Developmental Biology 360 (2011) 415–425

Contents lists available at SciVerse ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Genomes and Developmental Control

Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord

Owen J. Tamplin [a,b,1], Brian J. Cox [a], Janet Rossant [a,b,c,*]

[a] Program in Developmental and Stem Cell Biology, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8
[b] Department of Molecular Genetics, Medical Sciences Building, 1 King's College Circle, University of Toronto, Toronto, Ontario, Canada M5S 1A8
[c] Department of Obstetrics and Gynecology, 92 College Street, University of Toronto, Toronto, Ontario, Canada M5G 1L4

ARTICLE INFO

Article history:
Received for publication 29 November 2010
Revised 26 September 2011
Accepted 1 October 2011
Available online 8 October 2011

Keywords:
Noto
Foxa2
Enhancer
Notochord
ChIP-Seq
Microarray

ABSTRACT

The node and notochord are key tissues required for patterning of the vertebrate body plan. Understanding the gene regulatory network that drives their formation and function is therefore important. Foxa2 is a key transcription factor at the top of this genetic hierarchy and finding its targets will help us to better understand node and notochord development. We performed an extensive microarray-based gene expression screen using sorted embryonic notochord cells to identify early notochord-enriched genes. We validated their specificity to the node and notochord by whole mount *in situ* hybridization. This provides the largest available resource of notochord-expressed genes, and therefore candidate Foxa2 target genes in the notochord. Using existing Foxa2 ChIP-seq data from adult liver, we were able to identify a set of genes expressed in the notochord that had associated regions of Foxa2-bound chromatin. Given that Foxa2 is a pioneer transcription factor, we reasoned that these sites might represent notochord-specific enhancers. Candidate Foxa2-bound regions were tested for notochord specific enhancer function in a zebrafish reporter assay and 7 novel notochord enhancers were identified. Importantly, sequence conservation or predictive models could not have readily identified these regions. Mutation of putative Foxa2 binding elements in two of these novel enhancers abrogated reporter expression and confirmed their Foxa2 dependence. The combination of highly specific gene expression profiling and genome-wide ChIP analysis is a powerful means of understanding developmental pathways, even for small cell populations such as the notochord.

Introduction

The vertebrate embryo is defined by the notochord—a transient rod-like tissue that acts as a midline signaling centre (reviewed (Stemple, 2005)). It forms towards the end of gastrulation and is derived from the Spemann/Mangold organizer and node (Beddington, 1994; Kinder et al., 2001; Sulik et al., 1994; Yamanaka et al., 2007). The organizer is a highly conserved tissue in vertebrates that is necessary and sufficient for axis induction during gastrulation (reviewed (Niehrs, 2004)). The node is a late organizer population that forms at the distal tip of the embryonic day (E) 7.5 mouse embryo (Sulik et al., 1994); it can induce a partial posterior axis upon transplantation (Beddington, 1994), and is involved in establishment of the left–right axis (reviewed (Shiratori and Hamada, 2006)). More recently it has been shown that the cells of the notochord later contribute to the intervertebral disks (Choi et al., 2008), suggesting that defects in notochord development may be related to spinal disease. As notochord cells are critical for early development and contribute to mature tissues, it is of interest to understand the genes and regulatory mechanisms that are specific to the notochord.

The network of genes involved in patterning the notochord is only beginning to emerge, in part because the tissue is challenging to study (Robb and Tam, 2004; Tamplin et al., 2008); it is transient and at early stages (E8.0) is only composed of approximately 100 cells per embryo (Beddington, 1994). Our group and others have used micro-dissection and mutant analysis to identify genes expressed specifically in the node and notochord, however these screens only added a small number of novel genes (Frankenberg et al., 2007; Sousa-Nunes et al., 2003; Tamplin et al., 2008). An approach that has improved hit rates in embryonic expression screens is the use of transgenic mice carrying fluorescent reporters to sort small populations of cells; for example, endoderm at E8.25 (Sherwood et al., 2007), and mid-hindbrain organizer at E8.5 (Bouchard et al., 2005). Before the regulatory networks that underlie patterning of the embryonic notochord can be described, more complete gene expression data will need to be collected.

* Corresponding author at: The Hospital for Sick Children Research Institute, TMDT at MaRS, Room 13-305, 101 College Street, Toronto, ON, Canada M5G 1L7. Fax: +1 416 813 5085.
E-mail addresses: otamplin@enders.tch.harvard.edu (O.J. Tamplin), b.cox@utoronto.ca (B.J. Cox), janet.rossant@sickkids.ca (J. Rossant).
[1] Present address: Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115 USA.

0012-1606/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2011.10.002

The description of embryonic gene regulatory networks is essential for our understanding of developmental mechanisms. To complete these networks, various levels of information must be collected: the master regulator transcription factors (TFs) at the top of the hierarchy, the cis-regulatory modules (CRMs; we use this term interchangeably with "enhancer") that receive and integrate inputs from these TFs, and the downstream target genes that carry out specific functions in the developmental program. These networks have been very elegantly described in model organisms with relatively simple embryos, such as the sea urchin, because they have only a few well-defined cell types (Davidson et al., 2002).

Many TFs have been shown to be important for different phases of notochord development, including the following: Brachyury/T (Wilkinson et al., 1990); Noto (Abdelkhalek et al., 2004; Beckers et al., 2007; Yamanaka et al., 2007); Foxa2 (Ang and Rossant, 1994; Weinstein et al., 1994); Zic3 (Elms et al., 2004); Rfx3 (Bonnafe et al., 2004); Glis3 (Kim et al., 2003); TEAD (Sawada et al., 2005, 2008); Sox9 (Bagheri-Fam et al., 2006; Barrionuevo et al., 2006). However, we still have little information on how they interact at the gene regulatory network level to control notochord development.

The identification of functional CRMs in higher order vertebrate genomes presents particular challenges. Comparative genomics for CRM discovery has been successfully applied in vertebrates (Navratilova et al., 2009; Nobrega et al., 2003; Pennacchio et al., 2006; Woolfe et al., 2005) but the predictive value of sequence conservation is highly dependent on the phylogenetic distance between compared species (Fisher et al., 2006a; Odom et al., 2007). Thus extreme sequence conservation (e.g. human to zebrafish) can be overly stringent, thereby removing true positives from the analysis (Fisher et al., 2006a). Conversely, conservation within closely related species (e.g. mammals) can increase the number of candidates for screening by orders of magnitude (reviewed (Haeussler and Joly, 2010)).

A number of notochord CRMs have been identified in ascidian and teleost genomes (Corbo et al., 1997; Dickmeis et al., 2004; Matsumoto et al., 2007; Muller et al., 1999; Passamaneck et al., 2009; Rastegar et al., 2008; Woolfe et al., 2005). However, the number in mouse is small. Two examples of *bona fide* early embryonic notochord CRMs in the mouse include the *Foxa2* mNE enhancer (Nishizaki et al., 2001; Sasaki and Hogan, 1996; Sawada et al., 2005) and the *Sox9* E1 enhancer (Bagheri-Fam et al., 2006). A notochord and floor plate enhancer was found in the second intron of *Shh* (Epstein et al., 1999; Jeong and Epstein, 2003), but in the context of the broader locus only regulated expression in the floor plate (Jeong et al., 2006). With this set of data, there is not enough commonality across the different CRMs to be able to derive computational predictive models that could be used to search the genome for novel notochord CRMs; this approach has only been successful when a large set of training examples (>70) are available (Narlikar et al., 2010).

Recent studies have shown that ChIP-seq (i.e. chromatin immunoprecipitation (ChIP) combined with massively parallel sequencing) data can be used to successfully predict novel enhancers (Blow et al., 2010; Visel et al., 2009). ChIP-seq protocols are showing potential for enrichment of TF-associated chromatin from very small numbers of cells (Goren et al., 2010; Shankaranarayanan et al., 2011), however these techniques have yet to be proven for the few thousand cells that we have been able to purify from stage-matched embryos. Recently, an *in vitro* cell culture system was described that produced cells similar to embryonic notochord progenitors (Winzi et al., 2011), however further study will be needed to determine their homogeneity and equivalence to *in vivo* populations.

To overcome the limitations associated with studying a small and transient tissue, we hypothesized that binding events of a pioneer TF in an abundant adult tissue, such as liver, may indicate binding events in another tissue where the same TF functions, such as notochord. The pioneer factor Foxa2 is an excellent candidate to address this hypothesis because it is required for notochord formation and function (Ang and Rossant, 1994; Weinstein et al., 1994), as well as the development and homeostasis of many other tissues and organs, such as the liver (reviewed (Friedman and Kaestner, 2006)).

To test the above hypothesis, we first needed to expand the currently limited gene expression data for the early embryonic mouse notochord. We sorted GFP+ notochord progenitor cells from *Noto*$^{eGFP/+}$ embryos (Abdelkhalek et al., 2004; Yamanaka et al., 2007). After mRNA amplification, microarray analysis, and validation using WISH screening, we collected extensive novel expression profiles for the early embryonic mouse notochord. Next, we integrated this expression data with previously published Foxa2 ChIP-seq data for the adult liver (Wederell et al., 2008). Intersection of these two data sets identified a small set of Foxa2-bound regions in liver that were closely associated with genes expressed in the notochord. Lastly, we used a rapid transient reporter assay in zebrafish, which has been well validated as a readout of mammalian enhancer function (Fisher et al., 2006a; Kimura-Yoshida et al., 2004; Navratilova et al., 2009; Suster et al., 2009), to screen candidate notochord CRMs. Remarkably, we were able to identify 7 novel regions that functioned as notochord-specific enhancers and contained Foxa2 binding motifs.

Materials and methods

Mice and embryo dissection

Noto$^{eGFP/+}$ mice have an in-frame knock-in of *eGFP* to the *Noto* locus and were maintained on their original 129S3 (alternate name: 129 Sv/Imj) background (Abdelkhalek et al., 2004). A small percentage of pups born from *Noto*$^{eGFP/+}$ intercrosses produce viable and fertile homozygous null *Noto*$^{eGFP/eGFP}$ animals (Abdelkhalek et al., 2004; Yamanaka et al., 2007). *Noto*$^{eGFP/eGFP}$ male mice, maintained on either a 129S3 or mixed 129S3; ICR background, were crossed to ICR females. Timed pregnant females were dissected at E8.5 and litters of entirely heterozygous *Noto*$^{eGFP/+}$ embryos were obtained, which was confirmed using a fluorescence stereomicroscope. All embryos were dissected in phosphate buffered saline (PBS) with calcium and magnesium. Embryo stages were recorded based on somite number (Fig. 1B; median somite number, n = 6). Embryos were bisected in the transverse plane at the level of somite boundary SI–SII (Fig. 1) using 26G1/2 needles. Posterior regions were pooled in one well of a 4-well dish. The number of embryos used for each of 3 independent biological replicate experiments was as follows: replicate 1, n = 23; replicate 2, n = 39; replicate 3, n = 52. Anterior regions, which also contain a small number of Noto-GFP+ cells, were pooled and processed in parallel for FACS instrument calibration.

Cell sorting

The following protocol was based on the method by Bouchard and colleagues (Bouchard et al., 2005). Excess PBS was removed, 1% trypsin in PBS was added (200 μl), and embryos were incubated for 15′ at 37 °C for dissociation. Cells were pipetted up and down to obtain a mostly single cell suspension, and trypsin was stopped by adding DMEM with 15% FCS (500 μl). Cells were spun down for 5 min at 1500 rpm and the supernatant was removed. The cell pellet was resuspended in sorting media: phenol red-free DMEM; 10% FCS; 1 μg/μl propidium iodide (PI) (300 μl). The cell suspension was passed through a filter cap and collected in a 5 ml polypropylene vial for sorting. Cell sorting was performed using either a BD FACSArial or MoFlo instrument. GFP+ and GFP− cells were sorted from the PI channel and collected directly into Trizol or Trizol LS (Invitrogen). Final ratios of Trizol to cell suspension were 10:1 for Trizol and 3:1 for Trizol LS (the GFP+ sort was a very small volume and was adjusted using RNAse-free water). After sorting, collected cells in Trizol were vortexed for 30 s and stored at −80 °C.

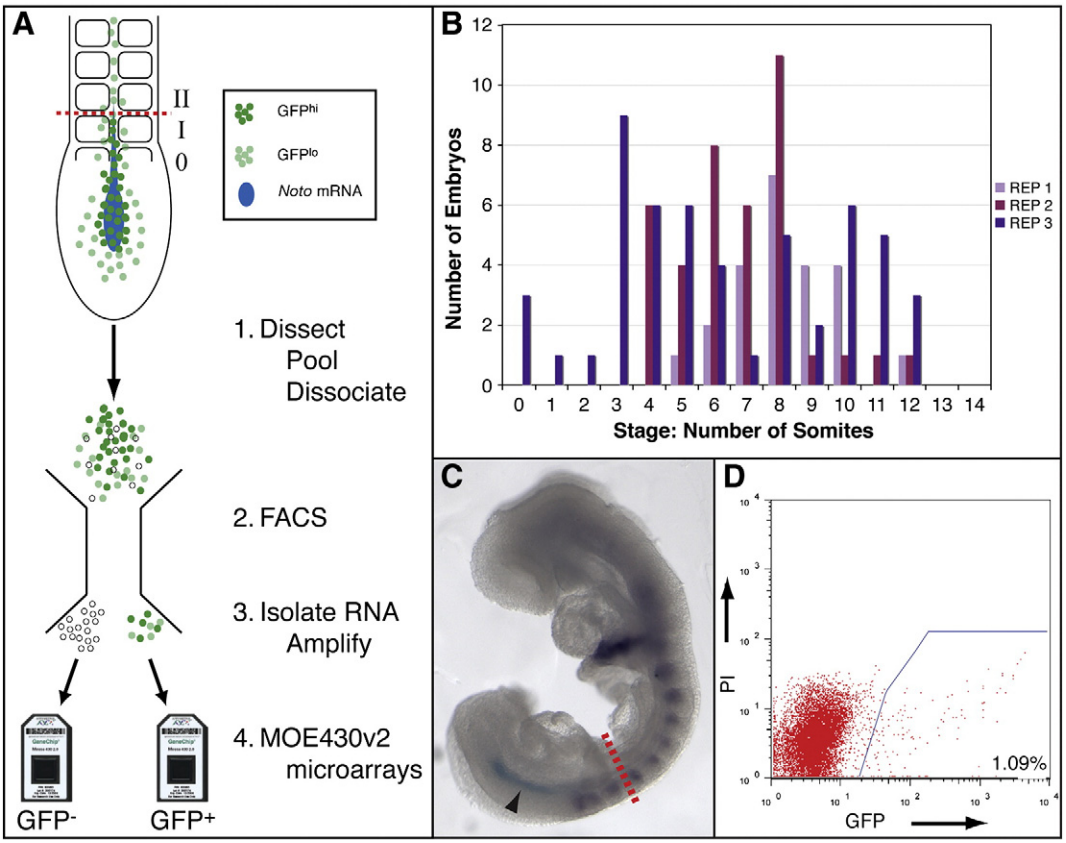

Fig. 1. Summary of screen to analyze gene expression in notochord progenitor cells from the mouse embryo. (A) A diagram summarizing the design of the screen. Early somite stage embryos were bisected at the SI–II boundary. The expression domain of *Noto* mRNA in the notochord is shown in blue. Strongly expressing GFP+ cells (GFP^{hi}) cells are found in the notochord. Weakly expressing GFP+ cells (GFP^{lo}) cells are found in the anterior notochord, paraxial mesoderm, and endoderm. The posterior regions of embryos were pooled, dissociated, and sorted. RNA from both GFP+ and GFP− sorted cells was isolated and amplified for hybridization on Affymetrix MOE430v2 microarrays. (B) Graph showing the staging and numbers of embryos collected for 3 biological replicate experiments: Replicate 1, n = 23; Replicate 2, n = 39; Replicate 3, n = 52. (C) Two-colour WISH to show the expression domain of *Noto* mRNA in relation to the developing somites. *Noto* mRNA is down-regulated in the notochord anterior to the SI–II boundary (dashed red line). *Noto* expression (light blue) in the posterior notochord (arrowhead), in relation to *Tbx18* expression in the somites (purple). Lateral view of a wild-type E8.5 embryo. (D) Representative pre-sort FACS plot of Noto-GFP+ cells isolated from a pool of dissociated posterior E8.5 $Noto^{eGFP/+}$ embryos (PI is propidium iodide to exclude dead cells).

Small-scale RNA isolation

Samples in Trizol were thawed, equilibrated to room temperature, and divided into 800 µl aliquots. 5 µg of Sigma "GenElute" linear polyacrylamide (LPA; Cat. No. 56575-1ML), diluted in DEPC water, was added to each aliquot and vortexed for 30 s. Chloroform extraction and 3 ethanol washes were performed as per the Trizol product insert (Invitrogen protocol Cat. No. 10296-010). The pellet was resuspended in 3 µl RNAse-free water and kept on ice until amplification.

RNA amplification and microarrays

The GFP+ RNA samples were too small for quality control and we proceeded directly to linear amplification. The GFP- RNA samples were more abundant and the quality was confirmed using an Agilent Bioanalyzer. All subsequent steps were as per standard Affymetrix protocols. Samples were prepared using the Two-Cycle Target Labeling kit. Samples were hybridized to Affymetrix MOE430v2 arrays and scanned using a GeneChip Scanner 3000.

Data analysis

Raw data was analyzed using Affymetrix GCOS software (v1.4), and pairwise comparisons were done between Noto-GFP+ and GFP − populations for each of the three biological replicates. We noted that there was good correspondence between replicates of genes with fold change significance calls ($p<0.003$; Table S1; microarray data submitted to NCBI GEO Series: GSE14211). Next we set criteria to generate a list of genes to screen by WISH. We built a list of 50 probe sets that corresponded to 34 known node and notochord marker genes, and then determined their expression values on our array (Table S1). We then applied a z-test (alpha ≤ 0.01) on the replicate fold change call p-values at different significance levels (Fig. 2A). This allowed us to establish a fold enrichment factor for known notochord probe sets within our pool of highly enriched genes. We established a cutoff for WISH screening of Noto-GFP+ enriched genes at an enrichment level greater than or equal to *Brachyury/T* (Fig. 2B; average signal $\log_2$ expression ratio ≥ 2.7; 230 genes total). This was a biologically significant rationale, because *Brachyury/T* is known to be highly expressed in the notochord (i.e. GFP+ sort), but is also expressed at lower levels in the surrounding tailbud mesoderm (i.e. GFP− sort) (Herrmann, 1991).

Whole mount in situ hybridization (WISH) screening

Both one and two-colour WISH was performed as previously described (Yamanaka et al., 2007). Details of clones and methods used to generate riboprobes are included in the supplementary files. WISH screening was automated using an INTAVIS InsituPro VS. Between 4 and 6 wild-type ICR embryos at stage E8.5 were used per basket/well for initial screening. Most positive notochord probes were repeated with embryos staged at E7.75, E8.0, E8.5, and E9.0. Gene expression patterns are representative of 3 or more stage-

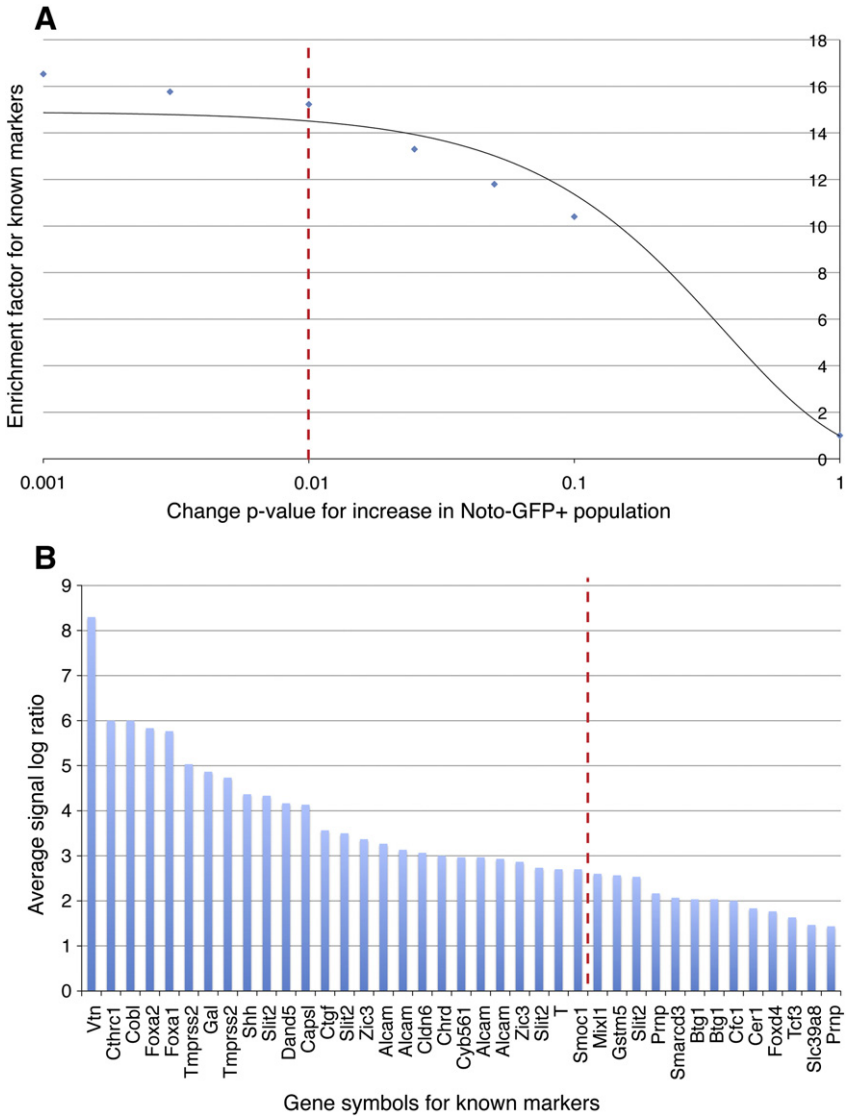

Fig. 2. Thresholds to establish a downstream screen based on the microarray analysis of Noto-GFP+ cells (see Tables S1 and S2 for details). (A) Enrichment factor for known markers expected to be found in Noto-GFP+ cells. Y-axis: (known markers above the threshold/all probe sets above the threshold)/(all known markers/all probe sets on the microarray). X-axis: enrichment factors are plotted at a number of change p-values ($p \leq 0.01$ was used as the threshold for further screening; dashed red line). (B) Known markers are rank ordered by average signal $\log_2$ expression ratio (highest to lowest), and the threshold for further screening was set at the level of *Brachyury*/*T* (average signal $\log_2$ expression ratio ≥ 2.7; dashed red line), which included 230 (redundant) probe sets.Thresholds to establish a downstream screen based on the microarray analysis of Noto-GFP+ cells (see Tables S1 and S2 for details). (A) Enrichment factor for known markers expected to be found in Noto-GFP+ cells. Y-axis: (known markers above the threshold/all probe sets above the threshold)/(all known markers/all probe sets on the microarray). X-axis: enrichment factors are plotted at a number of change p-values ($p \leq 0.01$ was used as the threshold for further screening; dashed red line). (B) Known markers are rank ordered by average signal $\log_2$ expression ratio (highest to lowest), and the threshold for further screening was set at the level of *Brachyury*/*T* (average signal $\log_2$ expression ratio ≥ 2.7; dashed red line), which included 230 (redundant) probe sets.

matched embryos. Imaging was done on a Leica stereo light microscope with Zeiss Axiocam and Axiovision software.

Cloning expression reporter constructs

Constructs were assembled using MultiSite Gateway and vectors from the Tol2kit (Kwan et al., 2007). Enhancers were PCR amplified from mouse genomic DNA using BD Biosciences Advantage 2 Polymerase (all primer sequences are included in Table S8). Our positive controls were amplified using the following primer sets: Sox9 E1 (Bagheri-Fam et al., 2006) forward CCATGCAATCGTGAGGATCT and reverse AGTCCAGGCATTTGGCTATG; Foxa2 mNE (Nishizaki et al., 2001) forward TCATCAGCAGGAAGCAGAGA and reverse TAATTCTGGCCCTTGTGAGG. Enhancer PCR products were directly ligated into a 5′ entry vector (pENTR 5′TOPO TA Cloning Kit from Invitrogen, Cat. No. K591-20), then sequence verified. Conventional cloning was used to generate one enhancer entry clone because of GC-rich regions that prevented PCR; a 2.4 kb EcoRV blunt end fragment from BAC RP23-100A9, which contains the 1.1 kb Cyb561-1 Foxa2 ChIP-seq region, was cloned into the SmaI site of p5E-MCS—a 5′ entry clone that contains the multiple-cloning site from pBluescript (Kwan et al., 2007). The middle entry vector was constructed by PCR amplifying and inserting the minimal promoter from the mouse β-globin gene, previously used for enhancer screening in zebrafish (Woolfe et al., 2005), upstream of GFP (into the NcoI site of Tol2kit #383 pME-EGFP; *NcoI sites* added in primers: β-globin forward primer, GGCACCATGGCCAATCTGCTCAGAGAGGACA; β-globin reverse primer, GGCACCATGGGATGTCTGTTTCTGAGGTTGC). The 3' entry vector was p3E-polyA (Tol2kit #302) and the destination vector was pDestTol2pA2 (Tol2kit #394). The four vectors were assembled using Gateway LR Clonase (Invitrogen, Cat. No. 11791-019) and confirmed by sequencing. An empty vector control to evaluate background expression in the zebrafish expression assay was assembled as above, using a 5′ entry clone with a multiple cloning site in place of an enhancer (Tol2kit #228 p5E-MCS). Assembled Tol2 reporter constructs

were further purified over QIAGEN PCR purification columns, eluted in RNAse/DNAse-free water, and quantified using a Thermo Scientific NanoDrop before injection. Site-directed mutagenesis was performed using the Stratagene QuikChange Multi Site-Directed Mutagenesis Kit (Cat. No. 200515). Mutagenesis primer sequences can be found in Supplementary Methods.

Generation of transient reporter zebrafish embryos

Capped Tol2 transposase mRNA, for co-injection with assembled Tol2 reporter constructs, was synthesized from NotI linearized pCS2FA-transposase vector (Kawakami and Shima, 1999), using the Ambion mMessage mMachine SP6 kit. Transposase mRNA was purified using a QIAGEN RNeasy Mini Kit, eluted in RNAse/DNAse-free water, and quantified using a Thermo Scientific NanoDrop. A detailed protocol of enhancer screening using Tol2 transgenesis in zebrafish has been previously described by Fisher et al. (2006b). Briefly, we injected one-cell stage embryos from crosses of wild-type AB strain zebrafish. We used 15 pg of Tol2 transposase mRNA and 25 pg Tol2 transposon reporter construct DNA. At 28 hpf, we used a Zeiss fluorescent dissecting microscope to score and image live GFP expression in injected embryos.

Results and discussion

A screen to identify co-expressed genes in the developing notochord

Following the method established by Bouchard et al. (2005), we used FACS to isolate GFP-positive notochord progenitors from early somite stage embryos (Figs. 1A, B, D; E8.5; median somite number, n = 6). *Noto* transcript is quickly down-regulated in the anterior notochord as it differentiates, at the level of the most newly formed somite (i.e. the SI–SII boundary; Figs. 1A,C; (Abdelkhalek et al., 2004; Plouhinec et al., 2004)). To enrich our population for notochord progenitors with active *Noto* transcription, we dissected *Noto*$^{eGFP/+}$ embryos at the SI–SII boundary and pooled the posterior tailbud regions before dissociating into a single cell suspension for sorting. We sorted and collected between ~3000 and 6000 Noto-GFP+ cells for each of three biological replicate experiments, which represented ~1.0–1.5% of the sorted cells (Fig. 1D). This was approximately 100 cells per embryo, which is consistent with previous estimates of the cell number in the node (Beddington, 1994). We collected the sorted GFP− cells from the posterior of the embryo as the comparator control population (Fig. 1A).

Affymetrix Two-Cycle linear amplification and GeneChip Mouse Genome 430 2.0 (MOE430v2) arrays were used for expression profile analysis of isolated RNA (Fig. 1A). We established criteria for generating a list of genes to screen by WISH by using a test set of known notochord expressed genes to minimize both false positives and false negatives (type I and type II errors; see Materials and Methods). Our final threshold criteria were a fold change ≥2.7 and a p-value of ≤0.01. At this threshold, we found 76% of genes with published expression patterns were expressed in notochord (n = 29/38; Fig. S1 and Table S2). Although there were known notochord markers expressed at or below the level of *Brachyury/T* (e.g. *Smoc1*, *Cfc1*, *Foxd4*, *Prnp*, *Smarcd3*; Fig. 2B), lowering thresholds to include all known markers would dilute additional novel hits amongst a majority of false positives (type I error). Thus we felt that we had set an optimal threshold for further validation of our data set.

We tested 152 genes that were above the threshold cutoff, and did not have a published E8.5 expression pattern (Fig. S1 and Table S2). Of these, 31% exhibited novel notochord expression patterns (n = 47/152), which was a 10-fold higher discovery rate than any previous similar expression screen for this tissue by others or us (Sousa-Nunes et al., 2003; Tamplin et al., 2008). If known and novel notochord markers were taken together, the percentage of notochord expression patterns among genes that met our thresholds increased even more to 39% (n = 74/190). Representative expression patterns are shown in Fig. 3 and include those in the posterior notochord only (e.g. *Vtn*, 3A; *Bicc1*, 3D; *Dpyd*, 3E; *Moxd1*, 3F), the node crown (e.g. *Cdo1*, 3B; *Upk3a*, 3C), and the entire notochord (e.g. *Ctgf, 3G; Pkd1l1, 3H; Sostdc1, 3I*). Additional gene expression data is noted in Table S2 and images have been deposited in the public EMAGE and MGI GXD databases. Consistent with our prediction, genes with lower enrichment levels that were closer to the *Brachyury/T* cutoff were also expressed more broadly in the posterior embryo. As expected, we also identified a number of paraxial mesoderm and definitive endoderm markers, consistent with a low level of GFP signal in those cell types (Table S2; (Yamanaka et al., 2007)). This screen has greatly increased the number of known embryonic notochord expression patterns in the mouse (64 known + 49 novel = 113 total; Table S7), and has allowed us to search for common CRMs among a large and novel set of genes.

Global analysis of notochord genes

We performed Gene Ontology (GO) analysis to determine if there were significantly enriched terms among the genes identified in our screen that were indicative of notochord function (p ≤ 0.01; Tables S3–S5; GOFFA (Sun et al., 2006)). As expected, some genes associated with terms such as “growth factor binding”, “negative regulation of BMP signaling pathway” and “negative regulation of signal transduction” were identified; these functions are normally associated with the role of the organizer as a source of antagonistic signals (reviewed (Niehrs, 2004)). While our microarray screen did detect known TFs expressed in the notochord (e.g. *T*, *Foxa2*, *Foxa1*; Table S2), we did not find any GO terms annotated to novel notochord genes that would indicate putative TFs. This could suggest that a small set of tissue-specific TFs work together with a larger set of ubiquitously expressed TFs, a model that has been recently proposed (Vaquerizas et al., 2009). Alternatively, there may not be more notochord-specific TFs than those already identified in the literature; if true, then determination of all consensus binding motifs for these notochord-specific TFs could bring us closer to computationally deriving a CRM model for the notochord.

We also searched the MGI Mammalian Phenotype database for terms commonly associated with node and notochord defects (Table S6; (Bult et al., 2010)). As expected, we found known genes among our screen hits with phenotypes related to notochord (*T*, *Foxa2*, *Shh*, *Zic3*, *Chrd*, *Gdf1*), nodal monocilia (*1700027A23Rik*, *Foxa2*, *Bicc1*), and left–right asymmetry (*Dand5*, *1700027A23Rik*, *Zic3*, *Chrd*, *Gdf1*). Encouragingly, during the course of our study, a number of novel node and notochord-related phenotypes were assigned to genes that we identified in our screen; for example, *Pitchfork* (*1700027A23Rik*; (Kinzel et al.)), *Bicc1* (Maisonneuve et al., 2009), and *Pkd1l1* (Vogel et al., 2010). This strongly supports our screen as having identified genes that are not only expressed in the node and notochord, but are also functionally relevant; we hope our screen will provide a rich resource for future phenotypic studies. Having identified broad classes of notochord genes in our screen, we next sought to find CRMs that would connect these genes within a larger regulatory network.

Predicted Foxa2-dependent CRMs Associated with Notochord Genes

As explained above, we could not apply the current repertoire of CRM discovery methods (reviewed (Haeussler and Joly, 2010)) to our newly expanded list of notochord-expressed genes. Instead, to identify CRMs associated with these genes, we decided to focus on Foxa2 as a critical TF near the top of this network (Tamplin et al., 2008). In mouse, Foxa2 is absolutely required for formation of the organizer and its derivatives—the node, notochord, floor plate and

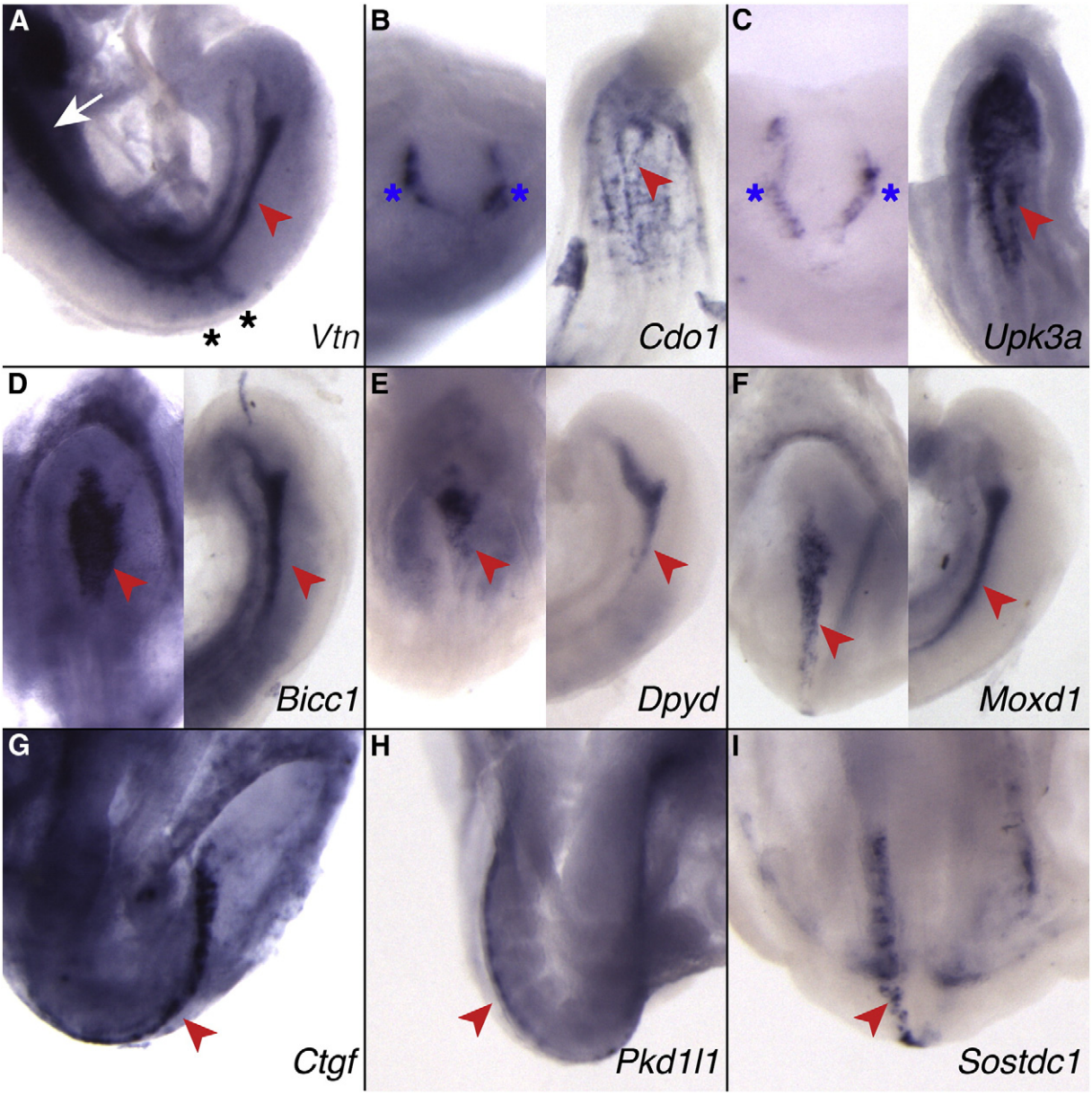

Fig. 3. 9 representative notochord gene expression patterns identified in the screen. *Vtn* (A) was the most highly enriched transcript in Noto-GFP+ cells; it is strongly expressed in the posterior notochord (red arrowhead), and is also expressed in gut endoderm (white arrow) and newly formed somites (black stars). *Cdo1* (B) and *Upk3a* (C) are expressed in the node crown at E8.0 (blue stars; left panels) and posterior notochord (right panels). *Bicc1* (D), *Dpyd* (E), *Moxd1* (F) are expressed in the posterior notochord at E8.0 (left panels) and E9.0 (right panels). *Ctgf* (G), *Pkd1l1* (H), and *Sostdc1* (I) are expressed throughout the notochord. Lateral views of the posterior of E9.0 embryos: (A; D–F, right panels). Ventral views of the posterior of E8.0–E8.5 embryos: (B and C, right panels; D–F, left panels; G–I). Distal view of E8.0 embryos: (B and C, left panels). Red arrowhead: notochord. White arrow: gut endoderm. Black stars: somites. Blue stars: node crown.

definitive endoderm (Ang and Rossant, 1994; Weinstein et al., 1994). Foxa2 is also an important regulator in adult endoderm-derived organs, such as liver, pancreas, and lung (reviewed (Friedman and Kaestner, 2006)). As discussed above, we hypothesized that Foxa2-bound chromatin in an adult tissue, such as the liver, could predict Foxa2-dependent CRMs in a different tissue where Foxa2 is required, such as the notochord.

We looked to a previously published Foxa2 ChIP-seq dataset generated from adult mouse liver tissue as a read-out of high confidence Foxa2 binding events in the genome (Wederell et al., 2008). First, to assign Foxa2 binding events to our candidate notochord genes (n = 113; Fig. 4 and Table S7), we selected a 50 kb window around each transcriptional start site (i.e. 40 kb upstream and 10 kb downstream). Within this total candidate genomic space of 5.65 Mb, we found 56 Foxa2 ChIP-seq binding events from adult liver that were associated with 33 genes. Using available mouse liver microarray data (NCBI GEO Sample: GSM161108; 'present' call using GCOS), we divided our candidates into two classes: 1) genes expressed in the embryonic notochord *and* the adult liver (n = 16 with 31 associated ChIP-seq regions); 2) genes expressed in the embryonic notochord but *not* the adult liver (n = 17 with 25 associated ChIP-seq regions; Fig. 4). The 56 Foxa2 ChIP-seq regions averaged 727 bp in length and ranged from 288 bp to 1485 bp (Table S8).

Next, we classified Foxa2 ChIP-seq regions based on whether or not they contained a predicted Foxa2 binding motif. After searching all 56 candidate regions for at least one predicted Foxa2 binding motif (MotifScanner program (Aerts et al., 2003) using the Foxa2 position weight matrix from JASPAR (Bryne et al., 2008) and MatInspector (Cartharius et al., 2005)), the results were as follows: 9 of 25 regions associated with notochord specific genes, and 21 of 31 regions associated

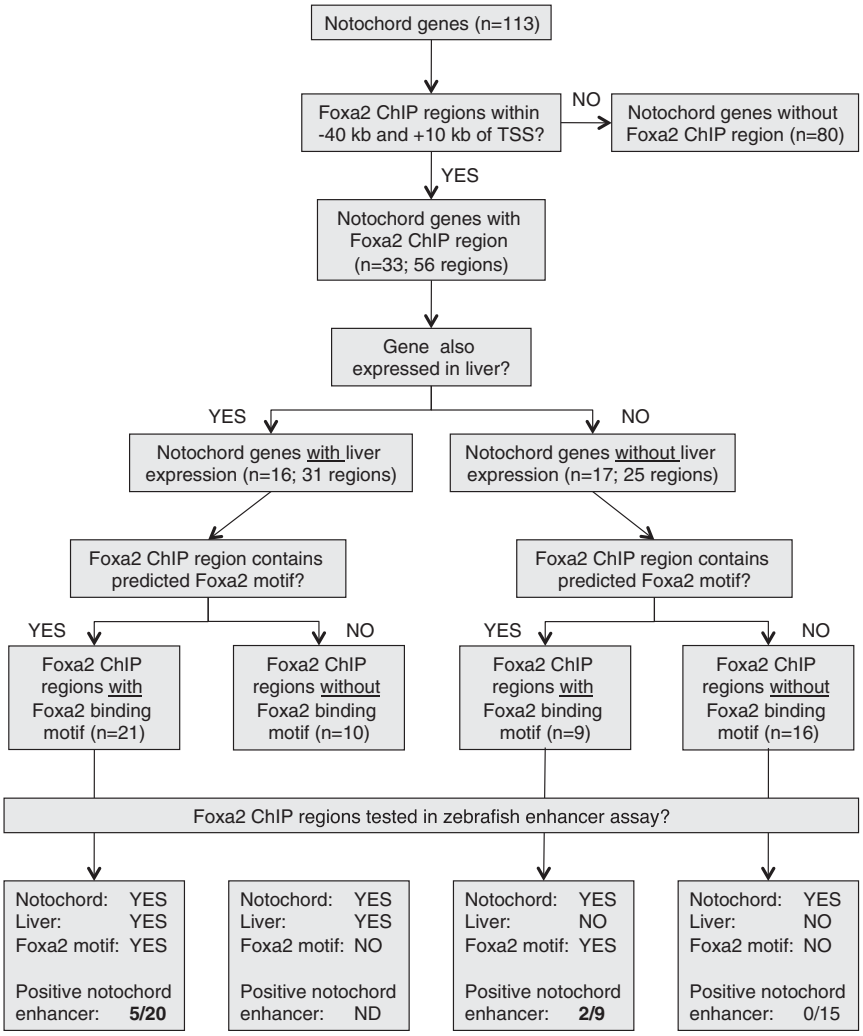

Fig. 4. A flowchart describing classification of notochord genes and their associated Foxa2 ChIP-seq regions from adult liver. Candidate regions tested in the zebrafish transient expression assay were grouped based on presence of a predicted Foxa2 binding motif and expression in the liver.

with notochord and liver expressed genes, contained at least one predicted Foxa2 binding motif. This was consistent with the original analysis of the Foxa2 ChIP-seq data set, which also found a proportion of Foxa2-bound regions that did not contain a predicted Foxa2 binding motif (22%, n = 2494/11,475; (Wederell et al., 2008)). Having applied a series of carefully chosen filters (Fig. 4), we reduced a large ChIP-seq data set (n = 11,475) to a small number of candidate regions (n = 56) that we could test *in vivo* for notochord enhancer function.

Functional validation of predicted notochord CRMs

We evaluated the *in vivo* function of these candidate Foxa2-dependent notochord regulatory regions using an established transient expression assay in zebrafish (Fisher et al., 2006a). Candidate regions from the mouse genome were cloned upstream of a minimal promoter in a GFP reporter vector with *Tol2* transposase sites. DNA reporter constructs were co-injected with *Tol2* transposase RNA into one-cell zebrafish embryos. An "empty" reporter vector control, with no enhancer cloned into the upstream multiple cloning site, produced only a small proportion of embryos with background expression in the notochord (Fig. 5A; 4%, n = 3/74). A positive control, the known *Sox9* E1 notochord enhancer cloned from the mouse genome (Bagheri-Fam et al., 2006) produced robust notochord expression in the zebrafish embryo without significant expression in ectopic cell types (Fig. 5B; 84%, n = 41/49). Similar results were obtained with the mouse notochord enhancer Foxa2 mNE (data not shown; 55%, n = 62/113 (Nishizaki et al., 2001)). We confirmed previous results that showed this zebrafish reporter assay is a robust cross-species assay of mammalian enhancer function (Fisher et al., 2006a; Kimura-Yoshida et al., 2004; Navratilova et al., 2009; Suster et al., 2009).

We then cloned and tested 44 of the 56 candidate regions identified above (Fig. 4 and Table S8), achieving coverage of most sequences with a predicted Foxa2 binding motif, both in the notochord-positive liver-negative set and the notochord-positive liver-positive set. We found 7 novel enhancers that regulated expression in the notochord (Fig. 5 and Table S8), including those with strong and frequent expression in the transient assay that was similar to the Sox9 E1 control (70–100% positive; F5-1 and Pkd1l1-1). We also found enhancers with moderate (40–70% positive; Bicc1-1, Prox1-2, Tgm2-1, and Vtn-2) or weak expression (10–40% positive; Foxa2-1). As expected (Passamaneck et al., 2009; Rastegar et al., 2008), regions that did not contain a predicted Foxa2 binding motif did not regulate notochord expression (n = 0/15, Fig. 4). Our filtering criteria above selected 56 candidate CRMs within a total of 5.65 Mb of the mouse genome. After *in vivo* testing of 44 of these regions in a zebrafish reporter assay, 7 regions tested positive for expression in the notochord, a positive hit rate of 16% that is on par with previous notochord enhancer screens (Dickmeis et al., 2004; Rastegar et al., 2008).

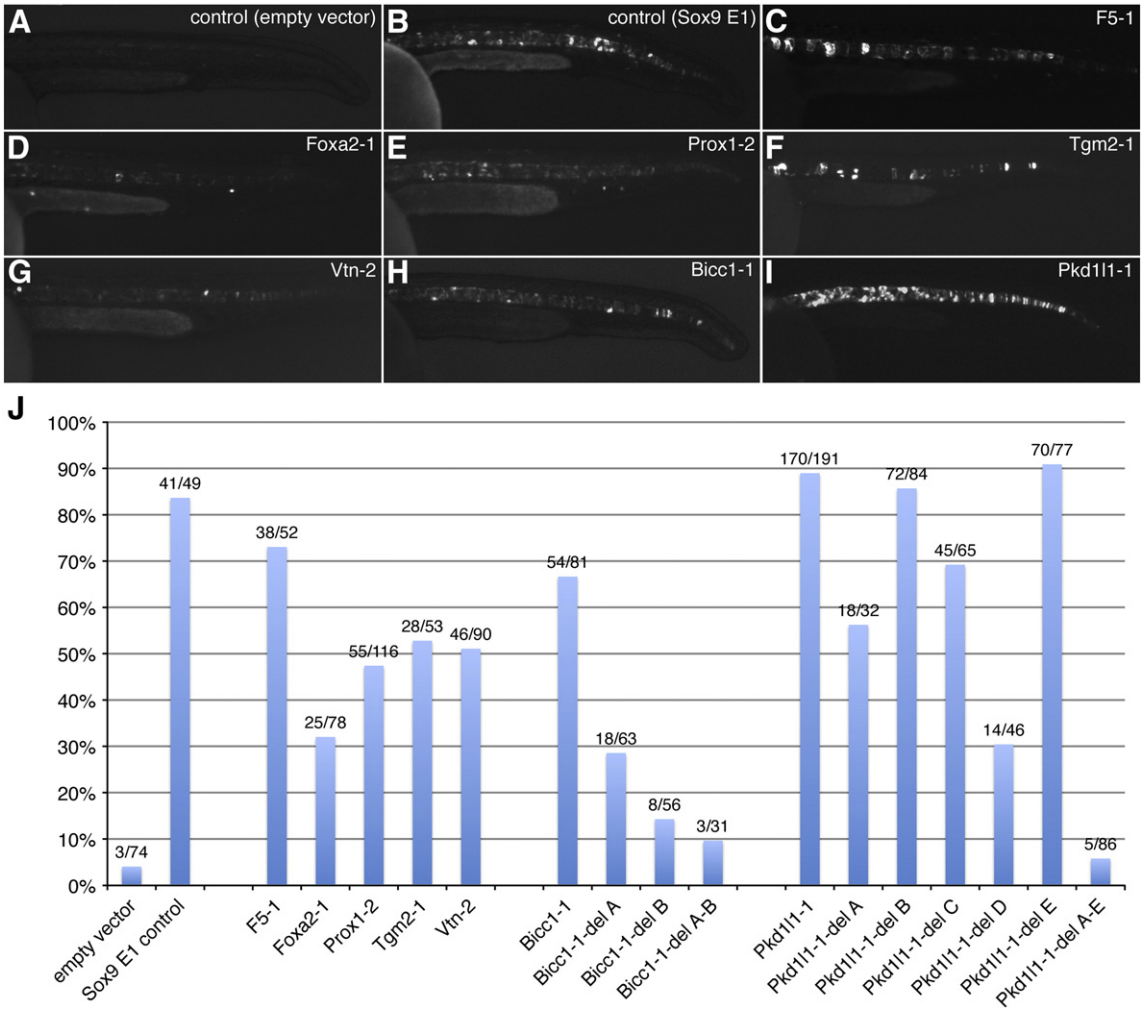

Fig. 5. Transient reporter assay in zebrafish to identify notochord enhancers. (A) Injection of an empty vector control construct with no upstream enhancer inserted produced only a few embryos with GFP positive cells in the notochord; this was considered the background level. (B) The control *Sox9* E1 enhancer from mouse (Bagheri-Fam et al., 2006) drives robust and specific notochord expression in the zebrafish embryo. Among the novel enhancers, we observed strong (70–100% positive; (C) F5-1 and (I) Pkd1l1-1), moderate (40–70% positive; (E) Prox1-2, (F) Tgm2-1, (G) Vtn-2, and (H) Bicc1-1), and weak expression levels (10–40% positive; (D) Foxa2-1). These levels were based on both the frequency of notochord-positive embryos amongst total injected (J), as well as the GFP signal intensity and mosaicism within positive embryos (C–I). Site-directed mutagenesis of Foxa2 binding motifs in the Pkd1l1-1 and Bicc1-1 enhancers is shown next to their wild-type counterparts (J). Note: all embryos shown are ~28 hpf, lateral views, anterior to the left. Data points in the chart (J) are labeled with the total number of notochord-positive embryos over the total number of embryos scored.

Motif content within the Bicc1-1 and Pkd1l1-1 notochord CRMs

Recent studies in zebrafish and ascidian showed a requirement for FoxA TF binding in notochord CRMs (Passamaneck et al., 2009; Rastegar et al., 2008). While Foxa2 alone was not sufficient to drive notochord expression in the zebrafish, it was at least necessary (Rastegar et al., 2008). All 7 of novel notochord enhancers we discovered were classified as Foxa2-bound regions in the liver that contained predicted Foxa2-binding motifs (Fig. 4). Accordingly, we reasoned that site-directed mutagenesis of these Foxa2 motifs should diminish overall enhancer function. We focused on the Bicc1-1 and Pkd1l1-1 enhancers because of their unique property of having bound Foxa2 in the adult liver, but rather than being expressed in that tissue, are expressed specifically in the embryonic notochord. We reasoned that these enhancers should have a high degree of specificity for the embryonic notochord. We mutated the two predicted Foxa2 motifs in the Bicc1-1 enhancer and five in the Pkd1l1-1 enhancer, both individually and together (Fig. 5J, Table S8 and Supplementary Methods). The core sequence of the Foxa2 motif is TTT or AAA, depending on the strand, and these sites were mutated to CCC or GGG, respectively.

Individual mutation of predicted Foxa2 binding motifs in the Bicc1-1 enhancer decreased reporter expression significantly (A or B gave 40–50% reduction in the number of notochord-positive embryos), whereas mutation of both motifs together (A and B) reduced expression to near background levels (Fig. 5J). The five Foxa2 motifs in the Pkd1l1-1 enhancer were mutated individually (A, B, C, D, or E); motif B and E had little to no effect on enhancer expression, however the frequency of notochord expression was reduced 20–60% when motifs A, C or D were mutated, with motif D having the strongest effect (Fig. 5J). Mutation of all 5 Foxa2 motifs together in the Pkd1l1-1 enhancer (A to E) resulted in near background levels of notochord expression. Our results are consistent with previous results in ascidian and zebrafish that demonstrated activation of notochord enhancers requires functional Foxa2 binding sites (Passamaneck et al., 2009; Rastegar et al., 2008).

Conservation of Foxa2-dependent notochord enhancers

As more genome-wide ChIP data is generated, it is becoming apparent that even within closely related mammals there is a minority of TF-bound sequences that are actually conserved (Odom et al.,

2007). For CRMs to retain their function during evolution as sequences diverge and motifs turn over, it appears that total binding motif content, regardless of position or orientation, are the most important criteria (Schmidt et al., 2010). The Bicc1-1 and Pkd1l1 mouse enhancers have retained some sequence conservation with humans (Fig. 6 and Supplementary Methods), however no aligned sequence could be found in more distant vertebrates, such as the zebrafish. Yet both fragments from the mouse genome very specifically regulated notochord expression in the zebrafish; this is consistent with previous observations that mammalian enhancers lacking extreme sequence conservation still function in the zebrafish (Fisher et al., 2006a). Even though the notochord is the defining feature of chordate body plan, the enhancers that regulate expression in this ancient tissue have not necessarily retained sequence level conservation across distantly related species. This highlights the need to find criteria other than sequence conservation alone to search for novel CRMs.

Significance of TF binding specificity in different tissues and species

Before general conclusions can be drawn regarding the binding profile of pioneer TFs, many more data sets will need to be collected and compared. Given the broad expression and function of Foxa2 (reviewed (Friedman and Kaestner, 2006)), it will be interesting to explore the overlap of Foxa2 binding events between different adult tissues (e.g. liver, pancreas, and lung), as well as other embryonic tissues (e.g. floor plate and endoderm). It would also be useful to collect and compare data for other TFs that have been described as having pioneer function, such as GATA4 (Cirillo et al., 2002). While our 7 examples of Foxa2-dependent notochord enhancers identified from adult liver Foxa2 ChIP-seq data suggest a degree of pioneer factor binding promiscuity, our approach was still not a general predictor of mouse midline enhancers; for example, the *Foxa2* mNE and *Sox9* E1 notochord CRMs, and the intronic Foxa2-dependent *Shh* floor plate CRM (Jeong and Epstein, 2003; Jeong et al., 2006), were not associated

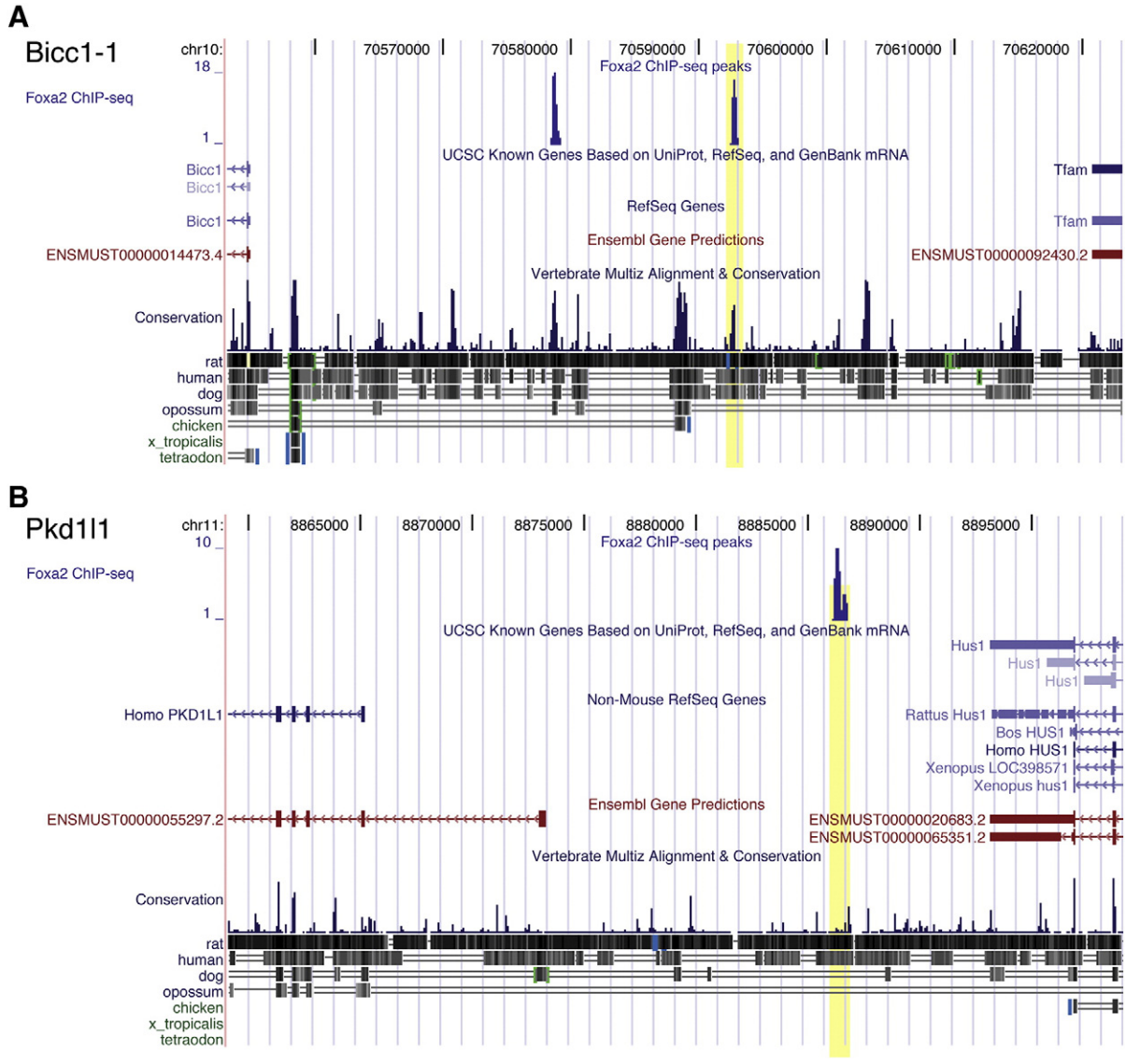

Fig. 6. UCSC Genome Browser views of Foxa2 ChIP-seq regions associated with notochord genes *Bicc1* and *Pkd1l1*. (A) A 70 kb genome browser window between *Bicc1* and its upstream neighbour (*Tfam*). There are 2 Foxa2 ChIP-seq regions, both of which are conserved in mammals. The more distal region, −38 kb from the transcriptional start site of *Bicc1*, regulated expression in the notochord (highlighted). (B) A 40 kb genome browser window between *Pkd1l1* and its upstream neighbour (*Hus1*). There is a single Foxa2 ChIP-seq −15 kb upstream of the *Pkd1l1* transcriptional start site (according to Ensembl mapping of *Pkd1l1* transcript ENSMUST00000055297). This region was positive for notochord expression in the zebrafish transient expression assay (highlighted). Note: UCSC genome build mm8 used for graphics; peaks were converted to mm9 for all other analysis; UCSC Genome Browser link and reference: http://genome.ucsc.edu/ (Kent et al., 2002).

with a Foxa2 ChIP-seq region in liver (data not shown). The observation that only a subset of notochord enhancers are bound by Foxa2 in the adult liver suggests that other levels of regulation, such as co-factors, repressors and chromatin modifications, will be needed to complete the genetic network for the notochord.

Conclusions

In this study we sought to expand our knowledge of the gene regulatory network that regulates formation and function of the embryonic mouse notochord. After collecting extensive expression data for highly purified notochord progenitor cells, we identified CRMs associated with our newly identified notochord-specific genes. We used a novel proxy method for CRM discovery—association of pioneer TF binding events in an abundant adult tissue with genes expressed in a rare embryonic tissue. Our results support a model with more TF binding events in a tissue than are required to regulate gene expression—an important consideration given the amount of genome-wide TF binding data that is becoming available. Our data represents a significant resource for further investigation of notochord development.

Supplementary materials related to this article can be found online at doi:10.1016/j.ydbio.2011.10.002.

Acknowledgments

Expert animal care was provided by Jorge Cabezas (Mount Sinai Hospital/Toronto Centre for Phenogenomics). Maxime Bouchard and David Grote (McGill University) offered advice regarding isolation of cells from embryos using FACS and linear amplification of RNA. Achim Gossler and Anja Beckers (Institute for Molecular Biology, Hannover) generated the *Noto*$^{eGFP/+}$ knock-in mouse. Sheyun Zhao and Pier-Andrée Penttilä (SickKids-UHN Flow Cytometry Facility) performed cell sorting. Chao Lu & Jane Yan Liu (The Centre for Applied Genomics, SickKids) performed RNA amplification and microarray hybridization. Pamela Hoodless and Marco Marra (BC Cancer Agency) shared their liver Foxa2 ChIP-Seq data set. Chi-bin Chien and Kristen Kwan (University of Utah) offered advice regarding application of the Tol2kit. Brian Ciruna and Ian Scott (SickKids/University of Toronto) provided advice in the design of zebrafish experiments. Brian Ciruna, Ian Scott, and Danielle Gelinas provided training, support and access to the SickKids zebrafish facility. We thank Brian Ciruna, Pamela Hoodless, and Achim Gossler for helpful comments on the manuscript. This work was supported by Canadian Institutes of Health Research (CIHR) grant (MOP-13426). O.J.T. was supported by a CIHR fellowship.

References

Abdelkhalek, H.B., Beckers, A., Schuster-Gossler, K., Pavlova, M.N., Burkhardt, H., Lickert, H., Rossant, J., Reinhardt, R., Schalkwyk, L.C., Muller, I., Herrmann, B.G., Ceolin, M., Rivera-Pomar, R., Gossler, A., 2004. The mouse homeobox gene Not is required for caudal notochord development and affected by the truncate mutation. Genes Dev. 18, 1725–1736.

Aerts, S., Thijs, G., Coessens, B., Staes, M., Moreau, Y., De Moor, B., 2003. Toucan: deciphering the cis-regulatory logic of coregulated genes. Nucleic Acids Res. 31, 1753–1764.

Ang, S.L., Rossant, J., 1994. HNF-3 beta is essential for node and notochord formation in mouse development. Cell 78, 561–574.

Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Gunther, T., Schule, R., Kemler, R., Mallo, M., Kanzler, B., Scherer, G., 2006. Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. Dev. Biol. 291, 382–397.

Barrionuevo, F., Taketo, M.M., Scherer, G., Kispert, A., 2006. Sox9 is required for notochord maintenance in mice. Dev. Biol. 295, 128–140.

Beckers, A., Alten, L., Viebahn, C., Andre, P., Gossler, A., 2007. The mouse homeobox gene Noto regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. Proc. Natl. Acad. Sci. U. S. A. 104, 15765–15770.

Beddington, R.S., 1994. Induction of a second neural axis by the mouse node. Development 120, 613–620.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Bristow, J., Ren, B., Black, B.L., Rubin, E.M., Visel, A., Pennacchio, L.A., 2010. ChIP-Seq identification of weakly conserved heart enhancers. Nat. Genet. 42, 806–810.

Bonnafe, E., Touka, M., AitLounis, A., Baas, D., Barras, E., Ucla, C., Moreau, A., Flamant, F., Dubruille, R., Couble, P., Collignon, J., Durand, B., Reith, W., 2004. The transcription factor RFX3 directs nodal cilium development and left–right asymmetry specification. Mol. Cell. Biol. 24, 4417–4427.

Bouchard, M., Grote, D., Craven, S.E., Sun, Q., Steinlein, P., Busslinger, M., 2005. Identification of Pax2-regulated genes by expression profiling of the mid-hindbrain organizer region. Development 132, 2633–2643.

Bryne, J.C., Valen, E., Tang, M.H., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., Sandelin, A., 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. 36, D102–6.

Bult, C.J., Kadin, J.A., Richardson, J.E., Blake, J.A., Eppig, J.T., 2010. The Mouse Genome Database: enhancements and updates. Nucleic Acids Res. 38, D586–D592.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., Werner, T., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21, 2933–2942.

Choi, K.S., Cohn, M.J., 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.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K.S., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell 9, 279–289.

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

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A genomic regulatory network for development. Science 295, 1669–1678.

Dickmeis, T., Plessy, C., Rastegar, S., Aanstad, P., Herwig, R., Chalmel, F., Fischer, N., Strahle, U., 2004. Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo. Genome Res. 14, 228–238.

Elms, P., Scurry, A., Davies, J., Willoughby, C., Hacker, T., Bogani, D., Arkell, R., 2004. Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. Gene Expr. Patterns 4, 505–511.

Epstein, D.J., McMahon, A.P., Joyner, A.L., 1999. Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. Development 126, 281–292.

Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., McCallion, A.S., 2006a. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312, 276–279.

Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., McCallion, A.S., 2006b. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat. Protoc. 1, 1297–1305.

Frankenberg, S., Smith, L., Greenfield, A., Zernicka-Goetz, M., 2007. Novel gene expression patterns along the proximo-distal axis of the mouse embryo before gastrulation. BMC Dev. Biol. 7, 8.

Friedman, J.R., Kaestner, K.H., 2006. The Foxa family of transcription factors in development and metabolism. Cell. Mol. Life Sci. 63, 2317–2328.

Goren, A., Ozsolak, F., Shoresh, N., Ku, M., Adli, M., Hart, C., Gymrek, M., Zuk, O., Regev, A., Milos, P.M., Bernstein, B.E., 2010. Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. Nat. Methods 7, 47–49.

Haeussler, M., Joly, J.S., 2010. When needles look like hay: how to find tissue-specific enhancers in model organism genomes. Dev. Biol. 350, 239–254.

Herrmann, B.G., 1991. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. Development 113, 913–917.

Jeong, Y., Epstein, D.J., 2003. Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. Development 130, 3891–3902.

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., Epstein, D.J., 2006. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development 133, 761–772.

Kawakami, K., Shima, A., 1999. Identification of the Tol2 transposase of the medaka fish *Oryzias latipes* that catalyzes excision of a nonautonomous Tol2 element in zebrafish *Danio rerio*. Gene 240, 239–244.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.

Kim, Y.S., Nakanishi, G., Lewandoski, M., Jetten, A.M., 2003. GLIS3, a novel member of the GLIS subfamily of Kruppel-like zinc finger proteins with repressor and activation functions. Nucleic Acids Res. 31, 5513–5525.

Kimura-Yoshida, C., Kitajima, K., Oda-Ishii, I., Tian, E., Suzuki, M., Yamamoto, M., Suzuki, T., Kobayashi, M., Aizawa, S., Matsuo, I., 2004. Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. Development 131, 57–71.

Kinder, S.J., Tsang, T.E., Wakamiya, M., Sasaki, H., Behringer, R.R., Nagy, A., Tam, P.P., 2001. The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. Development 128, 3623–3634.

Kinzel, D., Boldt, K., Davis, E.E., Burtscher, I., Trumbach, D., Diplas, B., Attie-Bitach, T., Wurst, W., Katsanis, N., Ueffing, M., Lickert, H., 2010. Pitchfork regulates primary cilia disassembly and left-right asymmetry. Dev. Cell 19, 66–77.

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., Chien, C.B., 2007. The Tol2kit: a multisite

gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev. Dyn. 236, 3088–3099.
Maisonneuve, C., Guilleret, I., Vick, P., Weber, T., Andre, P., Beyer, T., Blum, M., Constam, D.B., 2009. Bicaudal C, a novel regulator of Dvl signaling abutting RNA-processing bodies, controls cilia orientation and leftward flow. Development 136, 3019–3030.
Matsumoto, J., Kumano, G., Nishida, H., 2007. Direct activation by Ets and Zic is required for initial expression of the Brachyury gene in the ascidian notochord. Dev. Biol. 306, 870–882.
Muller, F., Chang, B., Albert, S., Fischer, N., Tora, L., Strahle, U., 1999. Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. Development 126, 2103–2116.
Narlikar, L., Sakabe, N.J., Blanski, A.A., Arimura, F.E., Westlund, J.M., Nobrega, M.A., Ovcharenko, I., 2010. Genome-wide discovery of human heart enhancers. Genome Res. 20, 381–392.
Navratilova, P., Fredman, D., Hawkins, T.A., Turner, K., Lenhard, B., Becker, T.S., 2009. Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. Dev. Biol. 327, 526–540.
Niehrs, C., 2004. Regionally specific induction by the Spemann–Mangold organizer. Nat. Rev. Genet. 5, 425–434.
Nishizaki, Y., Shimazu, K., Kondoh, H., Sasaki, H., 2001. Identification of essential sequence motifs in of Foxa2 (Hnf3beta) gene that are conserved. Mech. Dev. 102, 57–66.
Nobrega, M.A., Ovcharenko, I., Afzal, V., Rubin, E.M., 2003. Scanning human gene deserts for long-range enhancers. Science 302, 413.
Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., Fraenkel, E., 2007. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat. Genet. 39, 730–732.
Passamaneck, Y.J., Katikala, L., Perrone, L., Dunn, M.P., Oda-Ishii, I., 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.
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B.L., Couronne, O., Eisen, M.B., Visel, A., Rubin, E.M., 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–502.
Plouhinec, J.L., Granier, C., Le Mentec, C., Lawson, K.A., Saberan-Djoneidi, D., Aghion, J., Shi, D.L., Collignon, J., Mazan, S., 2004. Identification of the mammalian Not gene via a phylogenomic approach. Gene Expr. Patterns 5, 11–22.
Rastegar, S., Hess, I., Dickmeis, T., Nicod, J.C., Ertzer, R., Hadzhiev, Y., Thies, W.G., Scherer, G., Strahle, U., 2008. The words of the regulatory code are arranged in a variable manner in highly conserved enhancers. Dev. Biol. 318, 366–377.
Robb, L., Tam, P.P., 2004. Gastrula organiser and embryonic patterning in the mouse. Semin. Cell Dev. Biol. 15, 543–554.
Sasaki, H., Hogan, B.L., 1996. Enhancer analysis of the mouse HNF-3 beta gene: regulatory elements for node/notochord and floor plate are independent and consist of multiple sub-elements. Genes Cells 1, 59–72.
Sawada, A., Nishizaki, Y., Sato, H., Yada, Y., Nakayama, R., Yamamoto, S., Nishioka, N., Kondoh, H., Sasaki, H., 2005. Tead proteins activate the Foxa2 enhancer in the node in cooperation with a second factor. Development 132, 4719–4729.
Sawada, A., Kiyonari, H., Ukita, K., Nishioka, N., Imuta, Y., Sasaki, H., 2008. Redundant roles of Tead1 and Tead2 in notochord development and the regulation of cell proliferation and survival. Mol. Cell. Biol. 28, 3177–3189.
Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., Talianidis, I., Flicek, P., Odom, D.T., 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328, 1036–1040.
Shankaranarayanan, P., Mendoza-Parra, M.A., Walia, M., Wang, L., Li, N., Trindade, L.M., Gronemeyer, H., 2011. Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. Nat. Methods 8, 565–567.
Sherwood, R.I., Jitianu, C., Cleaver, O., Shaywitz, D.A., Lamenzo, J.O., Chen, A.E., Golub, T.R., Melton, D.A., 2007. Prospective isolation and global gene expression analysis of definitive and visceral endoderm. Dev. Biol. 304, 541–555.
Shiratori, H., Hamada, H., 2006. The left–right axis in the mouse: from origin to morphology. Development 133, 2095–2104.
Sousa-Nunes, R., Rana, A.A., Kettleborough, R., Brickman, J.M., Clements, M., Forrest, A., Grimmond, S., Avner, P., Smith, J.C., Dunwoodie, S.L., Beddington, R.S., 2003. Characterizing embryonic gene expression patterns in the mouse using nonredundant sequence-based selection. Genome Res. 13, 2609–2620.
Stemple, D.L., 2005. Structure and function of the notochord: an essential organ for chordate development. Development 132, 2503–2512.
Sulik, K., Dehart, D.B., Iangaki, T., Carson, J.L., Vrablic, T., Gesteland, K., Schoenwolf, G.C., 1994. Morphogenesis of the murine node and notochordal plate. Dev. Dyn. 201, 260–278.
Sun, H., Fang, H., Chen, T., Perkins, R., Tong, W., 2006. GOFFA: gene ontology for functional analysis—a FDA gene ontology tool for analysis of genomic and proteomic data. BMC Bioinformatics 7 (Suppl. 2), S23.
Suster, M.L., Kania, A., Liao, M., Asakawa, K., Charron, F., Kawakami, K., Drapeau, P., 2009. A novel conserved evx1 enhancer links spinal interneuron morphology and cis-regulation from fish to mammals. Dev. Biol. 325, 422–433.
Tamplin, O.J., Kinzel, D., Cox, B.J., Bell, C.E., Rossant, J., Lickert, H., 2008. Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. BMC Genomics 9, 511.
Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., Luscombe, N.M., 2009. A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 10, 252–263.
Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., Pennacchio, L.A., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858.
Vogel, P., Read, R., Hansen, G.M., Freay, L.C., Zambrowicz, B.P., Sands, A.T., 2010. Situs inversus in Dpcd/Poll−/−, Nme7−/−, and Pkd1l1−/− mice. Vet. Pathol. 47, 120–131.
Wederell, E.D., Bilenky, M., Cullum, R., Thiessen, N., Dagpinar, M., Delaney, A., Varhol, R., Zhao, Y., Zeng, T., Bernier, B., Ingham, M., Hirst, M., Robertson, G., Marra, M.A., Jones, S., Hoodless, P.A., 2008. Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. Nucleic Acids Res. 36, 4549–4564.
Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., Darnell Jr., J.E., 1994. The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell 78, 575–588.
Wilkinson, D.G., Bhatt, S., Herrmann, B.G., 1990. Expression pattern of the mouse T gene and its role in mesoderm formation. Nature 343, 657–659.
Winzi, M.K., Hyttel, P., Dale, J.K., Serup, P., 2011. Isolation and characterization of node/notochord-like cells from mouse embryonic stem cells. Stem Cells Dev. 2011 Apr 6. [Electronic publication ahead of print].
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3, e7.
Yamanaka, Y., Tamplin, O.J., Beckers, A., Gossler, A., Rossant, J., 2007. Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. Dev. Cell 13, 884–896.