

Functional and evolutionary insights from the *Ciona* notochord transcriptome

Wendy M. Reeves, Yuye Wu, Matthew J. Harder and Michael T. Veeman*

ABSTRACT

The notochord of the ascidian *Ciona* consists of only 40 cells, and is a longstanding model for studying organogenesis in a small, simple embryo. Here, we perform RNAseq on flow-sorted notochord cells from multiple stages to define a comprehensive *Ciona* notochord transcriptome. We identify 1364 genes with enriched expression and extensively validate the results by *in situ* hybridization. These genes are highly enriched for Gene Ontology terms related to the extracellular matrix, cell adhesion and cytoskeleton. Orthologs of 112 of the *Ciona* notochord genes have known notochord expression in vertebrates, more than twice as many as predicted by chance alone. This set of putative effector genes with notochord expression conserved from tunicates to vertebrates will be invaluable for testing hypotheses about notochord evolution. The full set of *Ciona* notochord genes provides a foundation for systems-level studies of notochord gene regulation and morphogenesis. We find only modest overlap between this set of notochord-enriched transcripts and the genes upregulated by ectopic expression of the key notochord transcription factor Brachyury, indicating that Brachyury is not a notochord master regulator gene as strictly defined.

KEY WORDS: *Ciona*, Transcriptome, Notochord

INTRODUCTION

The notochord is the eponymous feature of the chordates and is an essential organ in chordate development. Despite its importance in the innovation of the chordate lineage, the evolutionary origins of the notochord are unclear. The details of notochord morphology and morphogenesis vary somewhat between the cephalochordates, tunicates and vertebrates, but in all chordates the notochord forms from dorsal midline mesoderm cells that express the transcription factors Brachyury and FoxA, and undergo striking convergence and extension movements (Flood et al., 1969; Corbo et al., 1997a; Holland et al., 2004; Friedman and Kaestner, 2006; Satoh et al., 2012). These movements transform a relatively isodiametric primordium into a characteristic long thin rod that can have both structural and/or signaling roles depending on the species (Stemple, 2005; Corallo et al., 2015). This rod becomes encapsulated in a perinotochordal basement membrane (PBM) as it forms, and the notochord cells typically form vacuoles or extracellular lumen pockets that help to inflate the PBM sheath to act as a hydrostatic skeleton. This

structural role is particularly evident in species with swimming larval stages. Birds and mammals have smaller less-prominent notochords, but the notochord serves as a source of important midline patterning cues that control the dorsoventral patterning of the neural tube and the mediolateral patterning of the somite in all vertebrate embryos.

Notochord morphogenesis has been most studied in *Xenopus*, zebrafish and the ascidian *Ciona*. Non-canonical Wnt/PCP signaling has been implicated in mediolateral intercalation or related cell behaviors associated with convergent extension in all three species (Heisenberg et al., 2000; Wallingford et al., 2000; Jiang et al., 2005; Skoglund and Keller, 2010). Laminin, fibronectin and other extracellular matrix (ECM) components are known to be required for proper notochord morphogenesis in diverse chordates (Skoglund and Keller, 2007; Gansner et al., 2008; Veeman et al., 2008; Segade et al., 2016).

Ciona provides a particularly compelling model for studying notochord morphogenesis (Satoh et al., 2003; Passamanek and Di Gregorio, 2005). The *Ciona* notochord consists of only 40 cells that develop through fixed and relatively well-understood cell lineages. The *Ciona intestinalis* genome is extremely small at ~160 Mb and it diverged before the major genome duplications in the vertebrates (Dehal et al., 2002; Satoh, 2013). Modern molecular phylogenies show that it is the tunicates, including *Ciona*, and not the cephalochordates that are the closest outgroup of the vertebrates (Delsuc et al., 2006; Vienne and Pontarotti, 2006). *Ciona* was the second animal genome to be sequenced, and there are excellent community resources for genomics (Tassy et al., 2010; Brozovic et al., 2016).

Tissue-specific gene expression in the *Ciona* notochord has been of long-standing interest in the context of understanding cis-regulatory transcriptional control (Corbo et al., 1997b; Hudson and Yasuo, 2006; Hotta et al., 2008; Oda-Ishii et al., 2010; Katikala et al., 2013; Di Gregorio, 2017), the mechanisms of morphogenesis (Munro and Odell, 2002; Jiang et al., 2005; Hotta et al., 2007b; Veeman et al., 2008; Deng et al., 2013) and the enigmatic evolutionary origins of this defining innovation of the chordates (Satoh et al., 2012). Several previous studies have led to the systematic identification of *Ciona* notochord genes. Many of these came from a groundbreaking effort that used subtractive hybridization to identify transcripts upregulated in response to misexpression of the key notochord transcription factor Brachyury (Hotta et al., 1999; Takahashi et al., 1999). Others came from high-throughput *in situ* hybridization projects and microarray experiments (Satou et al., 2001b; Imai et al., 2004; Kugler et al., 2008; José-Edwards et al., 2011, 2013). Together, these projects identified 106 genes (summarized in Table S1) with either notochord-specific or notochord-enriched expression during intercalation and elongation. These genes have been a valuable resource for diverse functional studies, but we hypothesized that this was a far from complete list.

Division of Biology, Kansas State University, Manhattan, KS 66506, USA.

*Author for correspondence (veeman@ksu.edu)

 M.T.V., 0000-0002-0752-6537

Received 14 June 2017; Accepted 1 August 2017

RESULTS

RNAseq identifies 1364 putative notochord genes

FACS sorting of fluorescently labeled cells followed by microarray analysis was originally used to identify genes expressed in the border cells of *Drosophila* ovaries (Borghese et al., 2006) and has since become a popular strategy for tissue-specific transcriptional profiling. FACS/microarray approaches have been used in *Ciona* to identify genes expressed in heart precursors (Christiaen et al., 2008) and in pigment cells (Racioppi et al., 2014), and also to identify seven notochord-expressed transcription factors (José-Edwards et al., 2011). Here, we take advantage of the greater sensitivity and dynamic range of RNAseq (Malone and Oliver, 2011), combined with flow-sorted notochord-enriched and notochord-depleted cell populations from dissociated embryos, to gain a comprehensive and unbiased view of transcript-level notochord gene expression across the entire genome. *Ciona* notochord cells were fluorescently labeled by electroporation of fertilized eggs with the reporter construct *Bra*>GFP, which strongly expresses GFP in the notochord, although there can also be weak expression in the mesenchyme. Three timepoints were used: early in intercalation (late neurula, Hotta stage 16) (Hotta et al., 2007a); late in intercalation (early tailbud, Hotta Stage 19.5); and well after the end of intercalation (late tailbud, Hotta stage 23) (Fig. 1A–C). At each time point, embryos were dissociated to single cells in $\text{Ca}^{2+}/\text{Mg}^{2+}$ -free artificial seawater with trypsin, and then flow sorted to isolate populations of notochord-enriched GFP^{+} cells and notochord-depleted GFP^{-} cells (with the caveat that the notochord depletion is likely to be modest because of mosaic

transgene expression). Triple biological replicates of purified RNA were analyzed by RNAseq, and both normalized read counts and differential expression were calculated using the Cufflinks software suite (Trapnell et al., 2012). This provides quantitative, genome-wide information on normalized transcript abundance in sorted notochord cells and on the fold enrichment in notochord compared with the rest of the embryo. Our analysis here focuses on the genes enriched in the notochord, but the data can also be mined to identify genes that are expressed in notochord without necessarily being enriched in notochord. The stage 23 timepoint later proved to be partially contaminated with muscle-specific transcripts (described below), so we have removed it from most analyses.

We identified 891 genes with higher expression levels in the notochord-enriched population compared with the notochord-depleted cell population at stage 16 (Fig. 1D) and 933 upregulated genes at stage 19.5 (Fig. 1E). When data from these two stages were combined, a total of 1364 notochord-enriched genes were identified; 460 of which were enriched at both stages (Fig. 1F and Table S2). Notochord-enriched genes spanned a wide range of expression levels, with normalized read count (fragments per kilobase mapped, FPKM) values ranging from 1 to 4515. They represent 11.8% of the 11557 genes expressed in the notochord-enriched sample at one or both time points above a threshold of 1 FPKM. Our sequencing depth and biological replication were sufficient to detect statistical significance (corrected for multiple comparison) for genes that were only enriched in notochord by 1.47-fold. As seen in Fig. 1F,G, a smooth distribution of fold enrichment values was seen between these modestly enriched genes

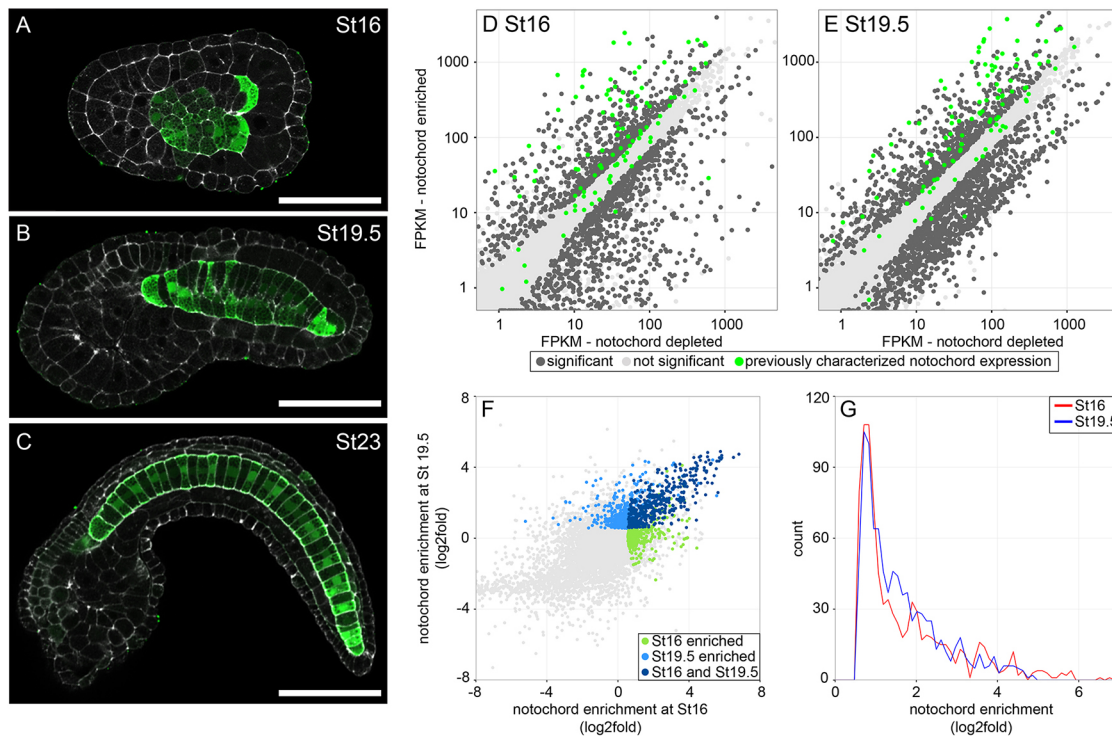


Fig. 1. Notochord-enriched genes identified by RNAseq. (A–C) *Ciona* embryos expressing a notochord-specific *Bra*>GFP reporter plasmid were harvested for dissociation, flow sorting and RNAseq at: (A) late neurula, stage 16; (B) early tailbud, stage 19.5; and (C) late tailbud, stage 23. Scale bars: 50 μm . (D,E) Gene expression in notochord-depleted and notochord-enriched cell populations at stage 16 and at stage 19.5. Dark gray indicates significant difference in expression between populations ($q < 0.05$); light gray indicates no significant difference; green indicates genes identified as being notochord expressed in previous screens as described in the text. (F) Comparison of notochord enrichment at stage 16 versus stage 19.5. Light green indicates genes enriched only at stage 16; light blue indicates genes enriched only at stage 19.5; dark blue indicates genes enriched at both stages; light gray indicates genes that are not enriched in notochord at either stage. (G) Distribution of fold enrichment values for statistically significant notochord-enriched genes at stage 16 (red) and stage 19.5 (blue).

and the most enriched genes. By comparison, the 20 most enriched genes at stage 16 and at stage 19.5 showed mean enrichments of 49- and 23-fold, respectively. This group includes the key notochord transcription factors *Brachyury* and *FoxAa*, as well as previously identified *Ciona* notochord genes *Noto4*, *Noto8*, *FGL2*, *TGFbeta2/3*, *SLC26Aa* and *THBS3/4* (thrombospondin) (Hotta et al., 1999, 2000; Imai et al., 2004; Katikala et al., 2013). The top hits also show considerable overlap, with half of the ten most enriched genes at stage 16 also in the top ten genes for stage 19.5 (Table 1).

As an initial validation, we examined the RNAseq expression data for the 106 genes previously identified as having notochord-specific or notochord-enriched expression (highlighted in green in Fig. 1D,E). Our genome-wide assay for notochord-enriched expression identified 81 of these 106 known notochord genes. Of the 23 that we missed, previously published *in situ* results showed that 16 were qualitatively expressed more strongly in another tissue compared with notochord (an expression pattern that we would expect to miss in our differential expression analysis) and five have conflicting published *in situ* hybridization data. This provided an initial confirmation that our RNAseq approach was likely to have few false negatives for genes with highly specific notochord expression.

Compared with previous screens, our RNAseq approach identifies a much greater range of expression levels, with less bias for highly expressed genes. At stage 19.5, for example, the mean notochord expression of the genes identified as notochord enriched by RNAseq was 130 FPKM, whereas it was 504 FPKM for the 106 previously characterized notochord genes and 697 FPKM for the 40 of those identified in the original subtractive hybridization screen (Takahashi et al., 1999). Many of the newly identified putative notochord genes were nevertheless highly expressed, with 14 of the 20 most highly expressed genes at stage 19.5 not found by the previous screens.

Validation by *in situ* hybridization

To further validate our RNAseq data, we selected 111 genes enriched in notochord at stage 16 and/or stage 19.5 and tested them by *in situ* hybridization at stages 16, 19.5 and 23. To better understand the data, we selected genes that spanned a broad range of notochord expression levels (from 3.5 to 4515 FPKM) and

enrichment values (from 1.5- to 90.9-fold) (Fig. S1). We included genes enriched at both single stages and at multiple stages. Predicted gene functions were not a criteria in selecting genes for *in situ* hybridization, with the exception of the 16 transcription factors discussed later. Of the 98 that were detectable by *in situ* hybridization, 88 (90%, Agresti-Coull 95% confidence interval: 82-95%) were expressed in notochord. These were sometimes expressed in other tissues as well, but notochord staining was distinct and usually predominant at the predicted stages. A representative sample of these genes is shown in Fig. 2A, and all results are summarized in Table S2. We were able to visualize notochord-specific expression of some genes, such as *ARF1/3/6* (5 FPKM and 8.7-fold enriched at stage 19.5, Fig. 2A), that were expressed at very low levels but highly enriched in the notochord. Genes with more modest notochord enrichment scores (enriched twofold or less) needed to have higher notochord expression levels (generally greater than 100 FPKM) to be detectable as enriched. *PTK2* is shown as an example in Fig. 2A. Both the timing of expression and the relative enrichment by *in situ* hybridization were well matched to the RNAseq data. Eight of the remaining ten genes were expressed strongly in mesenchyme cells, and two in muscle cells (examples in Fig. 2B,C). This modest contamination with mesenchymally expressed genes likely reflects the low levels of GFP expression induced in this tissue by the *Bra>GFP* construct. The stage 23 RNAseq data largely matched the *in situ* results for genes that first became enriched in notochord at stage 16 or stage 19.5.

We also tested 41 (36 detectable) genes that did not become enriched in notochord by RNAseq until stage 23, and found that these were largely expressed in mesenchyme and/or muscle (Fig. 2D,E). Only 11% (four genes) were notochord expressed. It is not clear whether this reflects failure to fully disassociate older embryos, muscle cell autofluorescence or suboptimal gating choices while FACS sorting. We therefore excluded the stage 23 RNAseq data from our subsequent analyses.

For stages 16 and 19.5, however, our 90% *in situ* validation rate indicates that the RNAseq data at these stages has a very low fraction of false positives. By comparison, the previous effort to systematically identify notochord genes by ectopic expression of *Brachyury* coupled with subtractive hybridization had an *in situ* validation rate of less than 10% and only identified 40 genes in total

Table 1. Highly enriched notochord genes

KH2012 gene model	<i>Ciona</i> gene name	Description	Stage 16			Stage 19.5		
			FPKM notochord enriched	FPKM notochord depleted	log ₂ fold-change	FPKM notochord enriched	FPKM notochord depleted	log ₂ fold-change
KH.S1404.1*	Brachyury	Transcription factor	674.8	6.0	6.81 ^{††}	565.6	21.2	4.74 ^{§§}
KH.C10.317 [‡]	FN1	Fibronectin	752.7	13.1	5.84 ^{††}	1575.1	68.0	4.53 ^{§§}
KH.L18.30 [§]	Noto4		781.9	13.6	5.84 ^{††}	2012.0	79.2	4.67 ^{§§}
KH.C12.115	SLC23A	Solute carrier	378.9	6.8	5.81 ^{††}	424.1	19.2	4.46
KH.C11.737 [§]	Noto8	Calcium-binding domain	1818.9	33.3	5.77 ^{††}	2798.9	94.4	4.89 ^{§§}
KH.C1.832 [¶]	FGL2	Fibrinogen	2452.4	46.7	5.72 ^{††}	3726.2	160.0	4.54 ^{§§}
KH.S396.5		No homology/domains	9.2	0.2	5.63 ^{††}	20.3	3.4	2.57
KH.C2.872	GPA1	G protein α	208.9	4.4	5.57 ^{††}	252.9	12.4	4.35
KH.C4.685	ITLN1/2	Intelectin/fibrinogen	33.7	0.7	5.53 ^{††}	53.6	3.5	3.92
KH.C6.164 ^{**}	THBS3/4	Thrombospondin	695.5	16.3	5.42 ^{††}	1369.7	119.0	3.52
KH.C10.529	CHDH	Choline dehydrogenase	15.2	0.5	5.05	21.6	0.8	4.83 ^{§§}
KH.C8.749 [‡]		No homology/domains	345.3	10.5	5.04	1878.0	77.9	4.59 ^{§§}
KH.C12.336	ODC1	Ornithine decarboxylase	51.8	2.1	4.66	125.9	5.1	4.63 ^{§§}
KH.C12.662		Has LDLR repeat	21.6	0.9	4.58	106.9	4.5	4.57 ^{§§}
KH.S391.1		No homology/domains	103.1	5.3	4.27	66.5	2.8	4.55 ^{§§}

*Carbo et al. (1997b). [‡]José-Edwards et al. (2013). [§]Hotta et al. (1999). [¶]Hotta et al. (2000). ^{**}Katikala et al. (2013). ^{††}Top ten enrichment at stage 16. ^{§§}Top ten enrichment at stage 19.5.

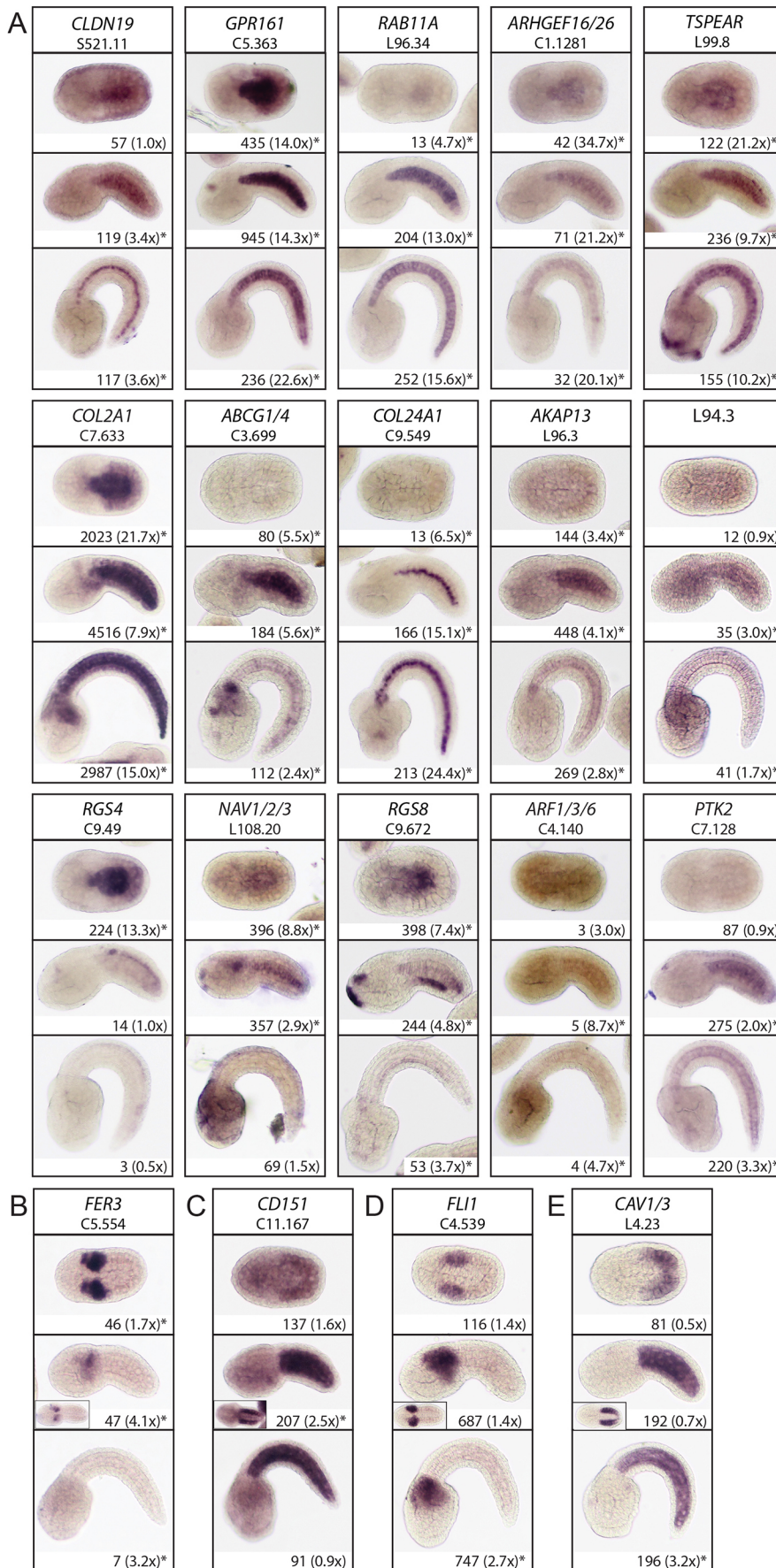


Fig. 2. *In situ* validation of RNAseq results.

Expression patterns of the indicated genes in embryos fixed at stage 16 (top), stage 19.5 (middle) and stage 23 (bottom). Expression levels (FPKM) in the notochord-enriched cell population and the fold enrichment compared with notochord-depleted cells are indicated. Statistically significant fold enrichments are labeled with an asterisk. For all images, anterior is towards the left. For the stage 16 embryos, dorsal is towards the viewer. For the stage 19.5 and 23 embryos, dorsal is towards the top of the page. Representative sample of genes identified at stage 16 and/or stage 19.5 with (A) confirmed notochord-enriched expression, (B) mesenchyme expression or (C) muscle expression. (D,E) Example of genes identified only at stage 23 showing mesenchyme (D) or muscle (E) expression. For B-E, insets show dorsal views.

(Hotta et al., 1999; Takahashi et al., 1999). Microarray screening for targets of the notochord-expressed transcription factor TBX2/3 had a moderately better validation rate of 29%, with 20/70 detectable genes confirmed as notochord expressed (José-Edwards et al., 2013). Our extensive validation predicts that about 90% of the 1364 notochord genes we have identified by RNAseq are likely to be bona fide. Our results support the tissue-specific transgene/FACS/RNAseq (FACSseq) approach for the systematic and relatively unbiased quantification of tissue-specific gene expression.

Massive enrichment of ECM-related GO codes

We used topGo (Alexa et al., 2006) to identify Gene Ontology (GO) terms enriched in our notochord-expressed gene sets. Gene model to GO code mappings were downloaded from the ANISEED ascidian community database (Brozovic et al., 2016) and are all inferred from sequence conservation and/or orthology with vertebrates. We ran the analyses for the group of genes that are notochord enriched at both stage 16 and stage 19.5 (460 genes), as well as for the 891 and 933 genes that are enriched at stage 16 or at stage 19.5 (Table 2 and Table S3).

For the genes enriched at both stages, the GO code analysis indicated a massive enrichment of genes associated with the ECM and cell-ECM interactions. This includes both commonplace ECM components, such as laminins, fibronectins, nidogen, perlecan and collagens, as well as many more exotic ECM components, such as fibulin, matrilin, hemicentin, FREM1, tenascin, fibrillin, papilin and thrombospondin. Genes involved in ECM modification were also common, including prolyl hydroxylase, lysyl hydroxylase, peroxidases, heparanase and serpins. Components of the focal adhesion complex linking the cytoskeleton and ECM were also enriched, including integrinA3/6/7, tensin, actinin, talin, parvin and fermitin. These are all consistent with the crucial role of the perinotochordal and intranotochordal ECM in notochord morphogenesis, both during intercalation and after (Skoglund and Keller, 2007; Gansner et al., 2008; Veeman et al., 2008; Segade et al., 2016).

Many of the other GO terms enriched with respect to the genes enriched at stage 16 and 19.5 were initially perplexing (neuromuscular junction, perception of sound, axon guidance, PDGF binding) but closer inspection indicated that these all reflected the enrichment of ECM-related genes. Although significant expression of ECM components was predicted given the known importance of the perinotochordal basement membrane, our data show that the most defining signature of the *Ciona* notochord transcriptome is the expression of a diverse set of ECM structural components, modifying enzymes and ECM-cytoskeleton linkers. We have identified many ECM components not previously known to be notochord enriched that will be useful in future studies of notochord morphogenesis. This is particularly important given the potential for functional redundancy in the ECM.

The GO codes enriched with respect to all genes enriched in the stage 16 notochord also include several associated with transcription, RNA processing, translation and the COPII ER to Golgi transport complex. These potentially reflect a general upregulation of the basic machinery for producing and secreting proteins, likely in preparation for the secretion of the perinotochordal ECM. Genes whose products are associated with GTPase activator activity (including both regulators of heterotrimeric G proteins and activators of small GTPases) were also over-represented in the notochord at this stage, and provide exciting candidates for regulatory roles in early notochord morphogenesis.

In the set of genes enriched at notochord at stage 19.5, we again found over-represented GO terms related to the ECM and cell-ECM binding. We also found over-represented GO terms related to both cell-cell adhesion and the actin cytoskeleton. In particular, we note that virtually the entire ARP2/3 complex (seven out of nine components) is enriched in the *Ciona* notochord at stage 19.5 (Table 2). This complex plays a key role in nucleating the growth of branched actin filaments (Swaney and Li, 2016), and has not been previously implicated in notochord morphogenesis. Stage 19.5 is quite late in the process of mediolateral intercalation, so we speculate that the ARP2/3 complex is

Table 2. Cellular component GO terms associated with *Ciona* notochord genes are enriched for ECM and cytoskeleton components

GO ID	Terms	Observed	Expected	P value
Genes enriched at both stage 16 and stage 19.5				
GO:0005604	Basement membrane	36	6.14	8×10^{-11}
GO:0005605	Basal lamina	11	0.93	5.6×10^{-7}
GO:0005925	Focal adhesion	15	4.17	0.000016
GO:0031093	Platelet α granule lumen	8	1.27	0.000027
GO:0005615	Extracellular space	46	23.12	0.00013
GO:0005592	Collagen type XI trimer	3	0.15	0.00022
GO:0005606	Laminin 1 complex	4	0.35	0.00024
GO:0031594	Neuromuscular junction	11	3.24	0.00036
Genes enriched at stage 16				
GO:0022627	Cytosolic small ribosomal subunit	16	3.35	5.7×10^{-8}
GO:0005604	Basement membrane	38	11.38	0.000001
GO:0005605	Basal lamina	11	1.71	0.000019
GO:0043025	Neuronal cell body	39	21.27	0.00009
GO:0005925	Focal adhesion	20	8.34	0.00019
GO:0005667	Transcription factor complex	36	22.75	0.0002
GO:0030864	Cortical actin cytoskeleton	18	6.16	0.00024
GO:0030127	COPII vesicle coat	4	0.47	0.00048
Genes enriched at stage 19.5				
GO:0005604	Basement membrane	42	12.07	7.1×10^{-8}
GO:0005885	Arp2/3 protein complex	7	0.7	5.3×10^{-7}
GO:0005605	Basal lamina	12	1.87	0.000005
GO:0005925	Focal adhesion	21	8.33	0.000064
GO:0005886	Plasma membrane	286	221.29	0.00035
GO:0031594	Neuromuscular junction	16	6.39	0.0005

Observed, number of genes significantly enriched in notochord; expected, number of genes expected by chance to be enriched in notochord.

involved in the subsequent morphogenetic event where the notochord becomes longer via the cells slowly changing from disk-shaped to drum-shaped, which is known to be driven by actomyosin contractility in the cell cortex (Dong et al., 2011).

Genes of interest: signaling, channels and transcriptional regulators

In addition to our unbiased assessment of enriched GO terms, we also searched the list of 1364 notochord-enriched genes for predicted functions in cell signaling, channel/transporter activity and transcriptional regulation. Signaling genes included numerous components of the noncanonical Wnt/PCP pathway, which has been previously implicated in *Ciona* notochord morphogenesis (Jiang et al., 2005). Enriched noncanonical Wnt/PCP components included Prickle, Strabismus, Daam1/2 and ROR1/2. There were also several Wnt ligands and receptors that could potentially be involved in canonical or noncanonical Wnt signaling, including Wnt5b, Wnt9, Frizzled5/8 and Frizzled4/9/10. Surprisingly, APC and Casein kinase I ϵ were also enriched, both of which are thought to be specifically involved in canonical Wnt/ β -catenin signaling (Hart et al., 1998; Peters et al., 1999; Angers and Moon, 2009).

Channels and transporters are of great interest with respect to the inflation of the notochord lumen and other aspects of notochord cell behavior. Even though these did not prove to be enriched GO codes, we found that many channels and transporters were enriched in the notochord (see Table S2). These included several predicted potassium channels (KCNB1/2, KCNH3/4/8, KCNJ4/12/8, KCNN1/2/3 and KCNQ2/3/4/5), 25 different members of the Solute Carrier (SLC) superfamily with diverse predicted

specificities (Na⁺, K⁺, Cl⁻, Zn⁺², sulphate, anions and cations) and three TMEM family channels. The anion transporter SLC26A α was previously identified as having notochord-enriched expression as well as playing an important role in notochord lumen expansion (Hotta et al., 1999; Deng et al., 2013). Deng et al. predicted that additional solute transporters and channels would also function during notochord tubulogenesis, and our data provide a list of candidates for future studies.

The notochord-enriched gene set also contained over 100 genes with potential roles in transcriptional regulation, including transcriptional activators and transcriptional repressors, as well as general transcriptional regulators such as chromatin remodeling factors, histone modifying enzymes, Mediator complex components and elongation factors. The transcriptional activators included several previously identified as expressed in the *Ciona* notochord, such as Bra, FoxAa, TBX2/3, Sall-a, STAT5/6, NFAT5 and Klf15 (Corbo et al., 1997b; Imai et al., 2004; José-Edwards et al., 2011), as well as many potential new notochord TFs. Some of these can be excluded based on published mesenchyme expression patterns, but we have performed *in situ* hybridization on 16 transcription factors with ambiguous or no published expression patterns at the relevant stages and found that 11 have subtle but bona fide notochord expression (the rest were expressed in the mesenchyme) (Fig. 3 and Table S2). Additional *in situ* validation will be needed to define the full set of notochord-enriched candidate transcriptional regulators.

Differential promoter usage and splicing

In addition to expression enrichment, the Cufflinks suite can test for differences in promoter use and differential splicing (Trapnell et al.,

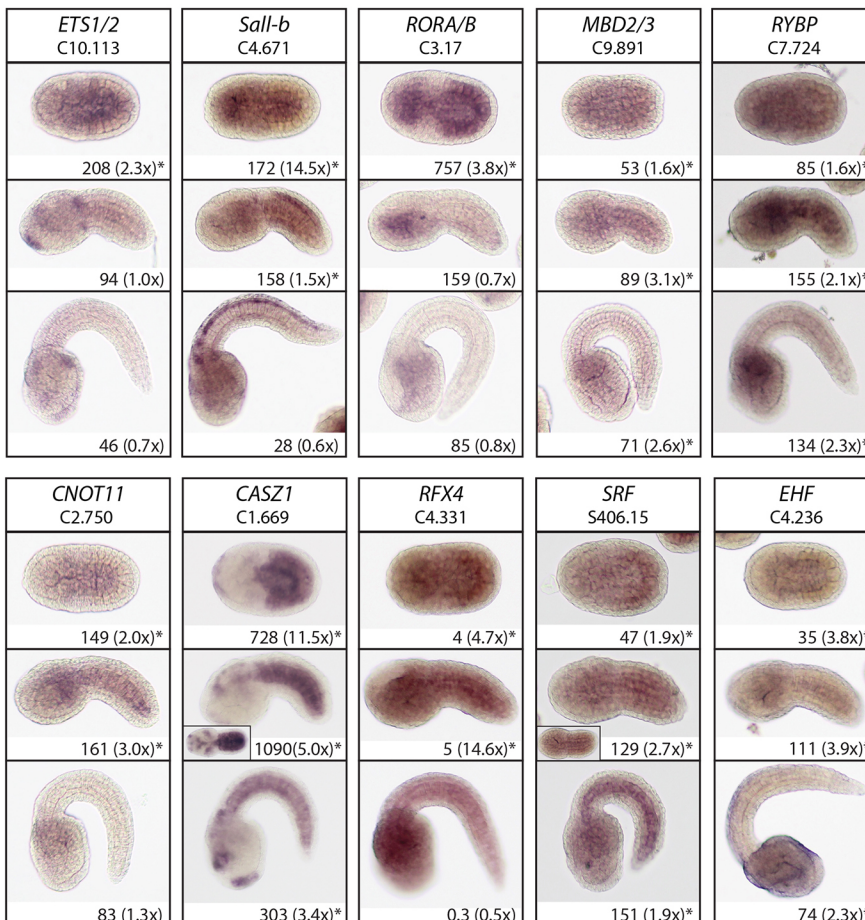


Fig. 3. Novel transcription factors with notochord-enriched gene expression. Expression patterns of the indicated genes in embryos fixed at stage 16 (top), stage 19.5 (middle) and stage 23 (bottom). Expression levels (FPKM) in the notochord-enriched cell population and the fold enrichment compared with notochord-depleted cells are indicated. Statistically significant fold enrichments are labeled with an asterisk. For all images, anterior is towards the left. For the stage 16 embryos, dorsal is towards the viewer. For the stage 19.5 and 23 embryos, dorsal is towards the top of the page. Insets at stage 19.5 are of dorsal views.

2013). At stage 16, 124 genes showed evidence of differential transcriptional start site usage in notochord-enriched versus -depleted cell pools (3.6% of the 3492 genes with more than one primary transcript annotated and expressed). Two-hundred and thirty-nine primary transcripts (3.15% of 7589 primary transcripts with sufficient alignments for testing) showed evidence of differential splicing between the two cell populations at this stage. Full results for both stages are shown in Tables S4 and S5. This is the first study to systematically estimate tissue-specific promoter use and alternative splicing in *Ciona*, so it is not clear at present if these rates are distinctive. One group previously estimated that about 15-20% of all genes in *Ciona* can be alternatively spliced, but did not look at individual tissues (Kim et al., 2007).

Only modest overlap found with genes induced by ectopic Brachyury expression

Transcription factor overexpression followed by transcriptional profiling provides an alternate strategy for identifying genes with tissue-specific expression. We identified many more notochord-enriched genes by FACSseq than were previously identified by Brachyury overexpression (Hotta et al., 1999; Takahashi et al., 1999), but it was unclear to what extent this reflected the more quantitative and genome-wide nature of RNAseq compared with subtractive hybridization. Alternatively, it might reflect a more fundamental difference between the genes induced by Brachyury misexpression and the normal notochord transcriptional profile. To

test this, we broadly misexpressed Brachyury under the control of the *FoxAa* enhancer in a similar fashion to Takahashi et al. (1999) (Fig. 4A,B), but then used RNAseq to identify upregulated genes as compared to GFP-expressing control embryos. Given that Brachyury is thought to be at the top of the notochord transcriptional cascade, we hypothesized that most of the 1364 notochord genes identified by FACSseq would be upregulated to some extent by ectopic Brachyury expression.

We identified 925 genes with increased expression in *FoxAa*>*Bra*-expressing embryos when compared with *Bra*>GFP-expressing controls (Table S6). There was unexpectedly little overlap, however, between these genes and the notochord FACSseq gene set, with only 264 found in both (Fig. 4E). Genome-wide there is a positive but weak correlation between notochord specificity (FACSseq fold-change) and upregulation in response to Brachyury misexpression (*FoxAa*>*Bra* fold-change) (Fig. 4C), but the majority of notochord-enriched genes show no statistically significant response to ectopic Brachyury expression (Fig. 4D). The sequencing depth and triple biological replicates used showed statistically significant differential gene expression for genes with a log2 fold-change as low as ~0.4, so this was a very sensitive test.

The *FoxAa* cis-regulatory region used drives expression in neural, endodermal and mesenchymal lineages in addition to notochord. It has previously been suggested that *FoxAa*>*Bra* expression transforms these other cells to notochord fate (Takahashi et al., 1999). Our RNAseq data, however, shows that only a subset of

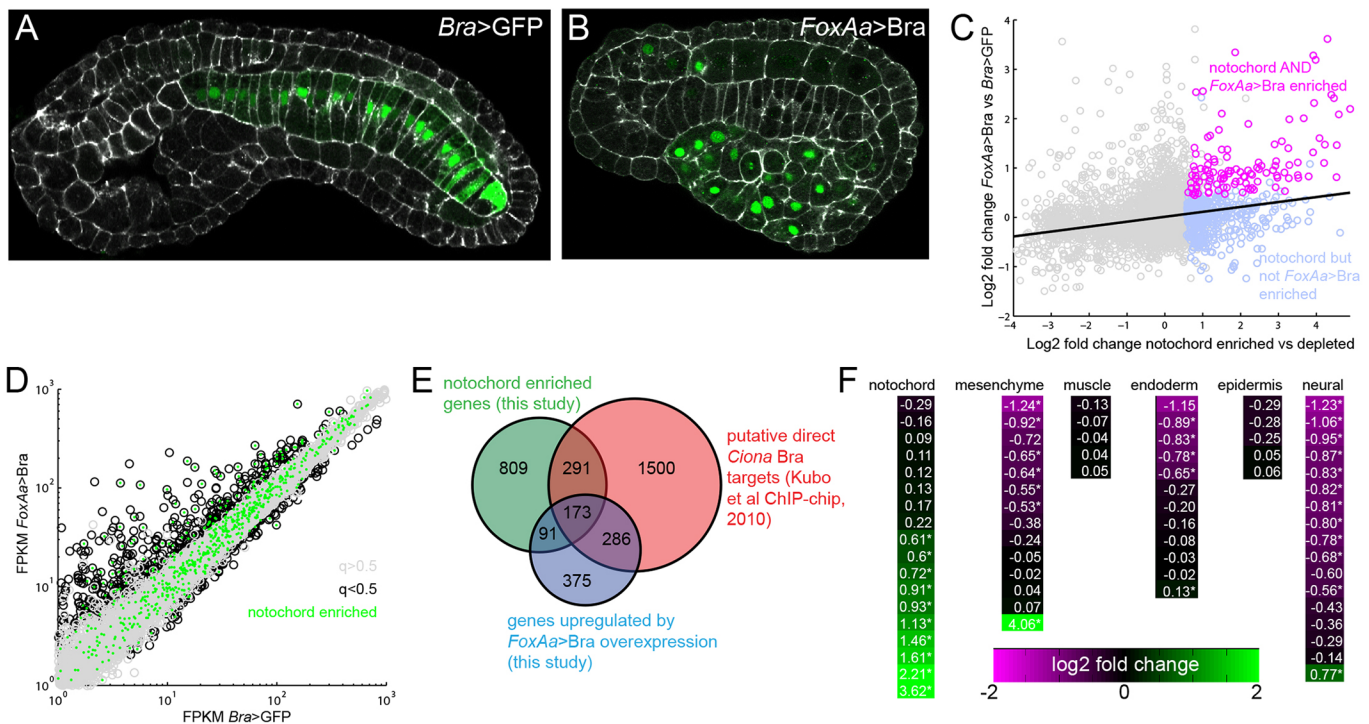


Fig. 4. Brachyury misexpression induces only a subset of notochord-enriched genes. (A) Control embryo expressing GFP under the control of the notochord-specific *Brachyury* enhancer. (B) Venus-tagged Brachyury expressed under the control of the *FoxAa* enhancer. *FoxAa*>*Bra*-expressing cells from the endoderm and other lineages form a disorganized mass. (C) Scatter plot showing the relationship between notochord enrichment by FACSseq and induction by ectopic Brachyury. Genes with statistical significance ($q < 0.05$) for being both enriched in the notochord and inducible by ectopic Brachyury are indicated in magenta. Genes with statistical significance for being enriched in the notochord but not for being induced by ectopic Brachyury are indicated in blue. (D) Scatter plot of normalized read counts (FPKM) for control *Bra*>GFP embryos compared with *FoxAa*>*Bra* embryos. Statistically significant ($q < 0.5$) differential expression is in black. Genes identified as notochord enriched by FACSseq are highlighted with green dots. (E) Venn diagram showing the overlap between genes identified as notochord enriched by FACSseq, genes upregulated by *FoxAa*>*Bra* expression and genes identified by Kubo et al. (2010) as Brachyury-bound by ChIP-chip. (F) Heat maps showing the fold change of expression of tissue-specific markers of notochord, mesenchyme, muscle, endoderm, epidermis and neural fate in *FoxAa*>*Bra*-expressing embryos compared with controls. The color map shows downregulated genes as magenta and upregulated genes as green. Statistically significant ($q < 0.5$) data points are marked with asterisks. Gene names are in Table S7.

notochord-enriched genes are upregulated. To better explore the transcriptional response, we defined panels of tissue-specific markers by screening the ANISEED gene expression database for genes with distinct tissue-specific expression in the main *Ciona* tissues at the relevant stages. We found that markers of muscle and epidermal fate were unaffected by Brachyury misexpression. Many markers of neural, endodermal and mesenchymal fate were downregulated, but others were unaffected and a few were paradoxically upregulated (Fig. 4F and Table S7). This again supports the hypothesis that Brachyury overexpression only partially reprograms other cells to notochord fate.

Of the 661 genes transcriptionally upregulated by Brachyury misexpression but not notochord enriched by FACSseq, many were either not expressed or only minimally expressed in control embryos (311 at FPKM<2, 428 at FPKM<5). We searched the ANISEED gene expression database and found that 18% of the 661 had previously reported *in situ* hybridization results at comparable stages. Of these 120 genes, only 14% had expression in notochord (typically in combination with some other tissue), 13% in mesenchyme, 8% in muscle, 13% in endoderm, 35% in neural cells and 26% in epidermis. Forty percent showed either no expression or diffuse/nonspecific expression. We performed *in situ* hybridization for 18 genes that had not been previously characterized and found that half of the genes had widespread expression (Fig. S2). Only three genes showed specific expression in the notochord, again in combination with other tissues. Most of the 18 genes had their strongest expression in muscle and/or mesenchyme. Together, these results show that only a small fraction of the genes that are upregulated by Brachyury overexpression but not notochord enriched by FACSseq are normally expressed in the notochord at all.

One possible explanation for why Brachyury misexpression only induces ~20% of notochord-enriched genes is that morphogenesis is severely perturbed in *FoxAa*>Bra embryos and the notochord GRN might perhaps be dependent on the normal progression of morphogenesis. To test this, we blocked morphogenesis entirely after notochord induction by treatment with cytochalasin B and then used RNAseq to look for differential gene expression as compared with control embryos. Genome-wide we found only 43 genes, showing significant downregulation at stage 19.5, and only four of these were notochord enriched by FACSseq (Table S8). Feedback from morphogenesis thus does not appear to be a major factor in the notochord GRN, at least not over the timeframe examined.

We also compared the notochord FACSseq and *FoxAa*>Bra RNAseq data with the Brachyury ChIP-chip data of Kubo et al., who used chromatin immunoprecipitation of a tagged Brachyury construct coupled with a sparse but genome-wide microarray to determine Brachyury occupancy at the early gastrula stage (soon after notochord induction) (Kubo et al., 2010). They identified 2252 genes with evidence of Brachyury occupancy, 464 of which are notochord enriched according to our FACSseq data. Two-hundred and eighty-six of the genes that are *FoxAa*>Bra induced but not notochord enriched are predicted to be Brachyury bound, but 375 are not (Fig. 4E and Table S9). This supports a complex model of notochord gene regulation in which many notochord genes are not direct Brachyury targets and most Brachyury-bound regulatory regions do not drive notochord-specific expression.

Notochord gene expression conserved across the Olfactores

There are both profound similarities and also distinct differences in notochord anatomy and morphogenesis between the cephalochordates, tunicates and vertebrates. To define a core set of genes with shared notochord expression, we compared our set of *Ciona* notochord-

enriched genes to the gene expression patterns cataloged in the mouse and zebrafish gene expression databases (Howe et al., 2013; Finger et al., 2017). We chose these species because they have the most comprehensive gene expression databases, but the mouse and zebrafish notochords are also quite different in their anatomy and development (Myers et al., 2002; Stemple, 2005; Yamanaka et al., 2007), and thus serve as diverse vertebrate comparison groups. For both species, we queried the databases for all genes expressed in the notochord during roughly comparable stages of notochord morphogenesis and then identified the *Ciona* orthologs (Brozovic et al., 2016).

We found many genes that are expressed in the notochord in both *Ciona* and in fish and/or mouse embryos. Of the 132 mouse notochord genes with orthologs in *Ciona*, 31 were present in our dataset (Table S10). This was significantly more overlap than the 14 genes that would be expected by chance ($P=2.1\times 10^{-5}$ by the hypergeometric distribution test). Similarly, the orthologs of 99/359 fish notochord genes were enriched in our sorted notochord cells (Table S11). This is again greater than the 39 genes expected by chance ($P=2.6\times 10^{-19}$). Overall, there were 12 genes identified as notochord expressed in all three species (Table 3).

Although we have identified more than twice as many genes with notochord expression shared between *Ciona* and vertebrates than would be predicted by chance alone, it is likely that the true overlap is considerably greater given the varying techniques used (RNAseq versus microarrays versus *in situ* hybridization; genome-wide studies versus gene-by-gene studies, etc.), the varying coverage in high-throughput *in situ* screens and the precise developmental stages examined. For example, many of the zebrafish notochord genes were identified in high-throughput *in situ* hybridization screens (Thisse et al., 2004), whereas the most common source of the mouse notochord genes was an *in situ* validated microarray experiment using flow-sorted *Noto*>eGFP-expressing cells cross-matched to FoxA2 ChIP data from adult mouse liver cells (Tamplin et al., 2011).

When compared with our full set of 1364 *Ciona* notochord-enriched genes, the 112 that are also notochord expressed in mice and/or fish have a similar distribution of notochord enrichment values, but higher absolute expression levels. The mean FPKM at stage 19.5 of genes with conserved notochord expression was 2.37-fold higher than for all genes enriched at that stage ($D=0.388$, $P=1.421\times 10^{-11}$ by the Kolmogorov–Smirnov test). This largely reflected the absence of genes with low notochord FPKMs in the conserved gene set, which is potentially of evolutionary significance but more likely reflects the more-sensitive RNAseq assay in which the *Ciona* notochord genes were here identified.

The 12 core genes shared by all three species fell into four distinct classes. *Brachyury* (Satoh et al., 2012) and *Sall-b* (also called *SALL1/3/4*) (Ott et al., 1996; Thisse et al., 2001) (Fig. 3) are both notochord transcription factors. Another key notochord transcription factor, *FoxAa*, is also expressed in notochord in all three species, but was absent from our high-throughput analysis because the mouse and *Ciona* genes are somewhat divergent and were not matched in the genome-wide ortholog database in ANISEED, despite their orthology having been previously confirmed (Yagi et al., 2003). *Collagen2A1*, *fibronectin1*, *prolyl hydroxylase* (also called *LEPREL1/2*) and *PAPS Synthetase* all encode either structural components of the ECM or proteins involved in ECM post-translational modifications (Faiyaz ul Haque et al., 1998; Hynes and Naba, 2016). *Prickle*, *Daam1/2* and *ERM* (the single *Ciona* ortholog of ezrin/radixin/moesin) all encode proteins involved in controlling the actin cytoskeleton, either directly or via roles in

Table 3. *Ciona* notochord genes with conserved notochord expression in both mice and zebrafish

KH2012 gene model	<i>Ciona</i> gene name	Mouse gene name	MGI ID*	Zebrafish gene name	ZFIN ID (ZDB-GENE)	<i>In situ</i> hybridization in <i>Ciona</i>
KH.C1.1224	LEPREL1/2	P3h3	1315208			
		P3h1	1888921	p3h1	031010-35	Takahashi et al. (1999)
		P3h2	2146663			
KH.C1.631	LRIG1/2/3	Lrig1	107935	lrig1	031113-23	
KH.C10.209	DAAM1/2	Daam2	1923691	daam1b	030131-4212	
KH.C12.129	ERM	Ezr	98931	msna	021211-2	Hotta et al. (2000)
KH.C2.187	PAPSS1/2	Papss1	1330587	papss2b	010323-5	Hotta et al. (2000)
KH.C3.416	CRIM1	Crim1	1354756	crim1	040312-2	
KH.C4.671	Sall-b	Sall3	109295	sall1a	020228-2	Fig. 3
KH.C7.633	COL2A1	Col2a1	88452	col2a1a	980526-192	Kugler et al. (2010) and Fig. 2
				col1a1a	030131-9102	
				prickle2b	030724-6	
KH.C8.316	Prickle	Prickle2	1925144	prickle1b	030131-2152	Hotta et al. (2000)
				prickle1a	030724-5	
				nav3	021205-1	Fig. 2
KH.L108.20	NAV1/2/3	Nav3	2183703	nav3	021205-1	Fig. 2
KH.S1404.1	Brachyury	T	98472	tb	081028-48	Corbo et al. (1997b)
KH.S417.6	FN1	Fn1	95566	fn1b	030131-6545	Kugler et al. (2008)

*FoxAa (KH.C11.331) is also expressed in the notochord of all three species, although the mouse and *Ciona* genes were not identified as orthologs in the ANISEED database (KH2012 orthologs).

noncanonical Wnt/PCP signaling (Tsukita and Yonemura, 1999; Skoglund and Keller, 2010). *Prickle* and *Daam1/2* have both been implicated in mediolateral intercalation (Habas et al., 2001; Takeuchi et al., 2003; Veeman et al., 2003), whereas a *Ciona* *ERM* morpholino had a later defect in the disk-to-drum notochord cell shape change that follows intercalation (Hotta et al., 2007b). The fourth class is an ad hoc grouping of *LRIG*, *CRIM1* and *NAV3*. All three have previously been implicated in aspects of cell signaling and/or the cytoskeleton, but have not been extensively characterized (Wilkinson et al., 2003; Kinna et al., 2006; van Haren et al., 2014; Simion et al., 2014). Their conserved expression across broad phylogenetic distances makes them attractive candidates for future functional studies of their roles in notochord morphogenesis.

DISCUSSION

Evolutionary origins of the notochord

The evolutionary origins of the notochord have long been debated. It is unique to the chordates and there are no compelling transitional fossils. The hemichordate stomochord is superficially reminiscent, but lacks *Brachyury* expression and a notochord-like basement membrane, and is now thought to be derived from pharyngeal endoderm (Peterson et al., 1999; Satoh et al., 2014). It has recently been proposed that a structure homologous to the notochord can be found in the ventral midline mesoderm of protostomes (Lauri et al., 2014; Brunet et al., 2015). This ‘axochord’ was characterized in the annelid *Platynereis* by virtue of its conserved expression of a set of transcription factors and other genes that are expressed in the notochord of many chordates. Not all were specific to the axochord, but the overlap of their complex expression patterns was relatively specific to this structure.

Given the considerable support for the dorsoventral inversion hypothesis at the level of gene expression (Arendt and Nübler-Jung, 1997), it is in many respects not surprising that ventral midline mesoderm in a protostome might have a conserved transcriptional signature with respect to dorsal midline mesoderm in chordates. It has been suggested, however, that the axochord actually engages in notochord-like behaviors such as mediolateral intercalation (Lauri et al., 2014), and that the notochord might in some respects thus have surprisingly deep origins before the divergence of protostomes and deuterostomes. We would argue, however, that many tissues engage

in intercalatory behaviors independently of the noncanonical Wnt/PCP-dependent tissue rearrangements that result in the massive convergence and extension of the chordate notochord (Zallen and Wieschaus, 2004; King et al., 2009; Walck-Shannon and Hardin, 2014). It is thus of great importance to know whether the axochord also expresses a broad range of notochord effector genes.

Ciona lacks notochord-specific expression of several genes, including *Noggin*, *Twist*, *SoxD*, *SoxE*, *Hedgehog* (*Hh*) and *Slit*, that were used as key markers of notochord identity by Lauri et al. (2014) (Table S12). Many of these are expressed in the notochord in *Amphioxus* as well as vertebrates, likely reflecting lineage-specific loss of notochord expression in the tunicates (Lauri et al., 2014). These genes were thus likely expressed in notochord in the last common ancestor of the chordates, but it is important to note that they do not all appear to be key effector genes for notochord morphogenesis per se. *Noggin* and *chordin*, for example, are involved in the function of the axial mesoderm as the Spemann/Mangold dorsal organizer (Smith and Harland, 1992; Smith et al., 1993; Sasai et al., 1994). *Netrin*, *Slit* and *Hh* are involved in later aspects of neural patterning and axon pathfinding (Echelard et al., 1993; Kennedy et al., 1994; Roelink et al., 1994; Yuan et al., 1999). Tunicates have lost the dorsal organizer as a key mediator of DV patterning (Lemaire et al., 2008) and appear to make less use of the notochord as a source of midline patterning cues (Yamada et al., 2009), but they nevertheless make a stereotypically chordate notochord that undergoes all of the expected morphogenetic behaviors. To better understand the evolutionary origins of the notochord, it will thus be important to determine the protostome expression patterns for a broad range of notochord-enriched effector genes conserved across the chordates, with a particular emphasis on genes likely to play essential conserved roles in notochord morphogenesis. Key questions suggested by our data are whether the protostome axochord is similarly enriched for *Prickle* or other PCP components, and whether or not it has a similar signature of ECM components beyond collagen.

Notochord morphogenesis

Ciona notochord morphogenesis involves a broad range of morphogenetic behaviors, including asymmetric division (Veeman and Smith, 2013), mediolateral intercalation (Munro and Odell, 2002;

Jiang et al., 2005), boundary capture (Veeman et al., 2008), controlled changes in aspect ratio (Dong et al., 2011) and tubulogenesis (Deng et al., 2013; Denker et al., 2013). This richness of cell behavior can be studied *in toto* in the 40 *Ciona* notochord cells, making it an attractive model for integrative studies of how tissue-specific gene expression controls tissue-specific morphogenetic behaviors. Particularly compelling candidate genes for loss-of-function studies include the large set of ECM structural components and modifying enzymes, the ARP2/3 complex components, and the various channels and transporters. *LRIG*, *CRIMI* and *NAV3* are also of particular interest given their conserved notochord expression in mice and fish.

Rethinking the role of Brachyury in the notochord GRN

Beginning with the early effort of Takahashi and colleagues to systematically identify notochord-specific genes by Brachyury overexpression and subtractive hybridization (Hotta et al., 1999; Takahashi et al., 1999), the *Ciona* notochord has become an important model for studying the mechanisms of tissue-specific gene expression (Passamaneck et al., 2009; José-Edwards et al., 2013, 2015; Katikala et al., 2013; Farley et al., 2016). The 1364 genes with notochord-enriched expression identified here represent a more than 10-fold increase in known notochord genes, and provide a comprehensive, genome-wide view of gene expression in the notochord during key stages of mediolateral intercalation. The *Ciona* notochord transcriptome provides a foundation for systematically dissecting the notochord effector gene regulatory network and systems-level studies of tissue-specific enhancer function.

Although used inconsistently in the literature, master regulator genes are typically defined as genes at the top of tissue-specific transcriptional cascades that are both necessary for initiating tissue-specific transcriptional programs and also capable of reprogramming other cell types to that fate when ectopically expressed (Chan and Kyba, 2013), as has been demonstrated with well known examples such as MyoD (Tapscott et al., 1988) and *Drosophila* *eyeless* (Halder et al., 1995). Brachyury has long been treated as a *de facto* master regulator gene for the *Ciona* notochord, in the sense of being both necessary for notochord fate but also sufficient to induce notochord fate when expressed in other cell types. Its necessity is well established, as various loss-of-function assays, including a likely null mutation show that it is required for notochord development (Satou et al., 2001a; Yamada et al., 2003; Chiba et al., 2009). If truly sufficient for notochord fate, however, misexpressing it outside the notochord would recapitulate the normal notochord transcriptional program in these other cell types. This clearly happens on a certain level, as Brachyury misexpression induces expression of many notochord-specific markers (Hotta et al., 1999; Takahashi et al., 1999). These arguments are confounded, however, by the fact that most previously known notochord markers were identified specifically by screening for transcriptional targets of Brachyury misexpression.

Here, we have identified a large and unbiased set of notochord-enriched genes and find that relatively few of them are upregulated in response to ectopic Brachyury (Fig. 4E) *FoxAa*>Bra-expressing cells instead appear to take on a hybrid character in which some but not all notochord markers are upregulated, some but not all markers of endodermal, mesenchymal and neural fate are downregulated, and many other genes are ectopically upregulated (Fig. 4F). This potentially reflects the combinatorial control of gene expression as influenced by Brachyury in combination with other transcriptional regulators expressed in *FoxAa*-expressing cells but not in wild-type

notochord cells. The idea that Brachyury misexpression only partially transforms other cell types to notochord fate is supported by the observation that most *FoxAa*>Bra-expressing cells outside the notochord form an overtly undifferentiated mass and never undergo mediolateral intercalation or other characteristic notochord cell behaviors. Notochord-enriched gene expression appears almost entirely normal in cleavage-blocked embryos, so the failure of *FoxAa*>Bra-expressing cells to morphologically differentiate as notochord likely reflects these transcriptional differences and not simply defects in embryonic architecture. An intriguing question for the future is whether the observation of numerous genes that are notochord enriched but not upregulated by *FoxAa*>Bra is more reflective of factors expressed outside the notochord that repress the notochord transcriptional regime, or whether other transcription factors might act in parallel to Brachyury at the top of the notochord GRN. In support of the latter hypothesis, we note that, despite being a likely null allele, the *Ciona* *brachyury* mutant *chobi* has a relatively mild phenotype in which a distinct notochord forms but has defects in intercalation and elongation (Chiba et al., 2009). In addition, at least one notochord transcription factor is expressed normally in *chobi* mutant embryos (José-Edwards et al., 2011).

MATERIALS AND METHODS

Ciona husbandry and methods

Ciona intestinalis type A were collected in San Diego (CA, USA) and shipped to Kansas State University by Marine Research and Educational Products. Although long identified as *Ciona intestinalis*, recent analysis has suggested that type A should potentially be considered a distinct species named *C. robusta* (Pennati et al., 2015). Adult *Ciona* were maintained in a recirculating aquarium. Standard fertilization, dechoriation and electroporation protocols were used (Veeman et al., 2011). For sorting experiments, 80 µg of *Bra*>GFP plasmid was electroporated.

Dissociation details

For each of the three biological replicates we collected 2000 electroporated embryos and 500 unelectroporated control embryos at each timepoint. Embryo dissociation was modified from methods described previously (Christiaen et al., 2008). Embryos were washed twice with Ca²⁺/Mg²⁺-free artificial seawater [CMF-ASW; 449 mM NaCl, 9 mM KCl, 33 mM Na₂SO₄, 2.15 mM NaHCO₃, 10 mM Tris-HCl (pH 8.0), 2.5 mM EGTA] in BSA-treated eppendorfs, spinning at 730 g for 1 min to pellet. Embryos were resuspended in 600 µl CMF-ASW and 60 µl of 1% trypsin (0.1% final concentration). Samples were pipetted for 3 min to dissociate embryos, then 600 µl of ice-cold CMF-ASW+0.1% BSA was added to stop the reaction. Cells were spun for 3 min at 730 g, washed once with 1 ml ice-cold CMF-ASW+BSA and resuspended in 300 µl ice-cold CMF-ASW+BSA. Cells were filtered through a BSA-treated 70 µm filter into 200 µl ice-cold CMF-ASW+BSA. Final volume was 500 µl.

FACS

Cell sorting was performed using a Biorad S3 at the Flow Cytometry Lab in Kansas State University College of Veterinary Medicine. For each timepoint, an unelectroporated cell sample was first analyzed to determine the levels of background autofluorescence, and gates on forward scatter and GFP fluorescence were conservatively selected to enrich for single cells with or without strong GFP signal. The electroporated sample was then sorted directly into RNA lysis buffer (Zymogen Quick-RNA miniprep kit R1054). For GFP-positive notochord-enriched cells, 2000–15,000 cells were purified into 500 µl buffer, while 100,000 GFP-negative (notochord depleted) cells were simultaneously sorted into 2 ml of lysis buffer. Samples were stored at –80°C until RNA purification.

Library prep and sequencing

RNA was purified with the Zymogen Quick-RNA miniprep kit (R1054) and eluted in 50 µl elution buffer. The yield was 0.6 to 1.25 µg of RNA from

notochord-depleted cells and 40-140 ng from notochord-enriched cells. RNA (10-60 ng per sample) was treated with DNase (Roche) and concentrated to 10 μ l with the Zymogen Clean and Concentrator kit (R1015). The final purified RNA concentration was 0.5-4.0 ng/ μ l.

Amplified cDNA was prepared using the NuGEN Ovation RNA-Seq System V2 (7102). Purified RNA (2.5 μ l) was used in half-reactions following their standard protocol. cDNA samples were fragmented with a Covaris S220 Ultrasonicator using the recommended settings to obtain an average fragment length of 300 bp and the NEBNext DNA Library Prep Master Mix set for Illumina from New England BioLabs (NEB) was used for library construction. The resulting libraries were quality checked with an Agilent 2100 Bioanalyzer and quantified by RT-PCR. Paired-end (2 \times 100 bp) sequencing was performed on an Illumina HiSeq 2500 at the Kansas University Genome Sequencing Core with a read depth of 30-80 million reads per sample.

FoxAa>Bra misexpression and RNAseq

Fertilized dechorionated eggs were electroporated with either 40 μ g of *FoxAa*>Venus-Bra plasmid or 40 μ g of control plasmid (*Bra*<GFP) on three separate days. For each of the replicates, 400 experimental and 400 control embryos were collected at stage 19.5 into RNA lysis buffer and stored at -80°C until RNA purification. RNA was purified with the Zymogen Quick-RNA miniprep kit (R1054), genomic DNA was removed using the in-column DNase treatment step in the kit and purified RNA was eluted in 50 μ l water, at a concentration of 52-60 ng/ μ l. Libraries were constructed with the TruSeq stranded mRNA library kit (Illumina) using standard protocols. Libraries were quality checked with an Agilent TapeStation and quantified by qPCR. Single-end (1 \times 100) sequencing was performed on an Illumina HiSeq 2500 at the Kansas University Genome Sequencing Core with a read depth of 33.5 to 38.5 million reads per sample.

Cytochalasin treatment and RNAseq

Dechorionated embryos were treated with 10 μ M cytochalasinB (1.4 μ l of 21 mM cytochalasin-DMSO stock in 3 ml of ASW) or with DMSO alone as a control. Drug was added at 3.75 hpf at 21.5C (64-cell stage). Embryo harvesting at stage 19.5, RNA purification, library preparation and sequencing were as described for *FoxAa* overexpression. Read depth was 8.5 to 11.6 million reads per sample.

Data analysis

Cufflinks (Trapnell et al., 2012, 2013) was used for expression quantitation. Reads were aligned to the *Ciona* KH2012 gene models (KH.KHGene.2012.gff3, retrieved from ANISEED, <http://www.aniseed.cnrs.fr/>) (Brozovic et al., 2016) with Tophat, abundances were calculated with Cuffquant, and differential expression, splicing and promoter usage were calculated with Cuffdiff using the default parameters. Statistical significance was reported as q-values adjusted for multiple comparisons using a false discovery rate of 0.05.

TopGO (Alexa and Rahnenfuhrer, 2016) was used to test for enrichment of GO terms in our notochord-expressed gene lists. GO annotations for the KH2012 gene models were downloaded from ANISEED. Enriched gene lists for each stage were tested against a limited universe of genes with an FPKM of at least 0.5 in the notochord-depleted and/or notochord-enriched cell populations at the appropriate stage. Enriched GO terms were identified using Fisher's exact test and the weight01 algorithm in topGo with a *P*-value cutoff of 5×10^{-4} . This algorithm does not require adjustment for multiple comparisons.

In situ hybridization

Probe synthesis, embryo collection and *in situ* hybridization were performed as described (Reeves et al., 2014). Clones used as templates for probe synthesis were obtained from the *Ciona* Gateway whole-ORF gene collection (Satou et al., 2002), or amplified from cDNA and cloned into pBSII SK(-). The gene collection IDs or PCR primers are listed in Table S13.

Notochord gene conservation

Genes expressed in the fish 'notochord' or 'axial chorda mesoderm' from shield to 22- to 25-somite stages were retrieved from the Zebrafish Model

Organism Database (ZFIN) (<http://zfin.org/>) in September 2016. Genes expressed in the mouse notochord from Thiler stage 12-21 were retrieved from the Gene Expression Database (GXD) (<http://www.informatics.jax.org/>) in September 2016. KH2012 orthologs were downloaded from ANISEED and used to identify genes with conserved notochord expression.

Acknowledgements

The authors thank the University of Kansas Genome Sequencing Core, the KSU CVM Confocal Core, the KSU CVM Flow Cytometry Lab and the KSU Integrated Genomics Facility for technical support.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: W.M.R., M.T.V.; Methodology: W.M.R., M.T.V.; Validation: W.M.R., Y.W., M.J.H.; Formal analysis: W.M.R., M.T.V.; Investigation: W.M.R., Y.W., M.J.H., M.T.V.; Data curation: W.M.R.; Writing - original draft: W.M.R., M.T.V.; Writing - review & editing: W.M.R., M.T.V.; Visualization: W.M.R., M.T.V.; Supervision: M.T.V.; Project administration: M.T.V.; Funding acquisition: M.T.V.

Funding

This work was supported by the National Institutes of Health (1R01HD085909), and subawards from the Kansas Idea Network for Biomedical Research Excellence (NIH P20GM103418) and the Center for Molecular Analysis of Disease Phenotypes COBRE (NIH P20GM103638). M.J.H. was initially supported by a National Science Foundation-REU fellowship to Kansas State University, Biology (NSF DBI-1156571). Deposited in PMC for release after 12 months.

Data availability

RNAseq datasets have been deposited in GEO (accession numbers GSE95333, GSE102141 and GSE102142). Gene expression patterns are available in the ANISEED ascidian community database (https://www.aniseed.cnrs.fr/aniseed/experiment/find_insitu).

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.156174.supplemental>

References

- Alexa, A. and Rahnenfuhrer, J. (2016). topGO: Enrichment analysis for Gene Ontology. R package version 2.28.0. <http://bioconductor.org/packages/release/bioc/html/topGO.html>.
- Alexa, A., Rahnenfuhrer, J. and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600-1607.
- Angers, S. and Moon, R. T. (2009). Proximal events in Wnt signal transduction. *Nat. Rev. Mol. Cell Biol.* **10**, 468-477.
- Arendt, D. and Nübler-Jung, K. (1997). Dorsal or ventral: similarities in fate maps and gastrulation patterns in annelids, arthropods and chordates. *Mech. Dev.* **61**, 7-21.
- Borghese, L., Fletcher, G., Mathieu, J., Atzberger, A., Eades, W. C., Cagan, R. L. and Rørth, P. (2006). Systematic analysis of the transcriptional switch inducing migration of border cells. *Dev. Cell* **10**, 497-508.
- Brozovic, M., Martin, C., Dantec, C., Dauga, D., Mendez, M., Simion, P., Percher, M., Laporte, B., Scornavacca, C., Di Gregorio, A. et al. (2016). ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. *Nucleic Acids Res.* **44**, D808-D818.
- Brunet, T., Lauri, A. and Arendt, D. (2015). Did the notochord evolve from an ancient axial muscle? The axochord hypothesis. *BioEssays* **37**, 836-850.
- Chan, S. S.-K. and Kyba, M. (2013). What is a master regulator? *J. Stem Cell Res. Ther.* **3**, 680-682.
- Chiba, S., Jiang, D., Satoh, N. and Smith, W. C. (2009). Brachyury null mutant-induced defects in juvenile ascidian endodermal organs. *Development* **136**, 35-39.
- Christiaan, L., Davidson, B., Kawashima, T., Powell, W., Nolla, H., Vranizan, K. and Levine, M. (2008). The transcription/migration interface in heart precursors of *Ciona intestinalis*. *Science* **320**, 1349-1352.
- Corallo, D., Trapani, V. and Bonaldo, P. (2015). The notochord: structure and functions. *Cell. Mol. Life Sci.* **72**, 2989-3008.
- Corbo, J. C., Erives, A., Di Gregorio, A., Chang, A. and Levine, M. (1997a). Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. *Development* **124**, 2335-2344.
- Corbo, J. C., Levine, M. and Zeller, R. W. (1997b). Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589-602.

- Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M. et al. (2002). The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157-2167.
- Delsuc, F., Brinkmann, H., Chourrout, D. and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965-968.
- Deng, W., Nies, F., Feuer, A., Bocina, I., Oliver, D. and Jiang, D. (2013). Anion translocation through an Slc26 transporter mediates lumen expansion during tubulogenesis. *Proc. Natl. Acad. Sci. USA* **110**, 14972-14977.
- Denker, E., Bocina, I. and Jiang, D. (2013). Tubulogenesis in a simple cell cord requires the formation of bi-apical cells through two discrete Par domains. *Development* **140**, 2985-2996.
- Di Gregorio, A. (2017). T-box genes and developmental gene regulatory networks in ascidians. *Curr. Top. Dev. Biol.* **122**, 55-91.
- Dong, B., Deng, W. and Jiang, D. (2011). Distinct cytoskeleton populations and extensive crosstalk control *Ciona* notochord tubulogenesis. *Development* **138**, 1631-1641.
- Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A. and McMahon, A. P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417-1430.
- Faiyaz ul Haque, M., King, L. M., Krakow, D., Cantor, R. M., Rusiniak, M. E., Swank, R. T., Superti-Furga, A., Haque, S., Abbas, H., Ahmad, W. et al. (1998). Mutations in orthologous genes in human spondyloepimetaphyseal dysplasia and the brachymorphic mouse. *Nat. Genet.* **20**, 157-162.
- 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. USA* **113**, 6508-6513.
- Finger, J. H., Smith, C. M., Hayamizu, T. F., McCright, I. J., Xu, J., Law, M., Shaw, D. R., Baldarelli, R. M., Beal, J. S., Blodgett, O. et al. (2017). The mouse gene expression database (GXD): 2017 update. *Nucleic Acids Res.* **45**, D730-D736.
- Flood, P. R., Guthrie, D. M. and Banks, J. R. (1969). Paramyosin muscle in the notochord of Amphioxus. *Nature* **222**, 87-88.
- Friedman, J. R. and Kaestner, K. H. (2006). The Foxa family of transcription factors in development and metabolism. *Cell. Mol. Life Sci.* **63**, 2317-2328.
- Gansner, J. M., Madsen, E. C., Mecham, R. P. and Gitlin, J. D. (2008). Essential role for fibrillin-2 in zebrafish notochord and vascular morphogenesis. *Dev. Dyn.* **237**, 2844-2861.
- Habas, R., Kato, Y. and He, X. (2001). Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. *Cell* **107**, 843-854.
- Halder, G., Callaerts, P. and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* **267**, 1788-1792.
- Hart, M. J., de los Santos, R., Albert, I. N., Rubinfeld, B. and Polakis, P. (1998). Downregulation of beta-catenin by human Axin and its association with the APC tumor suppressor, beta-catenin and GSK3 beta. *Curr. Biol.* **8**, 573-581.
- Heisenberg, C.-P., Tada, M., Rauch, G.-J., Saúde, L., Concha, M. L., Geisler, R., Stemple, D. L., Smith, J. C. and Wilson, S. W. (2000). Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76-81.
- Holland, L. Z., Laudet, V. and Schubert, M. (2004). The chordate amphioxus: an emerging model organism for developmental biology. *Cell. Mol. Life Sci.* **61**, 2290-2308.
- Hotta, K., Takahashi, H., Erives, A., Levine, M. and Satoh, N. (1999). Temporal expression patterns of 39 Brachyury-downstream genes associated with notochord formation in the *Ciona intestinalis* embryo. *Dev. Growth Differ.* **41**, 657-664.
- Hotta, K., Takahashi, H., Asakura, T., Saitoh, B., Takatori, N., Satou, Y. and Satoh, N. (2000). Characterization of Brachyury-downstream notochord genes in the *Ciona intestinalis* embryo. *Dev. Biol.* **224**, 69-80.
- Hotta, K., Mitsuhashi, K., Takahashi, H., Inaba, K., Oka, K., Gojobori, T. and Ikeo, K. (2007a). A web-based interactive developmental table for the ascidian *Ciona intestinalis*, including 3d real-image embryo reconstructions: I. from fertilized egg to hatching larva. *Dev. Biol.* **236**, 1790-1805.
- Hotta, K., Yamada, S., Ueno, N., Satoh, N. and Takahashi, H. (2007b). Brachyury-downstream notochord genes and convergent extension in *Ciona intestinalis* embryos. *Dev. Growth Differ.* **49**, 373-382.
- Hotta, K., Takahashi, H., Satoh, N. and Gojobori, T. (2008). Brachyury-downstream gene sets in a chordate, *Ciona intestinalis*: Integrating notochord specification, morphogenesis and chordate evolution. *Evol. Dev.* **10**, 37-51.
- Howe, D. G., Bradford, Y. M., Conlin, T., Eagle, A. E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S. A. T. et al. (2013). ZFIN, the Zebrafish model organism database: increased support for mutants and transgenics. *Nucleic Acids Res.* **41**, D854-D860.
- Hudson, C. and Yasuo, H. (2006). A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in *Ciona* embryos. *Development* **133**, 2855-2864.
- Hynes, R. O. and Naba, A. (2016). Overview of the matrisome—an inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* **4**, a004903.
- Imai, K. S., Hino, K., Yagi, K., Satoh, N. and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* **131**, 4047-4058.
- Jiang, D., Munro, E. M. and Smith, W. C. (2005). Ascidian prickle regulates both mediolateral and anterior-posterior cell polarity of notochord cells. *Curr. Biol.* **15**, 79-85.
- José-Edwards, D. S., Kerner, P., Kugler, J. E., Deng, W., Jiang, D. and Di Gregorio, A. (2011). The identification of transcription factors expressed in the notochord of *Ciona intestinalis* adds new potential players to the brachyury gene regulatory network. *Dev. Dyn.* **240**, 1793-1805.
- José-Edwards, D. S., Oda-Ishii, I., Nibu, Y. and Di Gregorio, A. (2013). Tbx2/3 is an essential mediator within the Brachyury gene network during *Ciona* notochord development. *Development* **140**, 2422-2433.
- José-Edwards, D. S., Oda-Ishii, I., Kugler, J. E., Passamaneck, Y. J., Katikala, L., Nibu, Y. and Di Gregorio, A. (2015). Brachyury, Foxa2 and the cis-regulatory origins of the notochord. *PLoS Genet.* **11**, e1005730.
- Katikala, L., Aihara, H., Passamaneck, Y. J., Gazdoui, S., José-Edwards, D. S., Kugler, J. E., Oda-Ishii, I., Imai, J. H., Nibu, Y. and Di Gregorio, A. (2013). Functional Brachyury binding sites establish a temporal read-out of gene expression in the *Ciona* notochord. *PLoS Biol.* **11**, e1001697.
- Kennedy, T. E., Serafini, T., de la Torre, J. R. and Tessier-Lavigne, M. (1994). Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425-435.
- Kim, E., Magen, A. and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* **35**, 125-131.
- King, R. S., Maiden, S. L., Hawkins, N. C., Kidd, A. R., Kimble, J., Hardin, J. and Walston, T. D. (2009). The N- or C-terminal domains of DSH-2 can activate the C. elegans Wnt/beta-catenin asymmetry pathway. *Dev. Biol.* **328**, 234-244.
- Kinna, G., Kolle, G., Carter, A., Key, B., Lieschke, G. J., Perkins, A. and Little, M. H. (2006). Knockdown of zebrafish *crim1* results in a bent tail phenotype with defects in somite and vascular development. *Mech. Dev.* **123**, 277-287.
- Kubo, A., Suzuki, N., Yuan, X., Nakai, K., Satoh, N., Imai, K. S. and Satou, Y. (2010). Genomic cis-regulatory networks in the early *Ciona intestinalis* embryo. *Development* **137**, 1613-1623.
- Kugler, J. E., Passamaneck, Y. J., Feldman, T. J., Beh, J., Regnier, T. W. and Di Gregorio, A. (2008). Evolutionary conservation of vertebrate notochord genes in the ascidian *Ciona intestinalis*. *Genesis* **46**, 697-710.
- Kugler, J. E., Gazdoui, S., Oda-Ishii, I., Passamaneck, Y. J., Erives, A. J. and Di Gregorio, A. (2010). Temporal regulation of the muscle gene cascade by *Macho1* and *Tbx6* transcription factors in *Ciona intestinalis*. *J. Cell Sci.* **123**, 2453-2463.
- Lauri, A., Brunet, T., Handberg-Thorsager, M., Fischer, A. H. L., Simakov, O., Steinmetz, P. R. H., Tomer, R., Keller, P. J. and Arendt, D. (2014). Development of the annelid axochord: insights into notochord evolution. *Science* **345**, 1365-1368.
- Lemaire, P., Smith, W. C. and Nishida, H. (2008). Ascidians and the plasticity of the chordate developmental program. *Curr. Biol.* **18**, 620-631.
- Malone, J. H. and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**, 34.
- Munro, E. M. and Odell, G. M. (2002). Polarized basolateral cell motility underlies invagination and convergent extension of the ascidian notochord. *Development* **129**, 13-24.
- Myers, D. C., Sepich, D. S. and Solnica-Krezel, L. (2002). Bmp activity gradient regulates convergent extension during zebrafish gastrulation. *Dev. Biol.* **243**, 81-98.
- Oda-Ishii, I., Ishii, Y. and Mikawa, T. (2010). Eph regulates dorsoventral asymmetry of the notochord plate and convergent extension-mediated notochord formation. *PLoS ONE* **5**, e13689.
- Ott, T., Kaestner, K. H., Monaghan, A. P. and Schütz, G. (1996). The mouse homolog of the region specific homeotic gene *spalt* of *Drosophila* is expressed in the developing nervous system and in mesoderm-derived structures. *Mech. Dev.* **56**, 117-128.
- Passamaneck, Y. J. and Di Gregorio, A. (2005). *Ciona intestinalis*: chordate development made simple. *Dev. Dyn.* **233**, 1-19.
- 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.
- Pennati, R., Ficetola, G. F., Brunetti, R., Caicci, F., Gasparini, F., Griggio, F., Sato, A., Stach, T., Kaul-Strehlow, S., Gissi, C. et al. (2015). Morphological differences between larvae of the *Ciona intestinalis* species complex: hints for a valid taxonomic definition of distinct species. *PLoS ONE* **10**, e0122879.
- Peters, J. M., McKay, R. M., McKay, J. P. and Graff, J. M. (1999). Casein kinase I transduces Wnt signals. *Nature* **401**, 345-350.
- Peterson, K. J., Cameron, R. A., Tagawa, K., Satoh, N. and Davidson, E. H. (1999). A comparative molecular approach to mesodermal patterning in basal

- deuterostomes: the expression pattern of Brachyury in the enteropneust hemichordate *Ptychodera flava*. *Development* **126**, 85-95.
- Racioppi, C., Kamal, A. K., Razy-Krajka, F., Gambardella, G., Zanetti, L., di Bernardo, D., Sanges, R., Christiaen, L. A. and Ristoratore, F.** (2014). Fibroblast growth factor signalling controls nervous system patterning and pigment cell formation in *Ciona intestinalis*. *Nat. Commun.* **5**, 4830.
- Reeves, W., Thayer, R. and Veeman, M.** (2014). Anterior-posterior regionalized gene expression in the *Ciona* notochord. *Dev. Dyn.* **243**, 612-620.
- Roelink, H., Augsburger, A., Heemskerk, J., Korzh, V., Norlin, S., Ruiz i Altaba, A., Tanabe, Y., Placzek, M., Edlund, T. and Jessell, T. M.** (1994). Floor plate and motor neuron induction by *vhh-1*, a vertebrate homolog of hedgehog expressed by the notochord. *Cell* **76**, 761-775.
- Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L. K. and De Robertis, E. M.** (1994). *Xenopus* chordin: a novel dorsaling factor activated by organizer-specific homeobox genes. *Cell* **79**, 779-790.
- Satoh, N.** (2013). *Developmental Genomics of Ascidiaceans*. Hoboken, NJ, USA: John Wiley & Sons.
- Satoh, N., Satou, Y., Davidson, B. and Levine, M.** (2003). *Ciona* intestinalis: an emerging model for whole-genome analyses. *Trends Genet.* **19**, 376-381.
- Satoh, N., Tagawa, K. and Takahashi, H.** (2012). How was the notochord born? *Evol. Dev.* **14**, 56-75.
- Satoh, N., Tagawa, K., Lowe, C. J., Yu, J.-K., Kawashima, T., Takahashi, H., Ogasawara, M., Kirschner, M., Hisata, K., Su, Y.-H. et al.** (2014). On a possible evolutionary link of the stomochord of hemichordates to pharyngeal organs of chordates. *Genesis* **52**, 925-934.
- Satou, Y., Imai, K. S. and Satoh, N.** (2001a). Action of morpholinos in *Ciona* embryos. *Genesis* **30**, 103-106.
- Satou, Y., Takatori, N., Yamada, L., Mochizuki, Y., Hamaguchi, M., Ishikawa, H., Chiba, S., Imai, K., Kano, S., Murakami, S. D. et al.** (2001b). Gene expression profiles in *Ciona intestinalis* tailbud embryos. *Development* **128**, 2893-2904.
- Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T., Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasakura, Y. et al.** (2002). A cDNA resource from the basal chordate *Ciona intestinalis*. *Genesis* **33**, 153-154.
- Segade, F., Cota, C., Famiglietti, A., Cha, A. and Davidson, B.** (2016). Fibronectin contributes to notochord intercalation in the invertebrate chordate, *Ciona intestinalis*. *Evodevo* **7**, 21.
- Simion, C., Cedano-Prieto, M. E. and Sweeney, C.** (2014). The LRIG family: enigmatic regulators of growth factor receptor signaling. *Endocr. Relat. Cancer* **21**, R431-R443.
- Skoglund, P. and Keller, R.** (2007). *Xenopus* fibrillin regulates directed convergence and extension. *Dev. Biol.* **301**, 404-416.
- Skoglund, P. and Keller, R.** (2010). Integration of planar cell polarity and ECM signaling in elongation of the vertebrate body plan. *Curr. Opin. Cell Biol.* **22**, 589-596.
- Smith, W. C. and Harland, R. M.** (1992). Expression cloning of *noggin*, a new dorsaling factor localized to the Spemann organizer in *Xenopus* embryos. *Cell* **70**, 829-840.
- Smith, W. C., Knecht, A. K., Wu, M. and Harland, R. M.** (1993). Secreted *noggin* protein mimics the Spemann organizer in dorsalizing *Xenopus* mesoderm. *Nature* **361**, 547-549.
- Stemple, D. L.** (2005). Structure and function of the notochord: an essential organ for chordate development. *Development* **132**, 2503-2512.
- Swaney, K. F. and Li, R.** (2016). Function and regulation of the Arp2/3 complex during cell migration in diverse environments. *Curr. Opin. Cell Biol.* **42**, 63-72.
- Takahashi, H., Hotta, K., Erives, A., Di Gregorio, A., Zeller, R. W., Levine, M. and Satoh, N.** (1999). Brachyury downstream notochord differentiation in the ascidian embryo. *Genes Dev.* **13**, 1519-1523.
- Takeuchi, M., Nakabayashi, J., Sakaguchi, T., Yamamoto, T. S., Takahashi, H., Takeda, H. and Ueno, N.** (2003). The prickle-related gene in vertebrates is essential for gastrulation cell movements. *Curr. Biol.* **13**, 674-679.
- Tamplin, O. J., Cox, B. J. and Rossant, J.** (2011). Integrated microarray and ChIP analysis identifies multiple *Foxa2* dependent target genes in the notochord. *Dev. Biol.* **360**, 415-425.
- Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P. F., Weintraub, H. and Lassar, A. B.** (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* **242**, 405-411.
- Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R. et al.** (2010). The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. *Genome Res.* **20**, 1459-1468.
- Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X. Q. et al.** (2001). Expression of the zebrafish genome during embryogenesis (NIH RO1 RR15402). *ZFIN Direct Data Submission* <http://zfinfo.org>
- Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., Kirchner, J., Parkhill, J.-P. and Thisse, C.** (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol.* **77**, 505-519.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L.** (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562-578.
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L.** (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46-53.
- Tsukita, S. and Yonemura, S.** (1999). Cortical actin organization: lessons from ERM (ezrin/radixin/moesin) proteins. *J. Biol. Chem.* **274**, 34507-34510.
- van Haren, J., Boudeau, J., Schmidt, S., Basu, S., Liu, Z., Lammers, D., Demmers, J., Benhari, J., Grosveld, F., Debant, A. et al.** (2014). Dynamic microtubules catalyze formation of navigator-TRIO complexes to regulate neurite extension. *Curr. Biol.* **24**, 1778-1785.
- Veeman, M. T. and Smith, W. C.** (2013). Whole-organ cell shape analysis reveals the developmental basis of ascidian notochord taper. *Dev. Biol.* **373**, 281-289.
- Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. and Moon, R. T.** (2003). Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680-685.
- Veeman, M. T., Nakatani, Y., Hendrickson, C., Ericson, V., Lin, C. and Smith, W. C.** (2008). Chongmague reveals an essential role for laminin-mediated boundary formation in chordate convergence and extension movements. *Development* **135**, 33-41.
- Veeman, M. T., Chiba, S. and Smith, W. C.** (2011). *Ciona* genetics. *Methods Mol. Biol.* **770**, 401-422.
- Vienne, A. and Pontarotti, P.** (2006). Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int. J. Biol. Sci.* **2**, 32-37.
- Walck-Shannon, E. and Hardin, J.** (2014). Cell intercalation from top to bottom. *Nat. Rev. Mol. Cell Biol.* **15**, 34-48.
- Wallingford, J. B., Rowling, B. A., Vogeli, K. M., Rothbacher, U., Fraser, S. E. and Harland, R. M.** (2000). Dishevelled controls cell polarity during *Xenopus* gastrulation. *Nature* **405**, 81-85.
- Wilkinson, L., Kolle, G., Wen, D., Piper, M., Scott, J. and Little, M.** (2003). CRIM1 regulates the rate of processing and delivery of bone morphogenetic proteins to the cell surface. *J. Biol. Chem.* **278**, 34181-34188.
- Yagi, K., Satou, Y., Mazet, F., Shimeld, S. M., Degnan, B., Rokhsar, D., Levine, M., Kohara, Y. and Satoh, N.** (2003). A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. III. Genes for Fox, ETS, nuclear receptors and Nf-kappaB. *Dev. Genes Evol.* **213**, 235-244.
- Yamada, L., Shoguchi, E., Wada, S., Kobayashi, K., Mochizuki, Y., Satou, Y. and Satoh, N.** (2003). Morpholino-based gene knockdown screen of novel genes with developmental function in *Ciona intestinalis*. *Development* **130**, 6485-6495.
- Yamada, S., Hotta, K., Yamamoto, T. S., Ueno, N., Satoh, N. and Takahashi, H.** (2009). Interaction of notochord-derived fibrinogen-like protein with Notch regulates the patterning of the central nervous system of *Ciona intestinalis* embryos. *Dev. Biol.* **328**, 1-12.
- Yamanaka, Y., Tamplin, O. J., Beckers, A., Gossler, A. and Rossant, J.** (2007). Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. *Dev. Cell* **13**, 884-896.
- Yuan, W., Zhou, L., Chen, J.-H., Wu, J. Y., Rao, Y. and Ornitz, D. M.** (1999). The mouse SLIT family: secreted ligands for ROBO expressed in patterns that suggest a role in morphogenesis and axon guidance. *Dev. Biol.* **212**, 290-306.
- Zallen, J. A. and Wieschaus, E.** (2004). Patterned gene expression directs bipolar planar polarity in *Drosophila*. *Dev. Cell* **6**, 343-355.

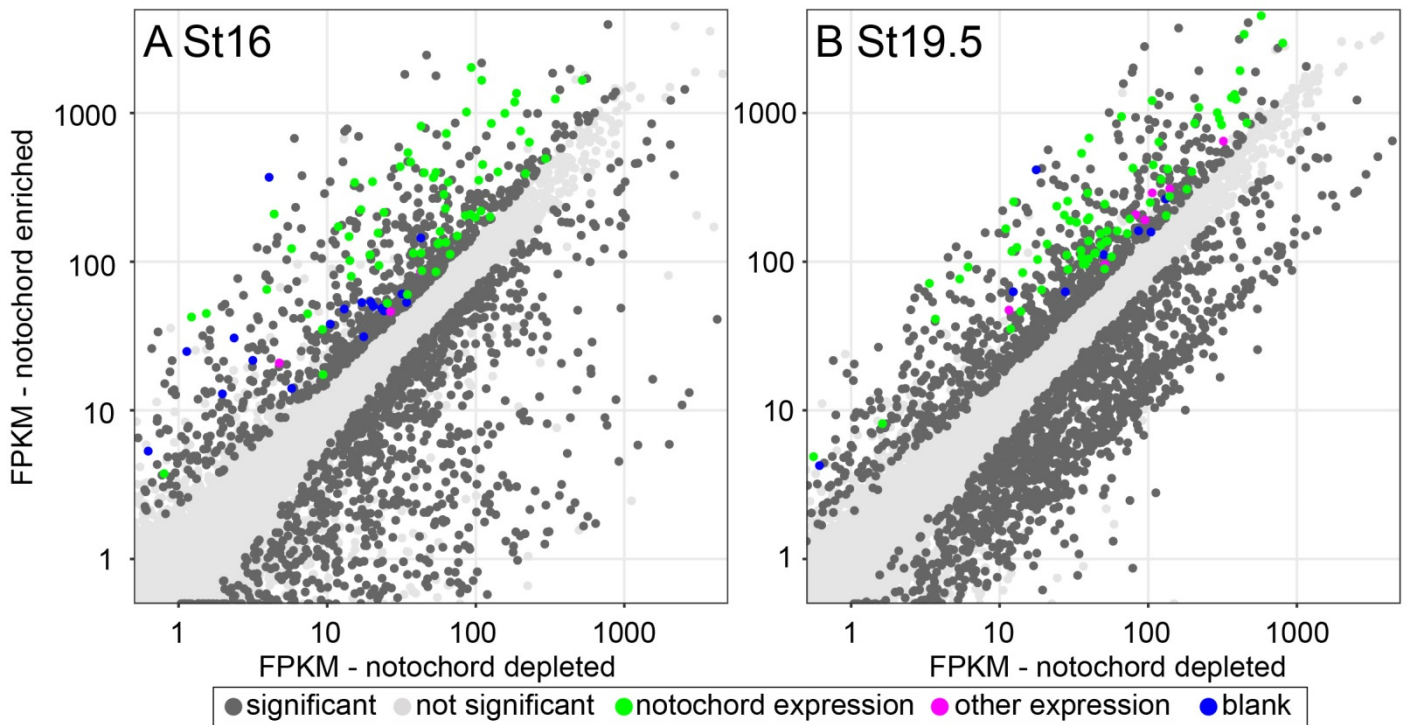


Figure S1. **In situ validation of notochord-enriched genes.** Gene expression in notochord depleted and notochord enriched cell populations at A) Stage 16 and at B) Stage 19.5. In situ results for tested genes are only shown at the stage(s) when they are notochord-enriched by FACSseq. dark grey = significant difference in expression between populations ($q < 0.05$), light grey = no significant difference, green = notochord enriched expression confirmed by in situ hybridization, magenta = expression in other tissues, blue = no expression detected by in situ.

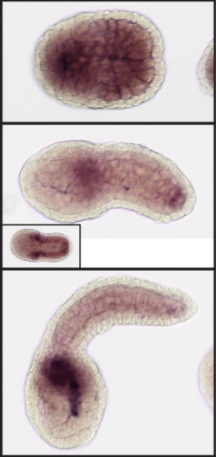
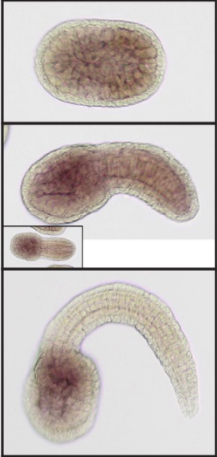
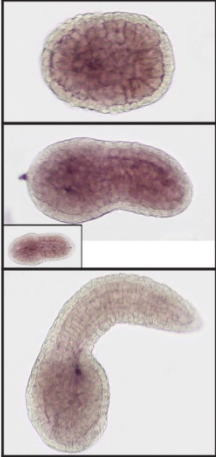
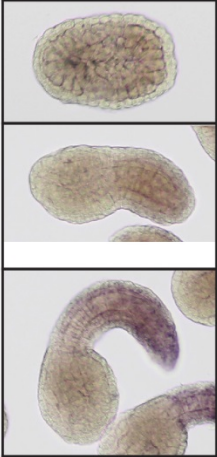
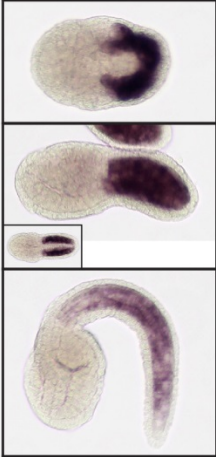
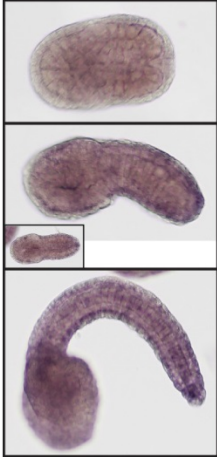

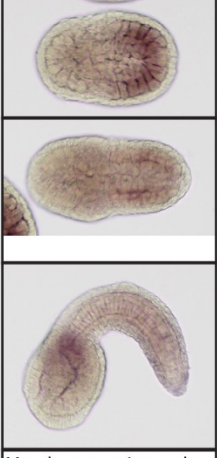
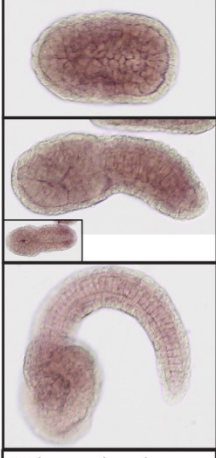
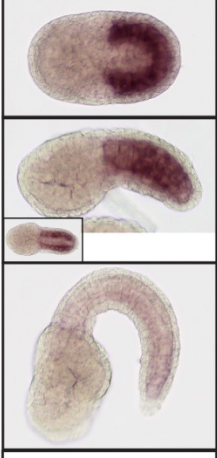
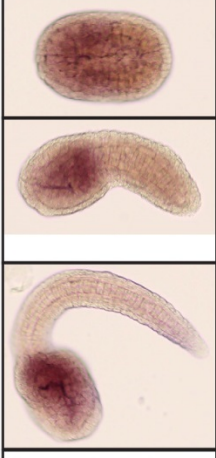
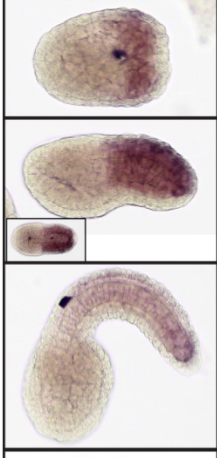
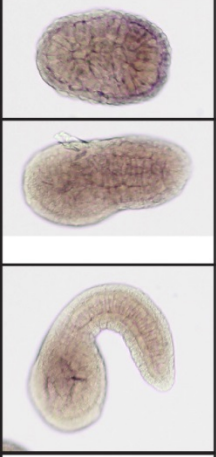
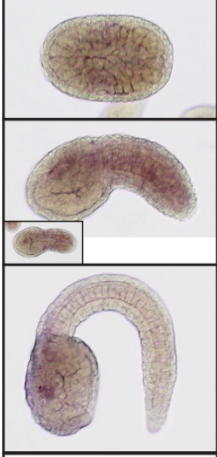
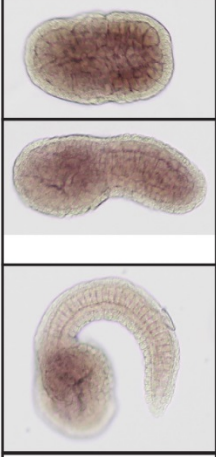
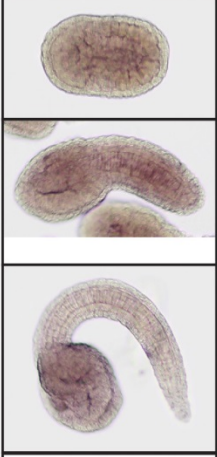
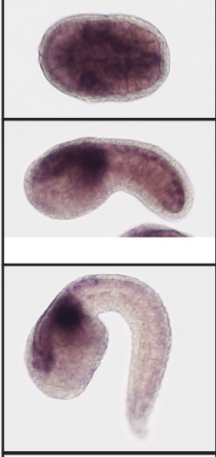
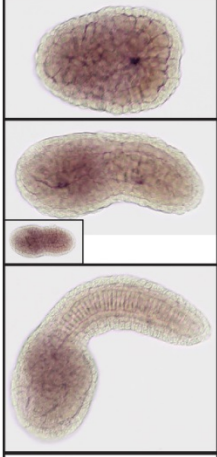
<p>C1.1101</p>  <p>Ubiquitous, except no staining in epidermis. Strongest in muscle and mesenchyme.</p>	<p>C1.73</p>  <p>Widespread expression early, with strongest expression in trunk later.</p>	<p>C1.860</p>  <p>Ubiquitous, except no staining in epidermis.</p>	<p>C10.8</p>  <p>No visible expression early. Tail epidermis late.</p>	<p>C11.121</p>  <p>Muscle</p>	<p>C12.20</p>  <p>Ubiquitous expression. Stronger in epidermis and tips of notochord late.</p>
<p>C12.51</p>  <p>Weak widespread expression early with strongest expression in mesenchyme.</p>	<p>C3.187</p>  <p>Muscle expression early, with increased expression in mesenchyme later.</p>	<p>C3.526</p>  <p>Widespread weak expression at all stages.</p>	<p>C4.190</p>  <p>Muscle.</p>	<p>C6.1</p>  <p>Strongest in trunk endoderm and mesenchyme.</p>	<p>C6.324</p>  <p>Muscle.</p>
<p>C7.106</p>  <p>Weak muscle early.</p>	<p>C8.265</p>  <p>Weak muscle and notochord early.</p>	<p>C8.429</p>  <p>Weak muscle early and expression in trunk later.</p>	<p>C9.424</p>  <p>Weak notochord early.</p>	<p>S403.5</p>  <p>Strongest in neural and mesenchyme. Weak widespread expression.</p>	<p>S598.2</p>  <p>Weak widespread expression. Slightly stronger in mesenchyme.</p>

Figure S2. Expression patterns of selected *FoxAa*>*Bra* induced genes. Expression patterns of the indicated genes in embryos fixed at Stage 16 (top), Stage 19.5 (middle) and Stage 23 (bottom). For all images, anterior is to the left. For the stage 16 embryos, dorsal is towards the viewer. For the stage 19.5 and 23 embryos, dorsal is towards the top of the page. Insets at stage 19.5 are of dorsal view.

Tables

Table S1. Notochord genes identified in previous screens

[Click here to Download Table S1](#)

Table S2. Genes with notochord enriched expression at stages 16 and 19.5

[Click here to Download Table S2](#)

Table S3. Gene Ontology code enrichment for notochord enriched genes

[Click here to Download Table S3](#)

Table S4. Stage 19.5: Differential promoter usage in notochord enriched vs notochord depleted populations

[Click here to Download Table S4](#)

Table S5. Stage 16: Differential splicing in notochord enriched vs notochord depleted populations

[Click here to Download Table S5](#)

Table S6. Genes induced by ectopic Brachyury expression

[Click here to Download Table S6](#)

Table S7. Tissue-specific marker expression in FoxAa>Venus-Bra embryos

[Click here to Download Table S7](#)

Table S8. Genes with altered expression in cytochalasin treated embryos

[Click here to Download Table S8](#)

Table S9. Comparison of notochord FACS-seq, FoxAa>Venus-Bra and Bra ChIPchip datasets

[Click here to Download Table S9](#)

Table S10. Genes with notochord expression in both mouse and Ciona

[Click here to Download Table S10](#)

Table S11. Genes with notochord expression in both zebrafish and Ciona

[Click here to Download Table S11](#)

Table S12. Expression of axochord genes in the Ciona notochord

[Click here to Download Table S12](#)

Table S13. In situ hybridization probes

[Click here to Download Table S13](#)