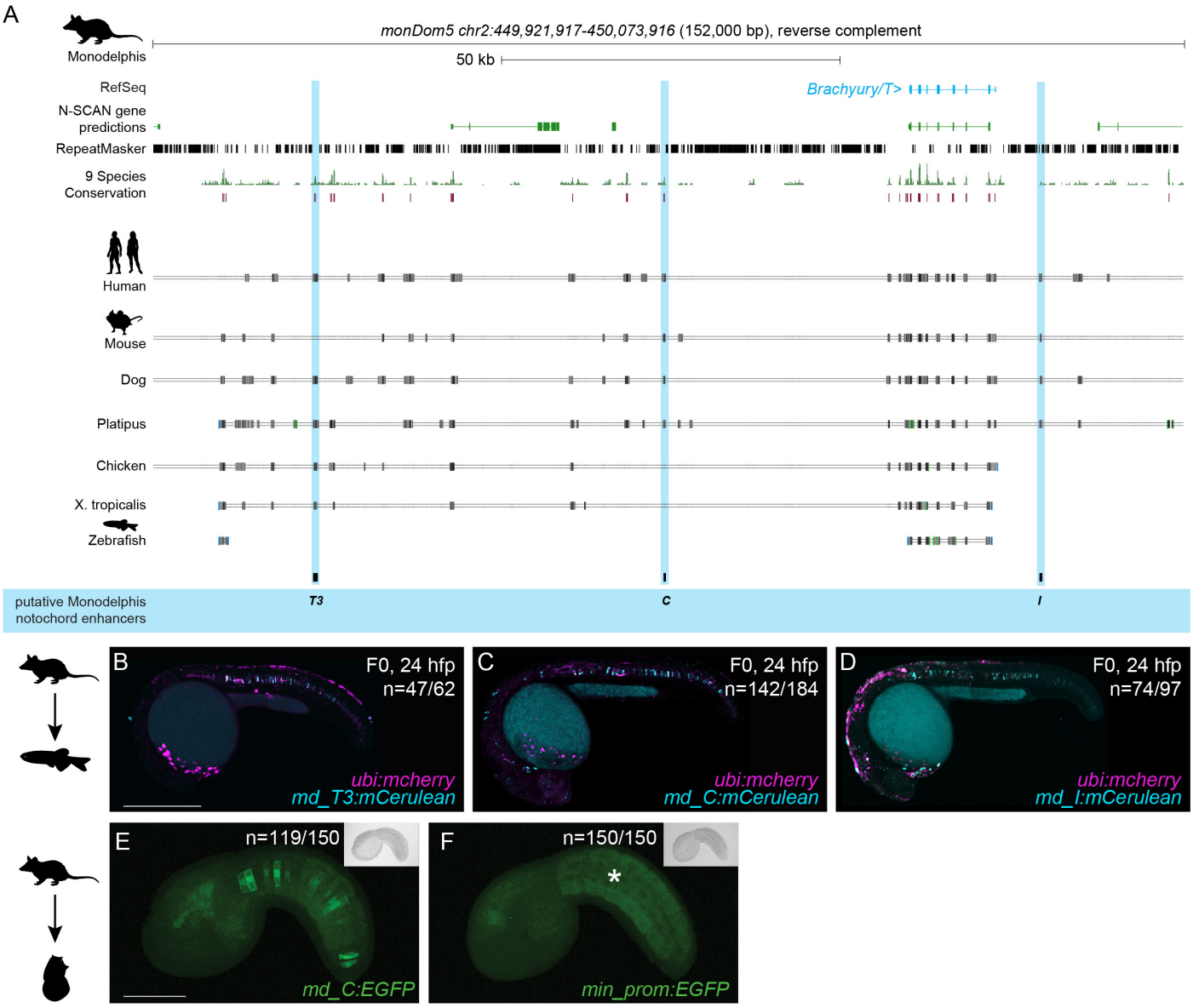


Figure 4: Monodelphis *Brachyury* enhancer elements are active in different species.
(**A**) Monodelphis *Brachyury/T/TBXTB* locus adapted from the UCSC genome browser. Repeats are marked in black using the RepeatMasker track. Further annotated are tracks containing N-SCAN gene predictions and 9 Species Conservation. The light blue highlighted boxes mark the Monodelphis enhancer elements *T3*, *C* and *I* and their conservation in other species.
(**B,C,D**) Representative F0 zebrafish embryos injected with the Monodelphis enhancer elements *md_T3* (**B**), *md_C* (**C**), and *md_I* (**D**) showing mosaic *mCerulean* reporter expression in the zebrafish notochord at 24 hpf. *ubi:mCherry* was used as injection control. N represent the number of animals expressing mCerulean in the notochord relative to the total number of animals expressing mCherry. Scale bar in **B**: 0.5 mm, applies to **B-C**.
(**E,F**) Representative images of *Ciona* embryos electroporated with Monodelphis enhancer element *md_C* (**E**), and minimal *forkhead* promoter (*fkh*) only as control (**F**). Monodelphis enhancer element *md_C* expresses EGFP throughout the entire *Ciona* notochord, compared to minimal *fkh* promoter only which does not express EGFP at all (asterisk in **F**). N represent the number of animals expressing EGFP in the notochord relative to the total number of animals. Inserts on the top right represent bright field images of respective embryos. Scale bar in **E**: 0.05 mm, applies to **E**,**F**.

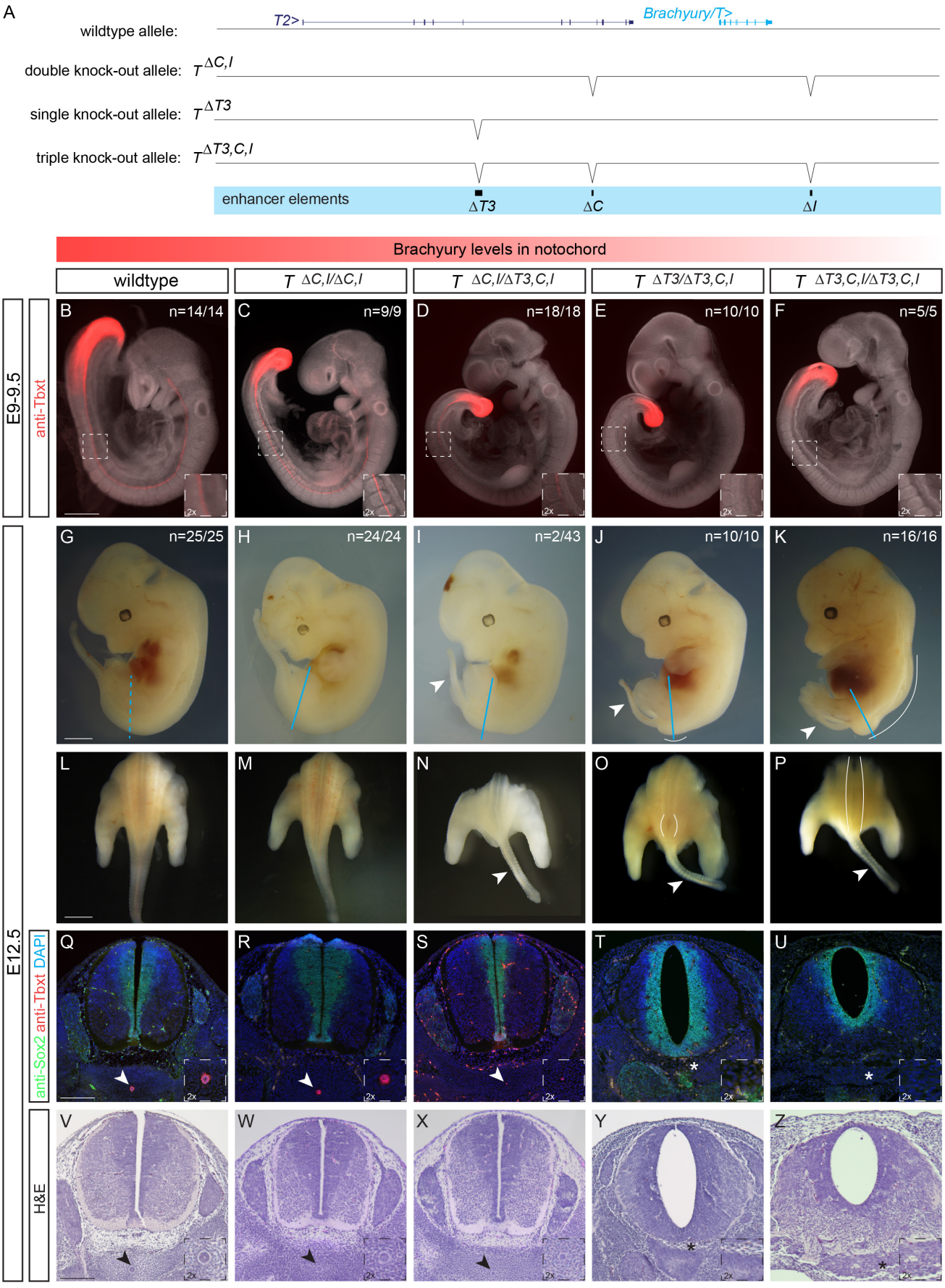


Figure 5: Deletion of the three enhancer elements *T3*, *C* and *I* results in selective loss of Brachyury protein expression in the notochord at E9.5 and posterior defects at E12.5.

(**A**) Overview of wildtype mouse *Brachyury/T/TBXTB* locus adapted from the UCSC genome browser and deletion alleles generated with CRISPR-Cas9 genome editing. Exact coordinates and sequences of target sites, deletions, and genotyping primer sequences can be found in **Supplemental Table 5**.

(**B-F**) Brachyury/T antibody staining (red) of E9.5 embryos. White dashed square in panels represents location of right

bottom inserts with 2x magnification. Brachyury/T protein expression in the notochord is dose-dependent on the three enhancer elements. Scale bar in **B**: 1 mm, applies to panels **B-F**.

(**G-K**) Overall morphology of E12.5 embryos with different genotypes. Blue lines indicate location of immunofluorescence and H&E sections. Spina bifida and tail defects are dose-dependent. Arrowheads marks small tails. White lines mark spina bifida. Scale bar in **G**: 1 mm, applies to panels **G-K**.

(**L-P**) Dorsal view of embryos (sectioned at blue line in **G-K**). White lines mark areas of spina bifida. Arrowheads mark small tails compared to tails in wildtype control and double knock-out allele. Scale bar in L: 2.5 mm, applies to panels **L-P**.

(**Q-U**) Immunofluorescence of mouse transverse sections. Anti-Sox2 labels the neural plate, anti-Tbxt the notochord, and DAPI marks nuclei. Sox2 expression is comparable amongst all genotypes, even in the genotypes with spina bifida, while there is loss of Brachyury/T staining in the notochord with increased loss of the enhancers. Arrowheads point to notochord. Asterisks mark absent notochord. Scale bar in **Q**: 0.2 mm, applies to panels **Q-U**.

(**V-Z**) H&E staining of transverse sections confirm the dose-dependent loss of the notochord and spina bifida. Arrowheads point to notochord. Asterisks mark absent notochord. Scale bar in V: 0.2 mm, applies to panels **V-Z**.

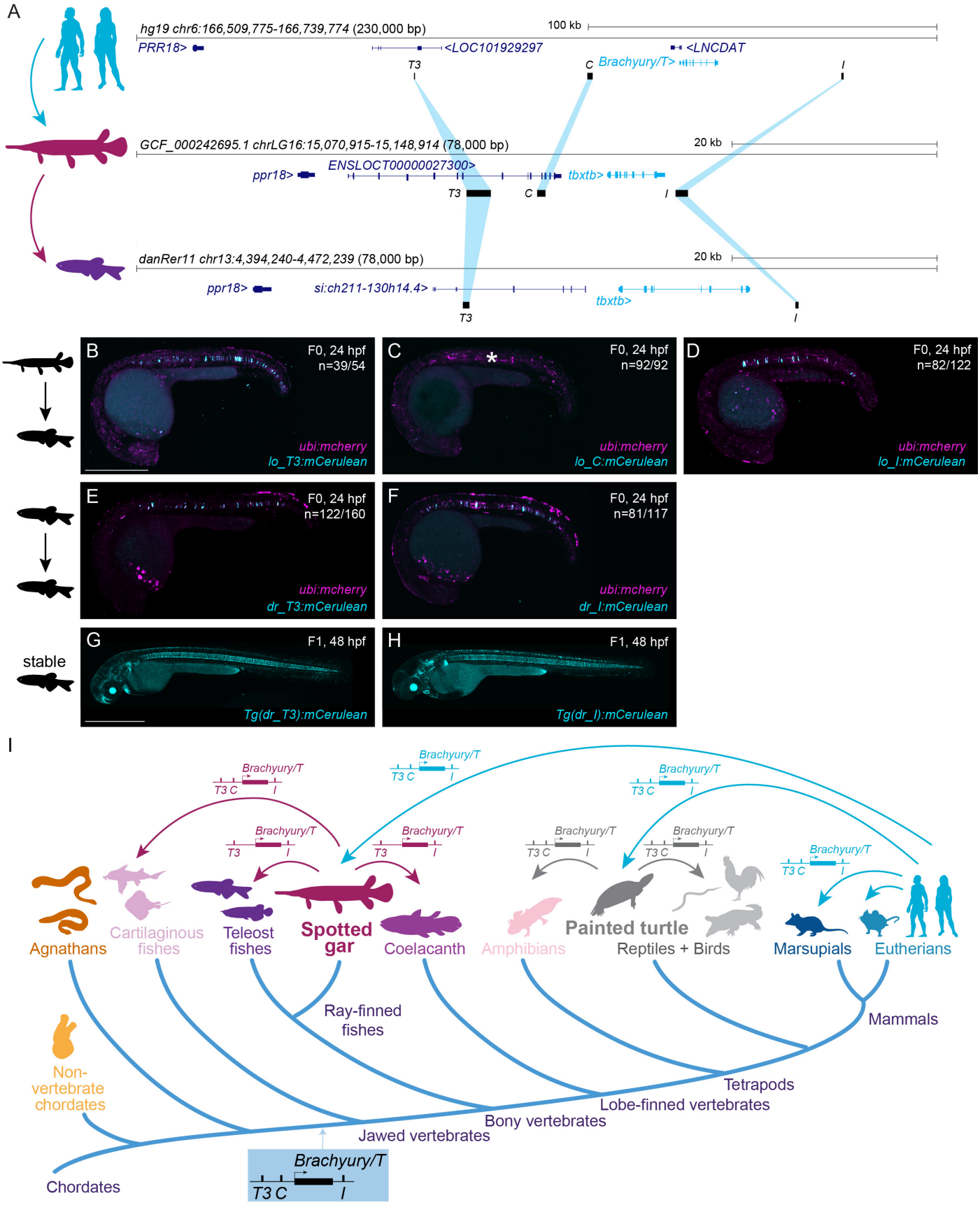
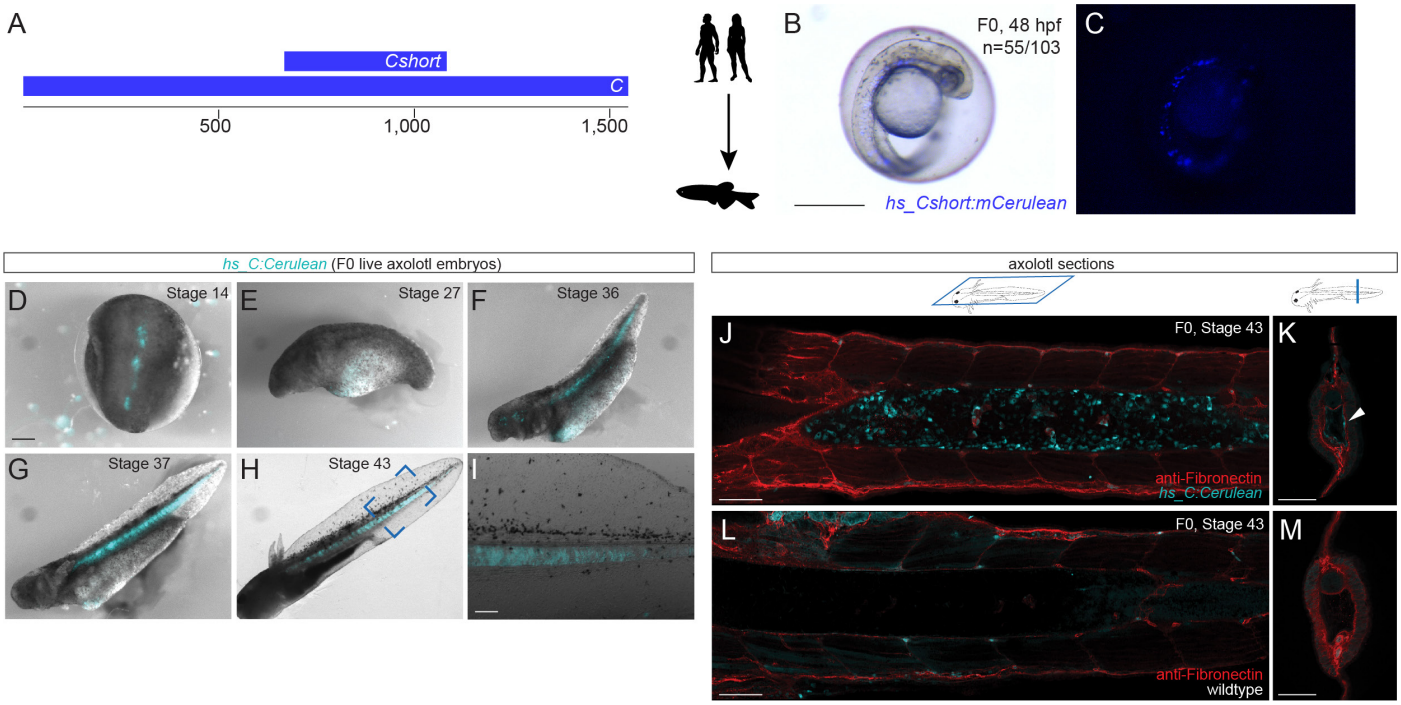


Figure 6: Bridge species establish the presence of *Tbxtb* enhancers across jawed vertebrates.
(**A**) Location of the enhancer elements in the human (top), gar (middle), and zebrafish (bottom) *Brachyury/T/Tbxtb* loci, adapted from the UCSC browser as established through the "gar bridge".
(**B,C,D**) Representative F0 zebrafish embryos injected with the gar enhancer elements *Io_T3* (**B**), *Io_C* (**C**), and *Io_I* (**D**). *T3* and *I* show mosaic *mCerulean* reporter expression in the notochord at 24 hpf compared to gar element *C* with is not active in the zebrafish notochord (asterisk). N represent the number of animals expressing mCerulean in the notochord relative to the total number of animals expressing mosaic *ubi:mCherry* as injection control. Scale bar in **B**: 0.5 mm, applies to **B-F**.

(**E,F**) Representative F0 zebrafish embryos injected with the conserved zebrafish enhancer elements *dr_T3* (**E**) and *dr_I* (**F**). *T3* and *I* show mosaic *mCerulean* reporter expression in the notochord at 24 hpf. N represent the number of animals expressing mCerulean in the notochord relative to the total number of animals expressing mosaic *ubi:mCherry* as injection control.
(**G,H**) Representative images of stable F1 embryos at 2 dpf of zebrafish enhancer elements *T3* and *I* recapitulate the F0 expression pattern in the notochord, with *dr_T3* (**E**) additionally expressing mCerulean in the brain, heart, and fin, and *dr_I* (**G**) in the proximal kidney close to the anal pore, pharyngeal arches, heart, fin, and spinal cord neurons. Scale bar in **G**: 0.5 mm, applies to **G**,**H**.
(**I**) Phylogenetic representation of species investigated using the bridging approach with spotted gar and painted turtle as anchor species within ray-finned fish and tetrapod lineages. Arrows indicate informative phylogenetic comparisons to uncover conservation of enhancer elements *T3*, *I*, and *C*.



Supplemental Figure 1: Human enhancer element *Cshort* in zebrafish and time course of human enhancer element *C* in axolotl.

(**A**) Schematic representation of human enhancer element *C* and *Cshort*.

(**B,C**) Representative F0 transgenic zebrafish embryo expressing *hs_Cshort:mCerulean* in the notochord at 2 dpf. Images shown are a merge of bright field and fluorescence (**B**) and fluorescence only (**C**). Scale bar in **B**: 0.5 mm, applies to **B**,**C**.

(**D-I**) Live images of representative F0 transgenic axolotl embryos expressing *hs_C:mCerulean* at stages 14 (**D**), 27 (**E**), 36 (**F**), 37 (**G**), 43 (**H**) and close up of H from the blue outline (**I**). Images shown are a merge of bright field and fluorescence. Scale bar in **D**: 1 mm; applies to panels **D-H**. Scale bar in **I**: 0.5 mm.

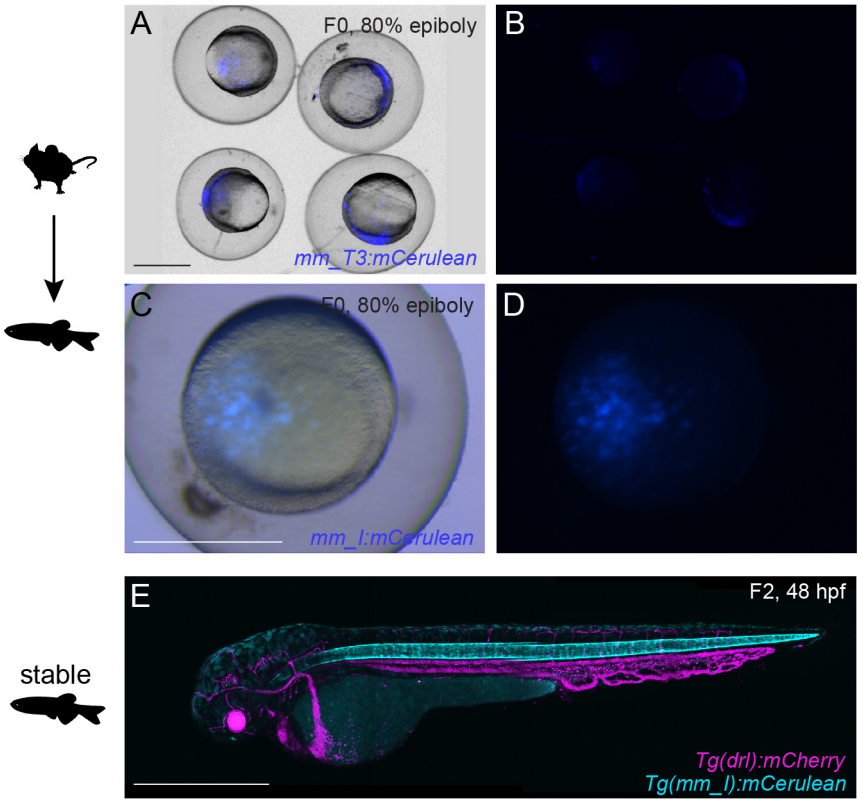
(**J-M**) Confocal images of horizontal (**J,L**) and cross (**K,M**) sections through the Axolotl embryo (stage 43) show mCerulean fluorescence in the notochord in transgenic *hs_C:mCerulean* embryos (**J,K**) compared to wildtype embryos (**L,M**), but not in the surrounding muscle which is highlighted by immunostaining of fibronectin in red. Scale bars in **J-M**: 0.5 mm.

A

B

Motif ID	Name	Input	Length	Start	Stop	Strand	Score	P value	Q value	Matched sequence
MA0009.2	TBXT	*hs_C*	415 bp	153	168	+	-20.2115	0.0341	1	TCATCCACATCTGCAA
MA0009.2	TBXT	*hs_C*	415 bp	153	168	-	-8.5	0.00532	0.995	TTGCAGATGTGGATGA
MA0009.2	TBXT	*hs_C*	415 bp	201	216	+	-9.55769	0.00648	0.995	TAGCACCAAACAGAGA
MA0009.2	TBXT	*hs_C*	415 bp	201	216	-	-10.6923	0.00793	0.995	TCTCTGTTTGGTGCTA
MA0009.2	TBXT	*hs_C*	415 bp	218	233	+	-10.0577	0.00709	0.995	AAGCAAACACAGGAGA
MA0009.2	TBXT	*hs_C*	415 bp	218	233	-	-21.1154	0.0385	1	TCTCCTGTGTTTGCTT

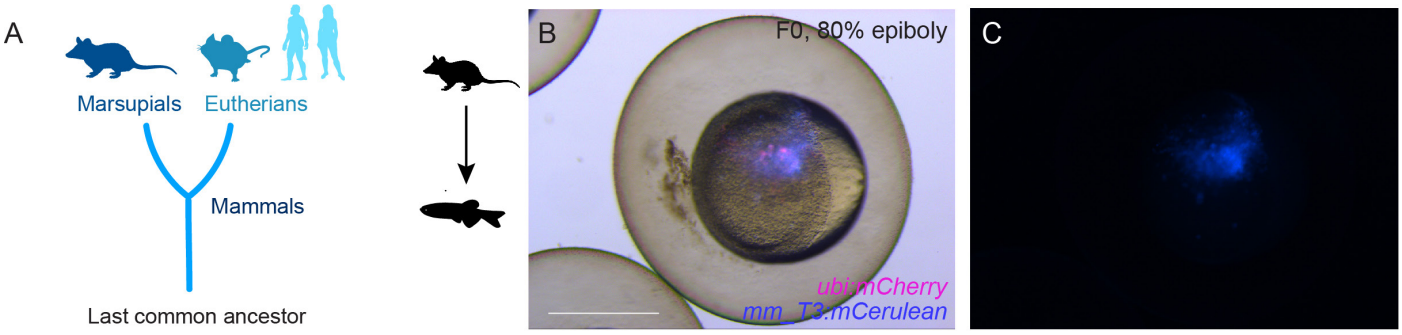
Supplemental Figure 2: Additional identified TBXT binding sites in enhancer *C*.
(**A**) Schematic depiction of human enhancer *C* with the TBXT binding site/T-box (pink box) and additional T-boxes (violet boxes).
(**B**) FIMO output with location of the T-boxes, statistical significance, and matched sequence within the enhancer elements.



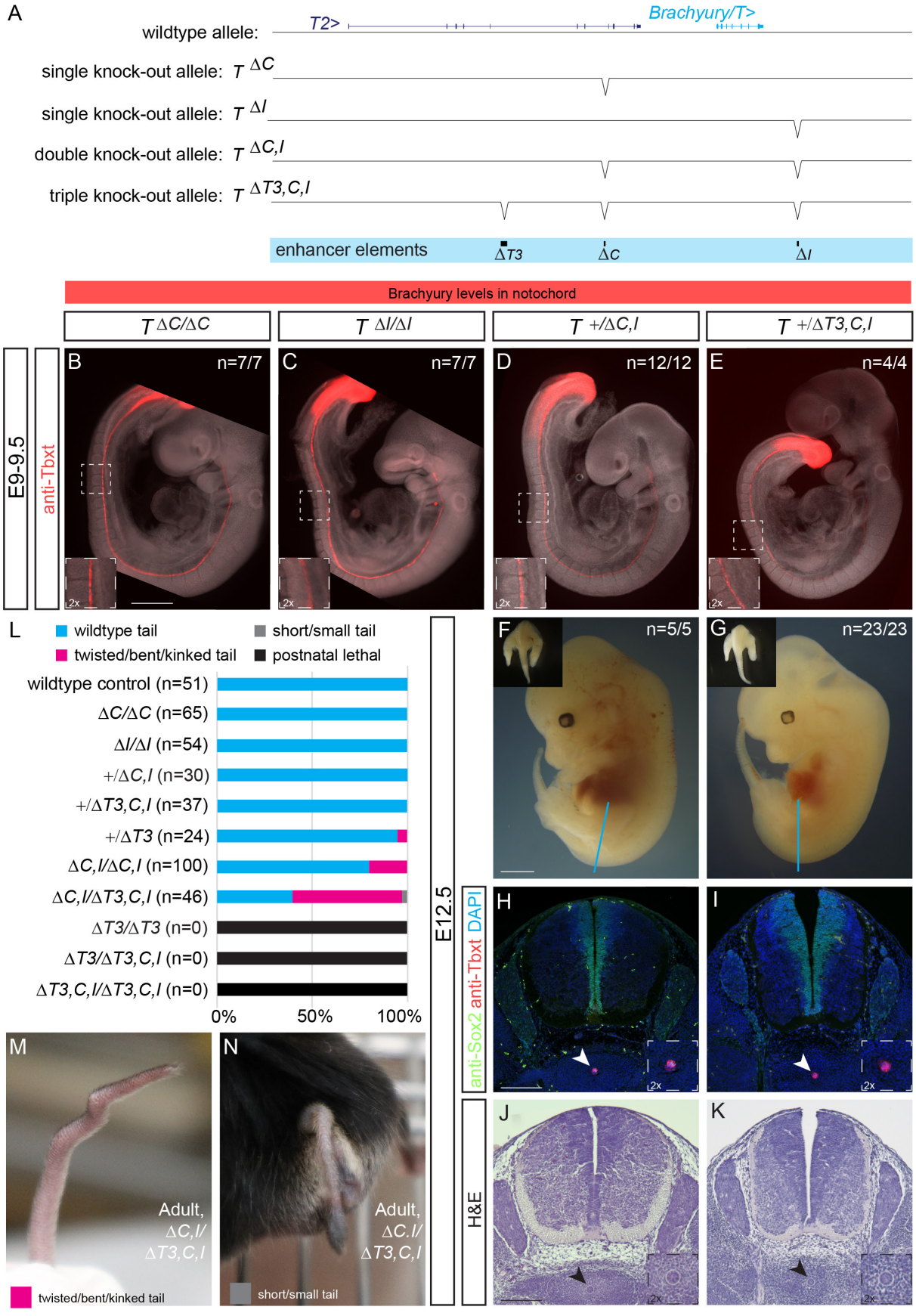
Supplemental Figure 3: Mouse enhancer elements *T3* and *I* at 80% epiboly in zebrafish and mouse enhancer element *I* stable zebrafish line.

(**A-D**) Mouse enhancer element T3 and I in zebrafish at 80% epiboly. Live images of representative F0 transgenic zebrafish embryos expressing *mm_T3:mCerulean* in the zebrafish embryo at 80% epiboly. Images shown are a merge of bright field and fluorescence (**A**) and fluorescence only (B). Further, live images of a representative F0 transgenic zebrafish embryo expressing *mm_I:mCerulean* in the zebrafish embryo at 80% epiboly. Images shown are a merge of bright field and fluorescence (**C**) and fluorescence only (**D**). Scale bars in **A**,**C**: 0.5 mm, applies to **A**,**B** and **C**,**D**.

(**E**) Representative image of a F2 embryo at 2 dpf from F1 stable line for mouse enhancer element *I* crossed to *Tg(drl:mCherry)* stable line labelling lateral plate mesoderm lineages. Transgenic F2 embryo recapitulates the F0 expression pattern in the notochord. Scale bar in **E**: 0.5 mm.



Supplemental Figure 4: Additional data to *Monodelphis domestica* enhancer elements.
(**A**) Mammalian phylogeny outlining the split into Marsupials and Eutherians.
(**B,C**) Monodelphis enhancer element *T3* in zebrafish at 80% epiboly. Live images of representative F0 transgenic zebrafish embryos expressing *md_T3:mCerulean* and *ubi:mCherry* in the zebrafish embryo at 80% epiboly. Images shown are a merge of bright field and fluorescence (**B**) and fluorescence only (**C**). Scale bar in **B**: 0.5 mm, applies to **C**.



Supplemental Figure 5: Additional data to deletion of the three enhancer elements.
(**A**) Mouse *Brachyury/T/TBXTB* locus adapted from UCSC browser and annotation of single knockout alleles *ΔC* and *ΔI*.
(**B-E**) E9.5 homozygous *ΔC* (**B**), homozygous *ΔI* embryos (**C**), heterozygous *+/ΔC,I* (**D**), and heterozygous *+/ΔT3,C,I* embryos (**E**) display normal Brachyury/T protein expression (red) in the notochord as depicted by anti-T immunofluorescence. White dashed square in panels represents location of right bottom inserts with 2x magnification. Scale bar in **B**: 1 mm, applies to panels **B-E**.
(**F,G**) Overall morphology of E12.5 embryos with different genotypes. Inserts in the left upper corner represent anterior

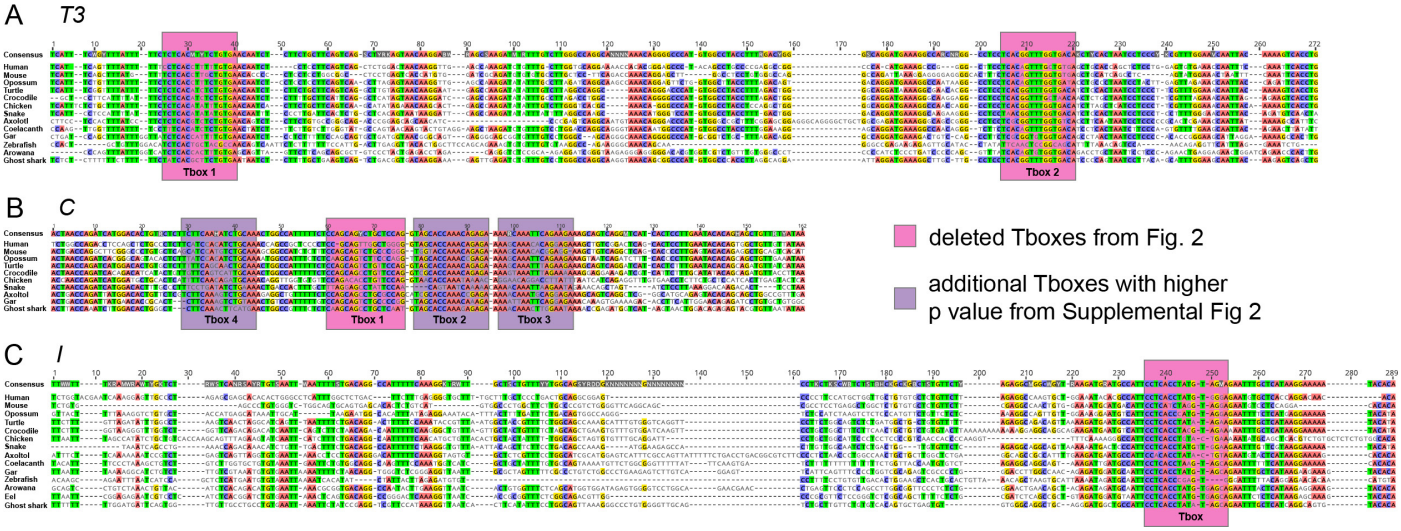
view of the trunk and tails. Blue lines indicate location of immunofluorescence and H&E sections. Inserts in the top left indicate wildtype looking tails. Scale bar in **F**: 1 mm, applies to panels **F**,**G**.

(**H,I**) Immunofluorescence of mouse transverse sections. Anti-Sox2 labels the neural plate, anti-Tbxt the notochord, and DAPI marks nuclei. Sox2 and Brachyury/T expression is comparable amongst the shown genotypes. Scale bar in **H**: 0.2 mm, applies to **H**,**I**.

(**J,K**) H&E staining of transverse sections confirm normal notochords. Arrowheads point to notochord. Scale bar in **J**: 0.2 mm, applies to **J**,**K**.

(**L**) Percentage of adult animals with tail phenotypes.

(**M,N**) Representative images of the kinked and short/small tail phenotype *in* $T^{\Delta C,I/\Delta T3,C,I}$ *trans*-heterozygous adult animals.



Supplemental Figure 6: Conservation of the enhancers *T3*, *C*, and *I* with their respective T-boxes.
(**A-C**) Conservation of the enhancers *T3* (**A**), *C* (**B**), and *I* (**C**) and respective T-boxes in different species. T-boxes with low p value are marked in pink (see **Fig. 2**) and T-boxes with higher p-value are marked in violet (see Supplemental **Fig. 2**).

Supplemental Data

Supplemental Table 1: Genomic features of the human enhancer elements.
Summary table listing the genomic features of the human enhancer elements, including length, location relative to transcription start (TS) site, ATAC- or T ChIP-sequencing peaks, and conservation in mouse and Monodelphis.

Supplemental Table 2: Reporter activity across animal models.
All numbers from the enhancer reporter experiments in zebrafish, Axolotl, mouse, and *Ciona*.

Supplemental Table 3: Coordinates of all cloned enhancer elements.
Summary table displaying the genomic coordinates of all enhancer elements from different species, as well as primer sequences used to amplify them, length, and reporter activity of the enhancers in the different species.

Supplemental Table 4: *Tbxtb* enhancer element conservation across vertebrates.
Genomic location and genome versions are provided for each species. BLAST bridging chain is indicated with -> showing BLAST hits from *Tbxtb* loci of one species to another and -x indicating lack chaining. (2x) indicate tetraploid species with up to two *tbxtb* loci.

Supplemental Table 5: Enhancer element deletions and primer sequences for genotyping.
Summary table with genomic coordinates and sequences of the used target sites, primer, and sequences of the three enhancer deletions.

Supplemental Table 6: Qualitative evaluation of Brachyury antibody staining in E9.5 embryos.
Summary table of qualitative evaluation of anti-Brachyury/T staining in E9.5 embryos.

Supplemental Data File to Fig. 6 and Supplemental Fig. 6: Sequence and alignment files of *T3*, *C*, and *I*.