Published in final edited form as:

Genesis. 2008 November ; 46(11): 683–696. doi:10.1002/dvg.20406.

Conservation of Notochord Gene Expression Across Chordates: Insights From the *Leprecan* Gene Family

Terence D. Capellini, Matthew P. Dunn, Yale J. Passamaneck, Licia Selleri, and Anna Di Gregorio*

Department of Cell and Developmental Biology, Weill Medical College of Cornell University, New York, New York

Summary

The notochord is a defining character of the chordates, and the T-box transcription factor Brachyury has been shown to be required for notochord development in all chordates examined. In the ascidian *Ciona intestinalis*, at least 44 notochord genes have been identified as bona fide transcriptional targets of Brachyury. We examined the embryonic expression of a subset of murine orthologs of *Ciona* Brachyury target genes in the notochord to assess its conservation throughout chordate evolution. We focused on analyzing the *Leprecan* gene family, which in mouse is composed of three genes, as opposed to the single-copy *Ciona* gene. We found that all three mouse *Leprecan* genes are expressed in the notochord. Additionally, while *Leprecan* expression in *C. intestinalis* is confined to the notochord, expression of its mouse orthologs includes dorsal root ganglia, limb buds, branchial arches, and developing kidneys. These results have interesting implications for the evolution and development of chordates.

Keywords

Ciona; *Brachyury*; notochord; ascidian; leprecan; chordates; mouse; prolyl 3-hydroxylase

INTRODUCTION

One of the defining features of the chordate clade is the presence of the notochord during embryonic development. The notochord, an axial organ of mesodermal origin, provides structural support to the developing embryo (Adams *et al.*, 1990; Stemple, 2005) and, in vertebrates, patterning signals to the adjacent neural tube and surrounding tissues (Cleaver and Krieg, 2001; Copp *et al.*, 2003). In invertebrate chordates (i.e., cephalochordates and urochordates), the notochord is the only axial structure sustaining the embryo; however, in vertebrates, it additionally serves as a precursor to the vertebral column, which gradually replaces it during development (Fleming *et al.*, 2001). Furthermore, the vertebrate notochord acts as an organizer on its neighboring tissues, inducing the floor-plate of the neural tube (Placzek *et al.*, 1990) and influencing the development of both endodermal (e.g., liver and pancreas) and mesodermal (e.g., heart and blood vessels) organs (Cleaver and Krieg, 2001; Fouquet *et al.*, 1997).

© 2008 Wiley-Liss, Inc.

* Correspondence to: Anna Di Gregorio, Department of Cell and Developmental Biology, Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10065, USA. and2015@med.cornell.edu.

Terence D. Capellini and Matthew P. Dunn contributed equally to this work.

Additional Supporting Information for this article may be found in the online version of this article.

Of the genes expressed during notochord development, *Brachyury* (*Bra* or *T*) has been shown to be essential for notochord formation in all chordates so far examined, including ascidians, zebrafish, and mice (Di Gregorio *et al.*, 2002; Schulte-Merker *et al.*, 1994; Wilson *et al.*, 1995). In the tunicate ascidian *Ciona intestinalis*, *Brachyury* (*Ci-Bra*) is expressed exclusively in the notochord, beginning at the 64-cell stage and persisting until the late tailbud stage (Corbo *et al.*, 1997). In contrast, *Bra* is first expressed pan-mesodermally early in mouse development, and only later, after the separation of axial and paraxial lineages, is its expression restricted to the notochord and the tailbud (Wilkinson *et al.*, 1990; Wilson *et al.*, 1993).

The Brachyury protein is an evolutionarily conserved T-box transcription factor, which specifically binds sequences found in the regulatory regions of its target genes (Casey *et al.*, 1998; Di Gregorio and Levine, 1999; Kispert *et al.*, 1995). Brachyury has been shown to control expression of the homeodomain transcription factor gene *Not* in mouse embryos (Ben Abdelkhalek *et al.*, 2004), and of the *Bix*, *FGF*, and *Xwnt11* genes in the frog embryo (Casey *et al.*, 1998, 1999; Tada and Smith, 2000; Tada *et al.*, 1998), but still little is known on the identity of structural notochord genes targeted by this transcription factor in vertebrates. In *C. intestinalis*, a subtractive hybridization screen between wild-type tailbud embryos and embryos ectopically expressing *Ci-Bra* identified 44 bona fide *Ci-Bra*-downstream genes, which are predominantly expressed in notochord cells (Hotta *et al.*, 1999, 2000; Takahashi *et al.*, 1999). We have identified *in silico* putative mouse orthologs of representative *Ci-Bra* target genes to examine whether their expression is also seen in the mouse notochord. Although these mouse genes had been previously identified, their expression in the notochord had not yet been assessed. We show that most of the evolutionarily conserved genes we analyzed are expressed in the notochord and that, differently from their notochord-specific ascidian counterparts, they are also frequently expressed in the mouse neural tube. We then examined in detail the expression patterns of three related genes, mouse *Leprecan*, *Leprecan-like1* (*Leprel1*), and *Leprecan-like2* (*Leprel2*), since they were still relatively uncharacterized in vertebrates, and we found that they were all expressed not only in the mouse notochord, but also in additional mesodermal and non-mesodermal derivatives.

We report an analysis of the embryonic expression of the mouse *Leprecan* genes throughout mid-gestation, and discuss how their transcriptional regulation might have changed during chordate evolution.

RESULTS

Expression of Putative Mouse Orthologs of *Ci-Bra* Target Genes

We chose a representative subset of mouse orthologs of *Ci-Bra* target genes (Takahashi *et al.*, 1999) to examine during mouse development. To ensure that we examined the expression of appropriate members of multi-gene families, we endeavored to identify the mouse ortholog(s) of each single-copy *Ciona* gene of interest. As BLAST search results yielded multiple similar hits in the mouse genome for each of the *Ciona* genes of interest, a simple one-to-one orthologous relationship was not readily apparent, as expected, because of general duplications in vertebrates. Therefore, we performed multiple sequence alignments followed by phylogenetic analyses to identify the closest orthologs from among the sometimes numerous murine sequences (data not shown). Once these sequences were identified, the corresponding cDNA clones were acquired from the IMAGE collection (Lennon *et al.*, 1996; Table S1) and used to generate digoxigenin-labeled antisense RNA probes, which were used for either whole-mount or section in situ hybridization experiments. Notochord expression was initially investigated by in situ hybridization on

transverse sections of gestational day E10.5 in embryos derived from C57/BL6 WT mice (Fig. 1).

Of the 12 putative Brachyury target genes examined, nine turned out to be expressed in the notochord. Specifically, strong notochord expression was identified for *Agrin*, *Papss1*, *Pellino1*, *Pellino2*, *SerpinC1*, *SerpinG1* (Fig. 1), and the three *Leprecan* genes (see below). In contrast, no notochord expression was observed for *Acly* (data not shown), *Fgl2* (Fig. S1), and *Papss2* (data not shown) at the stages analyzed. The vast majority of the genes examined were expressed in additional domains adjacent to the notochord, most notably the floor-plate of the developing neural tube (Fig. 1). For example, *Fgl2* (Fig. S1) and *Cdc45l* (data not shown) were expressed in the floor-plate, *Agrin* (Fig. 1) and *Acly* (data not shown) were expressed in the lateral ventral horns, and all other genes were expressed throughout the neural tube (Fig. 1). A detailed view of the staining in floor-plate and/or notochord for the genes *Leprecan* and *Fgl2*, as compared with the staining obtained for *Sonic hedgehog* (*Shh*), which labels both structures (Echelard *et al.*, 1993), is shown in Figure S1. Finally, we found that both *Serpin* genes analyzed were expressed in the lateral plate mesoderm, in addition to their axial expression domains (data not shown).

The *Leprecan* Gene Family

Based on our initial analysis of expression in the mouse notochord, we chose to examine the spatiotemporal expression of the *Leprecan* family of genes in greater detail, since all three mouse genes belonging to this family were expressed in the notochord and their embryonic expression was mostly uncharacterized.

Leprecan was first described as a basement membrane proteoglycan in rat yolk sac carcinoma L2 cells (Wassenhove-McCarthy and McCarthy, 1999), and two additional gene family members, *Leprecan-like1* and *Leprecan-like2*, were subsequently identified in both human and mouse (Järnum *et al.*, 2004). All *Leprecan* proteins contain a Prolyl 4-hydroxylase alpha (P4Halpha) domain and were therefore included in the 2-oxoglutarate- and Fe(II)-dependent oxygenase family (Aravind and Koonin, 2001), and *Leprecan* was eventually demonstrated to possess prolyl 3-hydroxylase activity (Vranka *et al.*, 2004). Although it was known that there are multiple *Leprecan* genes in vertebrates, as opposed to the single-copy gene found in the ascidian *C. intestinalis* (Hotta *et al.*, 2000), the broader evolution of *Leprecan* had not yet been explored.

Through a phylogenetic analysis based on sequence alignment of the highly conserved P4Halpha domain, we found that the ancestral *Leprecan* likely predates the bilaterians (Fig. 2), with a recognizable full-length copy found in the basal cnidarian *Nematostella vectensis*, and is also present in both the Ecdysozoans (*Apis mellifera* and *Tribolium castaneum*) and Lophotrochozoans (*Helobdella robusta*). Within the vertebrates, there are three well-supported clades of *Leprecan* genes, with *Leprecan-like2* being supported by a Bayesian posterior probability of 100 as sister to a clade of each of the other two *Leprecan* genes (Fig. 2, arrows). The relationships of the non-chordate Deuterostomes and invertebrates are less well supported, possibly, in part, due to minimal sampling or incomplete genome information.

The three ascidian sequences (from *C. intestinalis*, *C. savignyi*, and *Molgula tectiformis*) cluster together in a well-supported clade, and their weak representation here as sister to the vertebrates is in agreement with recent chordate phylogenies (Bourlat *et al.*, 2006). However, the clustering of the sea urchin, *Strongylocentrotus purpuratus*, sister to the cephalochordate, *Branchiostoma floridae*, in a polytomy with the vertebrates+ascidians and *Helobdella*, merely demonstrates that there is not enough information to determine the relationships between these groups based solely on this region of the *Leprecan* gene

sequence. This is also true for the polytomy at the base of the tree connecting *Nematostella* with *Apis* and *Tribolium*.

One point to note, however, is that the sequence obtained for *Leprecan* from *S. purpuratus* was truncated in the 5' region, which possibly led to its false placement next to *B. floridae*. Additionally, this clustering of sea urchin and amphioxus sequences is also likely due to the unavailability of a hemichordate sequence, as previously described in the case of similar comparisons (Bourlat *et al.*, 2006).

Finally, despite thorough genomic searches, no *Leprecan* sequences for any Dipterans (*Drosophila*, *Aedes*, or *Anopheles*) or for the nematode worm *Caenorhabditis elegans* were identified.

Whole-Mount In Situ Hybridization Reveals Conserved and Novel Domains of Expression of Mouse *Leprecan* Family Members

Expression of the three mouse *Leprecan* orthologs, *Leprecan*, *Leprel1*, and *Leprel2*, was further characterized at several time-points of mid-gestation by whole-mount in situ hybridization (WMISH). In general, it was found that while all three genes were expressed in the mouse notochord, each had acquired additional domains of expression, as also found for Brachyury in vertebrates (Wilkinson *et al.*, 1990). Details for each mouse *Leprecan* gene are presented below.

Leprecan—Expression of *Leprecan* was examined by WMISH at embryonic stages encompassing early and middle phases of notochord development (E9.5–E11.5). At E9.5, *Leprecan* exhibited little to no expression along the paraxial and lateral plate mesoderm of the trunk as well as in the forelimb buds and was faintly expressed in the developing notochord, otic capsule, and anterior craniofacial tissues (E9.5, Fig. 3a,b); expression in these territories was not detected in embryos hybridized with a sense probe (Fig. 3c). This pattern persisted through E10.5 (Fig. 3d), as expression became markedly visible in the notochord (Fig. 3e and red inset f), in the otic capsule (Fig. 3d), and craniofacial/branchial arch tissues (Fig. 3d), and was also present in the distal hindlimb but not in the forelimb bud (Fig. 3d,g). Furthermore, *Leprecan* expression was detected in the paraxial mesoderm (Fig. 3d). These territories were devoid of signal in control embryos hybridized with the sense probe (Fig. 3h). By E11.5, while otic capsule expression remained, *Leprecan* expression waned in craniofacial, limb, and paraxial mesoderm (Fig. 3i). Interestingly, notochord expression of this gene began to decrease in a rostral to caudal pattern, as it was not present at axial levels adjacent and/or anterior to the forelimb (Fig. 3j and black inset k) but remained in caudal domains, such as the hindlimb (Fig. 3l and red inset m). Finally, as a faint signal in the mid-brain region was seen at all gestational days and was moderately present also in embryos hybridized with the sense probe (Fig. 3c,h,n, and data not shown), we conclude that the staining seen in this region using the anti-sense probe is nonspecific signal most likely due to probe trapping within the head cavities.

Leprecan-like1—At E9.5, *Leprel1* was expressed in the otic capsule, similarly to *Leprecan*, although more robustly, in addition to being faintly expressed in the developing notochord (Fig. 4a,b). Expression was also specifically observed in the posterior-ventral region of the tail, presumably in the posterior-most section of the notochord and in the presomitic mesoderm (Fig. 4a,b), compared to sense control embryos (Fig. 4c). Unlike *Leprecan*, this gene was not expressed in other domains, even at the later gestational days E10.5 and E11.5 (Fig. 4d–m).

WMISH performed using a sense probe control revealed, particularly at E10.5, a weak signal in the otic capsule and in the head region (Fig. 4c,h,n), suggesting that the strong

signal detected in the otic capsule using the antisense probe may be partially reflecting background signal, whereas the strong expression in the head most likely results from probe trapping. Furthermore, some signal was detected in the posterior tail region at E10.5 by the sense probe, but was localized in the ectoderm (Fig. 4h, and data not shown), a pattern different from that observed using the antisense probe.

At E11.5, *Leprel1* expression persisted in the otic capsule and paraxial mesoderm (Fig. 4i), as well as throughout the notochord at all axial levels (Fig. 4j–m), while signal in the head region and distal limb ectoderm was likely due to probe trapping and/or background signal, as it was also seen in sense controls (Fig. 4n).

Leprecan-like2—To accurately examine the expression of *Leprel2*, three different probes, spanning different regions of the gene, were generated (see Methods section). Ultimately, we utilized for our WMISH studies sense and antisense probes spanning a 3' region of the *Leprel2* cDNA (*Leprel2^{Sma}*), as they provided reduced background signals relative to the probes spanning other regions (compare Figs. 5 and S2). For comparative purposes, Figure S2 depicts the *Leprel2* expression detected using the two alternative probes (*Leprel2^{Fl}* and *Leprel2^{Short}*). For all three probes utilized, we consistently found that unlike the other two *Leprecan* genes described earlier, *Leprel2* exhibited a much more widespread expression pattern at all gestational days analyzed.

At E9.5, *Leprel2^{Sma}* was only slightly expressed in anterior craniofacial, branchial arch, forelimb mesodermal, and ectodermal tissues (Fig. 5a,b). These patterns were more clearly detected when the alternative probes were employed (Fig. S2). Furthermore, differently from the other two *Leprecan* orthologs, which were not expressed in the paraxial mesoderm at this stage, *Leprel2* was already transcribed in this region (Fig. 5a,b); no expression was detected in this region in control embryos hybridized with the sense probe (Fig. 5c). Importantly, little to no genuine expression was detectable in the otic capsule and anterior notochord at this stage, a pattern that continued to be evident at both E10.5 and E11.5 (see below). By E10.5, *Leprel2^{Sma}* remained expressed at low levels in craniofacial tissues, such as the maxillary component of branchial arch 1 and the frontal nasal region (Fig. 5d). In addition, *Leprel2^{Sma}* was faintly expressed in both forelimb and hindlimb buds (Fig. 5d), in paraxial mesodermal tissues (Fig. 5d,g,h) and was also observed at this stage in the anterior notochord (Fig. 5e and red inset f). Interestingly, WMISH using alternative probes revealed more intense expression in all of these domains (Fig. S2). The expression identified in the head and in the otic capsules at these stages likely reflects probe trapping and background, as it was also apparent in sense controls (Fig. 5i). The same patterns of expression identified at E10.5 were markedly exhibited at E11.5, specifically in both the craniofacial and limb tissues (Fig. 5j–n), although paraxial mesoderm expression waned along the rostral to caudal axis (Fig. 5j). Finally, the faint signal seen in some mid-brain and frontal regions of the head was deemed as either background or the result of probe trapping, since it was also found in embryos hybridized with the sense probe (Fig. 5o).

Interestingly at E11.5, notochord expression of *Leprel2^{Sma}*, which was seen solely when the antisense probe was employed, became apparent, albeit at low levels, only in the more caudal domains of the embryo, at the level of the hindlimb bud (compare Fig. 5k and black inset 5l to m and red inset n). These data suggest a rostral-caudal asymmetry in the expression of this gene, similar to that seen at this gestational day for *Leprecan*.

Section In Situ Hybridization Highlights Differences in the Expression Domains of *Leprecan* Genes

To confirm and expand the results obtained on whole-mount mouse embryos, in situ hybridization was performed on sections obtained from embryos at stages E10.5 through

E13.5. Indeed, expression of all three *Leprecan* genes was identified in the notochord from E10.5 through E13.5. Interestingly, notochord expression of both *Leprecan* and *Leprel2* decreased in the anterior regions but persisted in the posterior regions by E13.5 (Fig. 6a' and data not shown).

Leprecan, while present in the notochord at all gestational days examined (Fig. 6a', E10.5–E13.5), was also expressed in the cartilages of the centra, laminae, and neural arches of the forming vertebrae (Fig. 6a' at E13.5; Fig. 6b', Panel c at E12.5; and Fig. 6b', Panels f and g at E13.5). Furthermore, this gene was markedly expressed in the precartilaginous mesenchyme and in the cartilage of the developing proximal humerus in the forelimb (Fig. 6b', Panel a at E12.5, and Panel d at E13.5), proximal femur in the hindlimb (Fig. 6b', Panel b at E12.5, and Panel e at E13.5), and pelvic girdle (Fig. 6b', Panel e at E13.5).

Leprel1 was also expressed in the developing notochord (Fig. 6a',b' Panel m), and was additionally expressed in the developing ventral spinal horns (Fig. 6a', E10.5) as well as in the olfactory epithelium (Fig. 6b', Panel j at E12.5), and in the proximal forelimb (Fig. 6b', Panel k at E13.5), proximal hindlimb and pelvis (Fig. 6b', Panel l at E13.5), and vertebral centra (Fig. 6b', Panel m at E13.5).

In agreement with the widespread expression seen by WMISH, *Leprel2* transcripts were also found in the dorsal root ganglia (Fig. 6a', E11.5–E12.5; and Fig. 6b' Panel s at E13.5), floor-plate (Fig. 6a', E12.5), mesenchyme, and cartilages of the proximal forelimb (data not shown), hindlimb (Fig. 6b', Panel n at E12.5, and Panel r at E13.5) and mandible (Fig. 6b', Panel o at E12.5), along with tissues of the developing eye (Fig. 6b', Panel p at E12.5) and olfactory epithelium (Fig. 6b', Panel q at E12.5). Finally, *Leprel2* was additionally expressed in the developing kidney (Fig. 6b', Panel t and red and blue insets at E13.5), specifically in clusters of mesenchymal cells within the cortex, likely representing the nephrogenic precursors (blue inset), as well as within the developing medulla, in epithelial cells lining the future ureteric ducts (red inset).

DISCUSSION

Evolutionarily Conserved Notochord Expression of Putative Brachyury Targets

The notochord plays a vital role in chordate embryonic development. A detailed knowledge of the effectors responsible for notochord specification and differentiation is necessary not only to understand how this structure serves all its functions, but also to gain insights on etiology, treatment, and prevention of developmental disorders and birth defects associated with its formation. In a broader perspective, a basic suite of evolutionarily conserved notochord genes likely provided a first stepping stone for the crucial transitions that ultimately led to the appearance of the backbone (Fig. 7). We took advantage of the information gained on Brachyury-downstream notochord genes in an ascidian chordate, *C. intestinalis*, to identify genes that have retained their expression in the notochord across the chordate clade and to document the expansion of their expression territories. We first identified putative mouse orthologs of *Ciona* Brachyury-downstream notochord genes via sequence comparisons and then examined their expression in E10.5 mouse embryos by section in situ hybridization. Next, we undertook a characterization of the dynamic expression patterns exhibited by the three members of the murine *Leprecan* gene family during mid-gestation.

Of the 12 mouse genes examined here (Table S1), nine (the three *Leprecan* orthologs, *Agrin*, *Papss1*, *Pellino1*, *Pellino2*, *SerpinC1*, and *SerpinG1*) showed expression in the notochord. However, several of the genes we surveyed also showed expression in some of its neighboring tissues. Specifically, all genes examined except *Papss1* showed expression in

the neural tube, frequently including the floor-plate (Fig. 1 and S1). Our results correlate well with the previously published data on expression patterns and functional roles for some of these genes, such as *Agrin*, which encodes a proteoglycan expressed at the neuromuscular junction and in neural tissues (Burgess *et al.*, 2000).

These findings are of particular interest given the concomitant expansion of *Brachyury* expression into the paraxial mesoderm in vertebrates and in amphioxus (Holland *et al.*, 1995), compared to the ascidians. In fact, in all the ascidians analyzed so far, including *Ciona*, *Molgula*, and *Halocynthia*, *Brachyury* is expressed solely in the notochord (Corbo *et al.*, 1997; Takada *et al.*, 2002; Yasuo and Satoh, 1993). In contrast, in the mouse, expression of *Brachyury* is initially pan-mesodermal, becoming restricted to the notochord at later stages (Wilkinson *et al.*, 1990). It is not yet clear whether ascidians have reduced the expression of *Brachyury* and its target genes to the notochord, or whether expression has expanded to the paraxial mesoderm in vertebrates. Given mouse *Brachyury*'s expression in this novel domain, it may be reasonable to expect the expression of its early target genes to have expanded to other mesodermal derivatives as well. Future studies will be required to verify that the expression of the genes presented here is controlled in the notochord by this transcription factor. However, through a preliminary analysis of the genomic loci of some of the genes shown in Figure 1, we have identified putative *Brachyury* binding sites in the *Pellino1* sequence (data not shown). In particular, the *Pellino1* locus contains two tandemly arranged half-sites spaced by 8 bp and one of these sites is part of an incomplete palindrome, consistent with the requirements identified for mouse *Brachyury* (Kispert *et al.*, 1995).

The *Leprecan* Gene Family

We took advantage of the broad variety of genomes publicly available to identify *in silico* sequences of *Leprecan* genes throughout the metazoans and generated a multisequence alignment of their conserved P4Halpa domains (see below). Phylogenetic analysis of this alignment resulted in the tree shown in Figure 2. The general structure of the tree is in agreement with a recent model of chordate phylogeny that places the cephalochordate *Branchiostoma* basal to the urochordates, which are in turn basal to the vertebrates (Delsuc *et al.*, 2006). The minimal sampling across highly divergent groups is likely responsible for the relatively undefined or poorly supported non-chordate relationships seen in the tree presented here. In addition, we noticed that the sea urchin *S. purpuratus* *Leprecan* gene that we identified *in silico* contained only the 3' half of the sequence encoding the P4Halpa domain. This truncation is likely an artifact of the genome assembly and it presumably led to the aberrant placement of the echinoderm sequence in our *Leprecan* phylogeny.

The two rounds of gene duplications along the branch leading to the vertebrates after the divergence of the tunicates is in line with the 2R hypothesis, which suggests that two rounds of genome duplications occurred in this lineage (Ohno, 1999). However, the presence of a single *Leprel2* gene in the organisms examined so far suggests that either one *Leprel2* paralog was lost in the vertebrate stem group or that *Leprel2* was not involved in a duplication event.

Leprecan was initially described as a basement membrane proteoglycan (Wassenhove-McCarthy and McCarthy, 1999) and as a potential tumor suppressor (Kaul *et al.*, 2000). Subsequently, it has been characterized as a functional prolyl 3-hydroxylase, which along with CRTAP and Cyclophilin B is part of a complex responsible for converting the proline in the Gly-Pro-4-Hyp motifs of procollagen peptides to 3-hydroxyproline (3-Hyp) (Vranka *et al.*, 2004). This modification of collagen occurs prior to the “zipping up” of the collagen triple helix in the rough endoplasmic reticulum (rER) (Gelse *et al.*, 2003; Myllyharju and Kivirikko, 2004).

Duplications of fibrillar collagens have also been the focus of recent discussion regarding the evolution of the chordate lineages (Boot-Handford and Tuckwell, 2003; Rychel and Swalla, 2007; Wada *et al.*, 2006). Correlating the non-overlapping domains of expression that we detected for the three *Leprecan* genes with those of different vertebrate fibrillar collagens could provide insights into possible substrate specificities of divergent *Leprecan* proteins. Outside of the chordates, the lack of *Leprecan* sequences in the Dipterans (e.g., *Drosophila*) and the nematode worm *C. elegans* is intriguing, since neither *Drosophila* nor *C. elegans* have any fibrillar collagens (Hynes and Zhao, 2000).

The presence of a distinct *Leprecan* ortholog in the cnidarian *N. vectensis* dates the origins of the *Leprecan* gene family as far back as the origins of the metazoans, approximating the origins of true collagens. However, while collagens have been found in several sponges (Aho *et al.*, 1993; Boute *et al.*, 1996; Exposito and Garrone, 1990), it is yet unclear whether these organisms contain *Leprecan* orthologs or whether their collagens are also modified with 3-hydroxyproline. As additional sequences become available from diverse organisms, these questions may soon be answered.

Analysis of *Leprecan* sequences revealed the presence of a P4Halp domain near the 3'-end of these genes (Aravind and Koonin, 2001). However, the 5' region of *Leprecan*, when analyzed separately, shows sequence similarity with both the *Leprecan* binding partner CRTAP and a poorly characterized nucleolar protein, No55 (Wassenhove-McCarthy and McCarthy, 1999). In fact, two phylogenies generated by TreeFam (<http://www.treefam.org/>, trees TF320837 and TF510574; Li *et al.*, 2006) assign the *Leprecan* gene family as sister to a clade of these two genes. Although these phylogenies demonstrate the overall evolutionary relationships of these genes, the interrelationships among the *Leprecan* copies vary between the two TreeFam trees, with little branch support distinguishing them from a polytomy. Additionally, the TreeFam taxon sampling includes numerous vertebrate species, but only a single invertebrate, *C. intestinalis*. And although *Leprecan* genes do share sequence homology with CRTAP and No55, these two genes, as well as the nematode and Dipteran genes used as outgroups in the TreeFam trees, lack any sequence related to the P4Halp domain. In this light, the phylogenetic analysis presented here provides a more in-depth examination of the relationships within the *Leprecan* gene family.

Expression of Murine *Leprecan* genes

Based on the conserved notochord expression of all three mouse *Leprecan* gene family members, and since the only published embryonic expression data for any vertebrate *Leprecan* gene was the characterization of *Leprecan* in Day-16 chick embryos (Vranka *et al.*, 2004), we chose to study their expression in mouse embryos in more detail (Figs. 3-6). Our analysis shows that, in addition to the notochord, novel expression domains are evident in mid-gestation mouse embryos, suggesting potential areas where *Leprecan* proteins might function during several organogenetic and skeletogenic processes.

Specifically, we have found that all three *Leprecan* genes are initially expressed at low levels during early mid-gestation, at E9.5/E10.5, in the notochord and paraxial mesoderm, with a much larger expression domain for *Leprel2*. Subsequently, by E10.5/E11.5, expression of each gene expands into wider domains. Furthermore, we observed that at E10.5 and E11.5 at least two, but often all three *Leprecans* are co-expressed, albeit at different levels, in several territories (i.e., paraxial mesoderm, notochord, limbs, craniofacial, and branchial tissues), suggesting that their reiterative presence might have a functional importance for the development of these regions. These dynamic patterns of expression are not restricted to early mid-gestation, as they are also observed at later stages. For example, at E12.5 and E13.5, *Leprel2* expression appears more widespread (e.g., kidney, olfactory epithelium, eye, skeletal tissues) and has a prolonged temporal progression

compared to other members, such as *Leprecan*, which is only co-expressed with this gene in skeletogenic tissues, and *Leprel1*, which exhibits an even more restricted spatio-temporal pattern. These patterns may reflect partially cooperative functions of different Leprecans in these tissues, for example, in the development of the axial (i.e., vertebral anlagen) and appendicular (e.g., humerus, femur, pelvic girdle) skeleton, as well as more specific contributions to other domains, as in the case of *Leprel2* in the developing Meckel's cartilage, eye, olfactory epithelium, and kidney.

Currently, little is known about the significance of 3-Hyp in collagen, although it has been shown that the hydroxylation of the central proline residue in the collagen Gly-Pro-4-Hyp repeats to 3-Hyp decreases the stability of its triple helices (Jenkins *et al.*, 2003). These previous studies have led to the hypothesis that the primary role of Leprecan is to locally perturb the collagen triple helix in basement membrane collagens, specifically allowing for their less rigorously structured macromolecular form. Our results show that *Leprecan* family members are expressed throughout various tissues of the developing embryo, the majority of which are coincident with regions of fibrillar collagen expression, especially in areas of skeletal cartilage condensations. *Leprecan* is found in the proximal humerus and femur, the pelvic girdle, and the developing vertebrae, whereas *Leprel2* is expressed in limb and mandibular cartilages. These results suggest that Leprecan proteins may widely function in processing of fibrillar collagens, rather than functioning solely in the basement membrane as previously proposed. This hypothesis is supported by the analysis of *CRTAP*^{-/-} mice, which are characterized by the loss of 3-Hyp in type I and II fibrillar collagens (Morello *et al.*, 2006). CRTAP functions in a complex with Leprecan (Vranka *et al.*, 2004), and in E13.5 mice is found localized to areas of chondrogenesis, similar to our findings with *Leprecan* family genes. Together, these results point to the importance of Leprecan proteins in the control of the macromolecular structure of fibrillar collagens.

Several studies have been devoted to elucidating *Leprecan* expression in the kidney in various organisms with differing results. For example, Wassenhove-McCarthy and McCarthy (1999) first described in adult rats the localization of Leprecan in the glomerular basement membrane, mesangial matrix, and Bowman's capsule of the nephron. It was later shown that the antibody used to detect Leprecan in rats had been raised against the highly conserved P4Halp domain, prompting concerns of cross-reactivity with other Leprecan family proteins (Lauer *et al.*, 2007). When a new, Leprecan-specific antibody was employed for immunofluorescence, a signal was detected also in podocytes, whereas in Western blot assays this antibody detected multiple bands, representing either alternatively spliced forms or differential post-translational modifications of Leprecan (Lauer *et al.*, 2007). In contrast, in humans, *Leprecan* (*LEPRE1*) was not detected in the adult kidney by Northern blot analysis (Kaul *et al.*, 2000), whereas *Leprecan-like1* (*LEPREL1*) was found to be expressed in this organ (Järnum *et al.*, 2004). The only prenatal examination of *Leprecan* genes expression to date was carried out in the chick embryo, where, in the kidney, *Leprecan* was found to be expressed in the calyx but neither in the glomeruli nor in the tubules (Vranka *et al.*, 2004). Interestingly, in our examination of sections of developing mouse embryos at E13.5, we noticed that, in the developing kidney, *Leprel2* is expressed in both the mesenchyme of the developing cortical nephrogenic precursors as well as within the epithelial cells of the ureteric buds, suggesting that this gene may have a role in prenatal renal development.

Evolution of the *Leprecan* Gene Family Across Chordates

Figure 7 provides a possible sequence of the events that led from the single-copy, notochord-specific *Leprecan* gene found in a tunicate, the ascidian *C. intestinalis*, to the tripartite gene family found in mouse and in all vertebrates so far investigated. In our simplified evolutionary scenario, the ancestral *Leprecan* gene found in non-chordates was

co-opted to the notochord in tunicate, likely through an individual notochord *cis*-regulatory module directly controlled by Ci-Bra (M.P.D. and A.D.G., unpublished results). Alternatively, *Leprecan* could have already been a Brachyury target at the time when Brachyury was co-opted to the notochord. As genome duplications occurred, the ancestral *Leprecan* and its *cis*-regulatory module(s) duplicated and independently started accumulating mutations, and this in turn led to the diversification of the individual genes. In parallel, the appearance of additional *cis*-regulatory elements/modules and the modification of existing ones ushered the expansion of their domains of expression.

Although the notochord expression of *Leprecan* genes might be considered, from the standpoint of parsimony, to have been retained from ascidians to mice, the argument for the conservation of the transcriptional regulation by Brachyury is obviously labile. In fact, while it is noteworthy that all three *Leprecan* genes are expressed in the notochord and the paraxial mesoderm, the timing of their expression, at least in the paraxial mesoderm, does not correlate well with a direct activation of *Leprecan* genes by Brachyury, as this gene is not expressed in the paraxial mesoderm at the developmental stages analyzed here. Additionally, the different temporal onset of expression seen for different *Leprecan* genes in the notochord, as well as the differences seen in their rostral-caudal expression, suggest the possibility that these genes might be either controlled by Brachyury through spatially and temporally localized co-factors, or even be controlled by a different transcription factor(s) altogether. Our preliminary searches of Brachyury binding sites along the genomic loci of the notochord genes examined in this study have focused on putative sites related to the palindromic consensus sequence identified by Kispert *et al.* (1995). Nevertheless, Brachyury proteins from *Xenopus* and *Ciona* are able to efficiently bind half-sites in vitro and presumably control their target genes in vivo through such sequences (Casey *et al.*, 1998; Di Gregorio and Levine, 1999); indeed, more recently, also the mouse protein has been shown to be able to bind half-sites in a cooperative fashion (Kusch *et al.*, 2002).

The information that is rapidly accumulating from multiple model systems on the binding properties of Brachyury as well as on its cofactors and intermediaries will aid future searches for the *cis*-regulatory modules controlling notochord expression of the genes that we have characterized.

METHODS

Plasmids and RNA Probe Synthesis

For all genes analyzed in this work, IMAGE mouse cDNA clones (Lennon *et al.*, 1996) were obtained from Open Biosystems (Huntsville, AL; <http://www.openbiosystems.com>). Clone numbers are listed in Table S1. The *Shh* antisense probe shown in Figure S1 was kindly provided by Dr. A. P. McMahon. IMAGE clones containing full-length ESTs for *Leprecan*, *Leprel1*, and *Leprel2* were modified via restriction digestion to generate gene-specific probes, as described later.

A 2.6-kb *XhoI* fragment was excised from IMAGE clone 5148677, containing the *Leprecan* coding sequence, and the plasmid was re-ligated upon itself. This procedure left only the initial 613 bp of *Leprecan* cDNA in the plasmid, which was then used for both sense and antisense probe synthesis.

A region corresponding to ~900 bp of the 3'-UTR of the *Leprel1* cDNA was PCR-amplified from IMAGE clone 2182260 using the following primers (nucleotides introducing restriction sites in the amplified DNA are indicated as lowercase):

mLeprel1.F: 5'-gtcagatattGGATCCCTTTATTAAACACTTCAGA-3' and *mLeprel1.R*: 5'-cagtcctcagGATCGGCTCGCTAGCTCTCGGAGTT-3'.

After digestion with *EcoRV* and *XhoI*, the PCR fragment was ligated into pBluescriptII KS+ to generate the plasmid, which was then used for both sense and anti-sense probe synthesis.

Three separate probes were generated for *Leprel2*: *Leprel2^{Fl}*, *Leprel2^{Short}*, and *Leprel2^{Sma}*. The *Leprel2^{Fl}* probe derives from the unmodified IMAGE clone 3497610, containing the full-length cDNA for *Leprel2*. To generate *Leprel2^{Short}*, a 1.8-kb *EcoRV* fragment was excised from IMAGE clone 3497610, removing the *Leprel2* coding sequence, and the plasmid was then religated, leaving only 426 bp of the *Leprel2* 3'-UTR in the insert, corresponding to bp 2268 through 2694 of Gen-Bank sequence AJ441086. To generate *Leprel2^{Sma}*, an 832 bp *SmaI* fragment containing both the 3' terminus of the *Leprel2* coding sequence as well as a portion of the 3'-UTR, corresponding to bp 1553 through 2384 of GenBank sequence AJ441086, was excised from IMAGE clone 3497610 and ligated into pBluescriptII KS+.

Whole-Mount and Section In Situ Hybridization

WMISH were performed on C57BL/6 WT mouse embryos at different gestational days using digoxigenin-labeled antisense and sense RNA probes, as previously described (Di Giacomo *et al.*, 2006; Selleri *et al.*, 2001).

For section in situ hybridization, C57BL/6 WT embryos were harvested at E10.5 to E13.5 and fixed O/N at 4°C in phosphate-buffered saline (PBS) containing 4% paraformaldehyde. Next, they were embedded in Tissue-Tek O.C.T. Compound (Sakura Finetek, Torrance, CA), as described in Di Giacomo *et al.* (2006). All section in situ hybridizations were performed on multiple, diverse transverse sections from embryos at different gestational days. For WMISH experiments alkaline phosphatase activity was detected using only the BCIP substrate to reduce background signals. Both antisense and sense probes were detected for approximately equal amounts of time. Images were captured using a Magna-fire digital camera (Optronix).

In Silico Identification of *Leprecan* Sequences

Sequences for all vertebrate *Leprecan* genes, as well as those for *Apis*, *C. savignyi*, and *Tribolium*, were obtained from the Ensembl database (<http://www.ensembl.org/>). The *M. tectiformis* *Leprecan* sequence was retrieved from the NCBI database (<http://www.ncbi.nlm.nih.gov/sites/entrez>). A full-length *C. intestinalis* *Leprecan* cDNA sequence was obtained via a 5'-RACE reaction on total RNA extracted from mid-tailbud embryos, using the GeneRacer kit (Invitrogen, Carlsbad, CA) and following the manufacturer's instructions. *Leprecan* sequences for *Nematostella* and *Branchiostoma* were obtained from the JGI web site (<http://www.jgi-doe.gov/>) via BLAST searches (Altschul *et al.*, 1990) using the *C. intestinalis* *Leprecan* sequence, whereas the *Helobdella* gene was identified using the *Apis mellifera* *Leprecan* sequence. The largest assembled genes were chosen from each search. The *S. purpuratus* sequence was identified via a BLAST search of the Human Genome Sequencing Center website (<http://www.hgsc.bcm.tmc.edu/projects/seaurchin/>) using the *C. intestinalis* *Leprecan* sequence. All sequences were verified as *Leprecan* orthologs by their monophyly in comparison to the P4Halpa domain sequences of vertebrate P4H genes (data not shown). GenBank accession numbers or Ensembl gene numbers for the sequences used in Figure 2 are listed in Table S2.

Leprean Phylogenetic Alignment and Analyses

Nucleotide sequences corresponding to the P4Halpa domain of each gene were manually aligned based on amino acid sequences in MACCLADE 4 (Maddison and Maddison, 2003). The entire aligned nucleotide matrix was used for subsequent phylogenetic analyses. Parameters for the Bayesian analyses were estimated by using MODELTEST 3.06 (Posada and Crandall, 1998). The Akaike Information Criterion (Akaike, 1973) recommended a general time reversible (GTR) model with added parameters for invariable sites and a g -distribution (GTR + i + g). Bayesian analyses were conducted using this model for the entire matrix. We used the Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) method as implemented in MrBayes, version 3.1.2 (Huelsenbeck and Ronquist, 2001), to run four chains (three heated). We ran 5 million generations, sampling every 1,000 generations, using a burn-in of 1,000 trees (1 million generations). A majority-rule consensus was calculated from the remaining trees in PAUP*, version 4.0b10 (Swofford, 2001), to determine the posterior probabilities. Sequence alignments are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors kindly thank members of the Di Gregorio, Selleri, and Nibu labs for assistance and helpful comments. We thank Dr. Billie Swalla for critically reading our manuscript and Dr. Yutaka Nibu for insightful discussion and comments. We thank Dr. D.G. Howarth for assistance with phylogenetic methods and for critical comments on the manuscript. A.D.G. and L.S. are Irma T. Hirschl Scholars.

Contract grant sponsor: NIH/NICHD; Contract grant numbers: R01HD050704, R01HD43997; Contract grant sponsor: March of Dimes Birth Defects Foundation (Basil O'Connor Starter Scholar Research Award); Contract grant number: 5-FY03-153.

Abbreviations

BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pair(s)
CNS	central nervous system
kb	kilobase(s), or 1,000 base pairs
NBT	nitro blue tetrazolium
PCR	polymerase chain reaction
WT	wild-type

LITERATURE CITED

- Adams D, Keller R, Koehl M. The mechanics of notochord elongation, straightening and stiffening in the embryo of *Xenopus laevis*. *Development*. 1990; 110:115–130. [PubMed: 2081454]
- Aho S, Turakainen H, Onnela M, Boedtker H. Characterization of an intronless collagen gene family in the marine sponge *Microciona prolifera*. *Proc Natl Acad Sci USA*. 1993; 90:7288–7292. [PubMed: 8346246]
- Akaike, H. Information theory as an extension of the maximum likelihood principle. In: Petrov, BN.; Csaki, F., editors. *Second International Symposium on Information Theory*. Akademiai Kiado; Budapest: 1973. p. 267-281.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–410. [PubMed: 2231712]

- Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* 2001; 2:RESEARCH0007. [PubMed: 11276424]
- Ben Abdelkhalek H, Beckers A, Schuster-Gossler K, Pavlova MN, Burkhardt H, Lickert H, Rossant J, Reinhardt R, Schalkwyk LC, Muller I, Herrmann BG, Ceolin M, Rivera-Pomar R, Gossler A. The mouse homeobox gene *Not* is required for caudal notochord development and affected by the truncate mutation. *Genes Dev.* 2004; 18:1725–1736. [PubMed: 15231714]
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature.* 2006; 444:85–88. [PubMed: 17051155]
- Boot-Handford RP, Tuckwell DS. Fibrillar collagen: The key to vertebrate evolution? A tale of molecular incest. *Bioessays.* 2003; 25:142–151. [PubMed: 12539240]
- Boute N, Exposito JY, Boury-Esnault N, Vacelet J, Noro N, Miyazaki K, Yoshizaki K, Garrone R. Type IV collagen in sponges, the missing link in basement membrane ubiquity. *Biol Cell.* 1996; 88:37–44. [PubMed: 9175266]
- Burgess RW, Skarnes WC, Sanes JR. Agrin isoforms with distinct amino termini: Differential expression, localization, and function. *J Cell Biol.* 2000; 151:41–52. [PubMed: 11018052]
- Casey ES, O'Reilly MA, Conlon FL, Smith JC. The T-box transcription factor Brachyury regulates expression of eFGF through binding to a non-palindromic response element. *Development.* 1998; 125:3887–3894. [PubMed: 9729496]
- Casey ES, Tada M, Fairclough L, Wylie CC, Heasman J, Smith JC. Bix4 is activated directly by VegT and mediates endoderm formation in *Xenopus* development. *Development.* 1999; 126:4193–4200. [PubMed: 10477288]
- Cleaver O, Krieg PA. Notochord patterning of the endoderm. *Dev Biol.* 2001; 234:1–12. [PubMed: 11356015]
- Copp AJ, Greene ND, Murdoch JN. The genetic basis of mammalian neurulation. *Nat Rev Genet.* 2003; 4:784–793. [PubMed: 13679871]
- Corbo JC, Levine M, Zeller RW. Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development.* 1997; 124:589–602. [PubMed: 9043074]
- Delsuc F, Brinkman H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature.* 2006; 439:965–968. [PubMed: 16495997]
- Di Gregorio A, Levine M. Regulation of *Ci-tropomyosin-like*, a Brachyury target gene in the ascidian, *Ciona intestinalis*. *Development.* 1999; 126:5599–5609. [PubMed: 10572037]
- Di Giacomo G, Koss M, Capellini TD, Brendolan A, Pöpperl H, Selleri L. Spatio-temporal expression of *Pbx3* during mouse organogenesis. *Gene Expr Patterns.* 2006; 6:747–57. [PubMed: 16434237]
- Di Gregorio A, Harland RM, Levine M, Casey ES. Tail morphogenesis in the ascidian, *Ciona intestinalis*, requires cooperation between notochord and muscle. *Dev Biol.* 2002; 244:385–395. [PubMed: 11944945]
- Echelard Y, Epstein DJ, St-Jacques B, Shen L, Mohler J, McMahon JA, McMahon AP. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell.* 1993; 75:1417–1430. [PubMed: 7916661]
- Exposito JY, Garrone R. Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. *Proc Natl Acad Sci USA.* 1990; 87:6669–6673. [PubMed: 2395869]
- Fleming A, Keynes RJ, Tannahill D. The role of the notochord in vertebral column formation. *J Anat.* 2001; 199(Part 1–2):177–180. [PubMed: 11523820]
- Fouquet B, Weinstein BM, Serluca FC, Fishman MC. Vessel patterning in the embryo of the zebrafish: Guidance by notochord. *Dev Biol.* 1997; 183:37–48. [PubMed: 9119113]
- Gelse K, Pöschl E, Aigner T. Collagens—Structure, function, and biosynthesis. *Adv Drug Deliv Rev.* 2003; 55:1531–1546. [PubMed: 14623400]

- Holland PW, Koschorz B, Holland LZ, Herrmann BG. Conservation of *Brachyury* (*T*) genes in amphioxus and vertebrates: Developmental and evolutionary implications. *Development*. 1995; 121:4283–4291. [PubMed: 8575328]
- Hotta K, Takahashi H, Asakura T, Saitoh B, Takatori N, Satou Y, Satoh N. Characterization of Brachyury-downstream notochord genes in the *Ciona intestinalis* embryo. *Dev Biol*. 2000; 24:69–80. [PubMed: 10898962]
- Hotta K, Takahashi H, Erives A, Levine M, Satoh N. Temporal expression patterns of 39 Brachyury-downstream genes associated with notochord formation in the *Ciona intestinalis* embryo. *Dev Growth Differ*. 1999; 41:657–664. [PubMed: 10646795]
- Hotta K, Takahashi H, Satoh N, Gojobori T. Brachyury-downstream gene sets in a chordate, *Ciona intestinalis*: Integrating notochord specification, morphogenesis and chordate evolution. *Evol Dev*. 2008; 10:37–51. [PubMed: 18184356]
- Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001; 17:754–755. [PubMed: 11524383]
- Hynes RO, Zhao Q. The evolution of cell adhesion. *J Cell Biol*. 2000; 150:F89–F95. [PubMed: 10908592]
- Järnum S, Kjellman C, Darabi A, Nilsson I, Edvardsen K, Aman P. LEPREL1, a novel ER and Golgi resident member of the Leprecan family. *Biochem Biophys Res Commun*. 2004; 317:342–351. [PubMed: 15063763]
- Jenkins CL, Bretscher LE, Guzei IA, Raines RT. Effect of 3-hydroxyproline residues on collagen stability. *J Am Chem Soc*. 2003; 125:6422–6427. [PubMed: 12785781]
- Kaul SC, Sugihara T, Yoshida A, Nomura H, Wadhwa R, Gros L, a potential growth suppressor on chromosome 1: Its identity to basement membrane-associated proteoglycan, leprecan. *Oncogene*. 2000; 19:3576–3583. [PubMed: 10951563]
- Kispert A, Koschorz B, Herrmann BG. The T protein encoded by *Brachyury* is a tissue-specific transcription factor. *EMBO J*. 1995; 14:4763–4772. [PubMed: 7588606]
- Kusch T, Storck T, Walldorf U, Reuter R. Brachyury proteins regulate target genes through modular binding sites in a cooperative fashion. *Genes Dev*. 2002; 16:518–529. [PubMed: 11850413]
- Lauer M, Scruggs B, Chen S, Wassenhove-McCarthy D, McCarthy KJ. Leprecan distribution in the developing and adult kidney. *Kidney Int*. 2007; 72:82–91. [PubMed: 17495866]
- Lennon G, Auffray C, Polymeropoulos M, Soares MB. The IMAGE consortium: An integrated molecular analysis of genomes and their expression. *Genomics*. 1996; 33:151–152. [PubMed: 8617505]
- Li H, Coghlan A, Ruan J, Coin LJ, Hériché JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, Wong GK, Zheng W, Dehal P, Wang J, Durbin R. TreeFam: A curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res*. 2006; 34(Database issue):D572–D580. [PubMed: 16381935]
- Maddison, DR.; Maddison, WP. *MacClade: Analysis of phylogeny and character evolution*. Sinauer; Sunderland, MA: 2003.
- Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M, Castagnola P, Rauch F, Glorieux FH, Vranka J, Bächinger HP, Pace JM, Schwarze U, Byers PH, Weis M, Fernandes RJ, Eyre DR, Yao Z, Boyce BF, Lee B. CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell*. 2006; 127:291–304. [PubMed: 17055431]
- Myllyharju J, Kivirikko KI. Collagens, modifying enzymes and their mutations in humans, flies, and worms. *Trends Genet*. 2004; 20:33–43. [PubMed: 14698617]
- Ohno S. Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. *Semin Cell Dev Biol*. 1999; 10:517–522. [PubMed: 10597635]
- Placzek M, Tessier-Lavigne M, Yamada T, Jessell T, Dodd J. Mesodermal control of neural cell identity: Floor plate induction by the notochord. *Science*. 1990; 250:985–988. [PubMed: 2237443]
- Posada D, Crandall KA. MODELTEST: Testing the model of DNA substitutions. *Bioinformatics*. 1998; 14:817–818. [PubMed: 9918953]
- Rychel AL, Swalla BJ. Development and evolution of chordate cartilage. *J Exp Zool B Mol Dev Evol*. 2007; 308:325–335. [PubMed: 17358002]

- Schulte-Merker S, van Eeden FJ, Halpern ME, Kimmel CB, Nüsslein-Volhard C. No tail (*ntl*) is the zebrafish homologue of the mouse T (*Brachyury*) gene. *Development*. 1994; 120:1009–1015. [PubMed: 7600949]
- Selleri L, Depew MJ, Jacobs Y, Chanda SK, Tsang KY, Cheah KS, Rubenstein JL, O'Gorman S, Cleary ML. Requirement for *Pbx1* in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development*. 2001; 128:3543–3557. [PubMed: 11566859]
- Stemple DL. Structure and function of the notochord: An essential organ for chordate development. *Development*. 2005; 132:2503–2512. [PubMed: 15890825]
- Swofford, DL. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer; Sunderland, MA: 2001.
- Tada M, Casey ES, Fairclough L, Smith JC. Bix1, a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. *Development*. 1998; 125:3997–4006. [PubMed: 9735361]
- Tada M, Smith JC. Xwnt11 is a target of Xenopus Brachyury: Regulation of gastrulation movements via dishevelled, but not through the canonical Wnt pathway. *Development*. 2000; 127:2227–2238. [PubMed: 10769246]
- Takada N, York J, Davis JM, Schumpert B, Yasuo H, Satoh N, Swalla BJ. Brachyury expression in tailless *Molgulid* ascidian embryos. *Evol Dev*. 2002; 4:205–211. [PubMed: 12054293]
- Takahashi H, Hotta K, Erives A, Di Gregorio A, Zeller RW, Levine M, Satoh N. Brachyury downstream notochord differentiation in the ascidian embryo. *Genes Dev*. 1999; 13:1519–1523. [PubMed: 10385620]
- Vranka JA, Sakai LY, Bächinger HP. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J Biol Chem*. 2004; 279:23615–23621. [PubMed: 15044469]
- Wada H, Okuyama M, Satoh N, Zhang S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol Dev*. 2006; 8:370–377. [PubMed: 16805901]
- Wassenhove-McCarthy DJ, McCarthy KJ. Molecular characterization of a novel basement membrane-associated proteoglycan, Leprecan. *J Biol Chem*. 1999; 274:25004–25017. [PubMed: 10455179]
- Wilkinson DG, Bhatt S, Herrmann BG. Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature*. 1990; 343:657–659. [PubMed: 1689462]
- Wilson V, Manson L, Skarnes WC, Beddington RSP. The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development*. 1995; 121:877–886. [PubMed: 7720590]
- Wilson V, Rashbass P, Beddington RS. Chimeric analysis of *T* (*Brachyury*) gene function. *Development*. 1993; 117:1321–1331. [PubMed: 8404534]
- Yasuo H, Satoh N. Function of vertebrate T gene. *Nature*. 1993; 364:582–583. [PubMed: 8350918]

FIG. 1.

Expression of mouse genes in notochord and CNS of E10.5 embryos. In situ hybridization of putative mouse orthologs of Ci-Bra target genes on E10.5 transverse sections at several different axial levels reveals that *Agrin*, *Papss1*, *Pellino1*, *Pellino2*, *SerpinC1*, and *SerpinG1* are expressed in the notochord (black arrowheads). With the exception of *Papss1* (empty black arrow), all other genes are also expressed in the developing neural tube (black arrows). Dorsal is up in all panels.

FIG. 2.

Phylogenetic analysis of the *Leprecan* gene family across metazoans. The relationships between the three vertebrate *Leprecan* genes suggest a history of two gene duplications, the first leading to *Leprecan-like2* and a precursor of *Leprecan/Leprecan-like1*, and the second duplication leading to *Leprecan* and *Leprecan-like1*, with each gene strongly supported as monophyletic within the vertebrates (arrows). The ascidian *Leprecan* gene clearly appears to be most closely related to the single ancestral *Leprecan* and not to a specific single copy of the vertebrate family, further supporting a hypothesis of gene duplication within the vertebrate lineage in this gene family. Additional metazoan *Leprecan* orthologs are included as outgroups. Note the absence of *Drosophila* or *C. elegans* sequences. Numbers above the lines indicate Bayesian posterior probabilities. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

FIG. 3.

Expression of mouse *Lepreca* in whole-mount embryos. WMISH of *Lepreca* at stages E9.5 (**a–c**), E10.5 (**d–h**), and E11.5 (**i–n**) using antisense and sense probes shows that, beginning at E9.5 (**a, b**), and continuing at E10.5 (**e** and red inset shown in **f**) and E11.5 (**l** and red inset shown in **m**), *Lepreca* is expressed in the developing notochord (black arrowheads). This signal was absent in sense controls (**c, h**, and **n**). At E11.5, expression is absent from the notochord in anterior domains (**j** and black inset **k**, empty black arrowheads). Furthermore, while at E9.5 *Lepreca* is weakly expressed in the otic capsule (**a** and **b**, short red arrows) and craniofacial tissues (**a** and **b**, long red arrows), it becomes more strongly expressed in these tissues at E10.5 and E11.5 (**d** and **i**, short and long black/red arrows, respectively). Finally, while at E10.5 *Lepreca* is expressed in the paraxial mesoderm (**d**, green arrows) and hindlimb (**d** and **g**, red arrows), at E11.5 its expression decreases in these structures (**i**, short red and white arrows, respectively). Dashed red lines in **d** and **i** indicate the level of the sections shown in **e** and **l**, respectively, whereas the dashed black line in **i** depicts the level of section **j**. Small red arrowheads depict areas of probe trapping and background signal, as revealed by comparisons with sense controls (**c, h, n**).

FIG. 4.

Expression of mouse *Leprecan-like1* in whole-mount embryos. WMISH of *Leprell1* at stages E9.5 (a–c), E10.5 (d–h), and E11.5 (d–h)) using antisense and sense probes reveals that, beginning at E9.5 (a, b), and continuing from E10.5 (e and red inset shown in f) through E11.5 (j and black inset k, as well as l and red inset shown in m), *Leprell1* is expressed in the developing notochord (black arrowheads). At E9.5, it is also specifically expressed in the distal tail mesoderm (b, short red arrow), compared to sense controls (c, empty red arrowhead). At E10.5–E11.5 it is also expressed in the paraxial mesoderm (d, g, and i, green arrows), specifically compared to sense controls (h and n). Furthermore, from E9.5 to E11.5, it is also strongly expressed in the otic capsule (a, b, d, and i, short black arrows), although a faint signal was detected using the sense probe at E10.5 (h) but not at E9.5 (c) or E11.5 (n). Dashed red lines in d and i indicate the level of the sections shown in e and l, respectively, whereas the dashed black line in i depicts the level of section j. Small red arrowheads depict areas of probe trapping and background, as revealed by comparisons with sense controls (c, h, n).

FIG. 5.

Expression of mouse *Leprecan-like2* in whole-mount embryos at stages E9.5, E10.5, and E11.5. WMISH of *Leprel2* at stages E9.5 (a–c), E10.5 (d–i), and E11.5 (j–o) using *Leprel2*^{Sma} anti-sense and sense probes indicates that from E9.5 (b) to E10.5 (e and red inset shown in f), *Leprel2* is expressed in the developing notochord (black arrowheads). Interestingly, at E11.5, *Leprel2* is absent in anterior domains (k and black inset l, empty black arrowheads) and minimally expressed in the notochord at more caudal domains of the embryo (m and red inset n, gray arrowheads). At E11.5, expression appears markedly reduced across all axial levels compared to earlier gestational days. *Leprel2* is also expressed in numerous other domains including: at E9.5, forelimb and hindlimb buds (albeit at very low levels; a, short red arrows), branchial arch 1 (a, long red arrow), and paraxial mesoderm (b, green arrows); at E10.5, at very low levels in both limb types (d, short red arrows), in branchial arch 1 (d, long red arrow), and along the paraxial mesoderm (d, g and h, green arrows); and at E11.5, at higher levels, in both limb types (i, short black arrows), branchial arch and frontal nasal tissues (j, long black arrows), and in a reduced manner in the paraxial mesoderm (j, short red arrows). Dashed red lines in d and j indicate the level of the sections shown in e and m, respectively, whereas the dashed black line in j depicts the level of section k. Small red arrowheads depict areas of probe trapping and background, as revealed by comparisons with sense controls (c, i, o).

FIG. 6.

Section in situ hybridization of mouse *Leprecan* genes from stages E10.5 to E13.5. **(a')** Section in situ hybridization confirms that from E10.5 to E13.5 all three mouse *Leprecan* genes are expressed in the developing notochord (red insets and black arrowheads). All sections shown are at the lumbar-sacral level, except for the E13.5 section shown for *Leprel2*, which is more caudal (ca). Importantly, all three genes were expressed in additional domains, as at E13.5, *Leprecan* was expressed in the neural arches (black arrow) and laminae (red arrow) of the developing vertebrae, at E10.5 and E13.5 *Leprel1* was expressed in the neural tube (black arrows), and at E11.5–E12.5, *Leprel2* was expressed in the dorsal root ganglia (black arrows) and floor-plate (red arrow). **(b')** Additional domains of expression are revealed by sections at varying axial levels. *Leprecan* was expressed in the developing forelimb (a, d, black arrowheads), hindlimb (b, e, black arrowheads), pelvic girdle (e, black arrow), and vertebral mesenchyme (c, black arrowheads) and at E12.5 and E13.5 in vertebral cartilages, such as the centrum (f, red arrow), transverse processes (f, black arrows), laminae (f, black arrowhead), and neural arches (g, black arrowheads). *Leprel1*, while absent from forelimb (h, empty black arrowhead) and hindlimb buds (i, empty black arrowhead) at E12.5, was present at low levels in these tissues (k and l, red arrowheads respectively) and the pelvic girdle (l, red arrow) at E13.5. In addition, it was present in the olfactory epithelium (j, black arrowhead) at E12.5, as well as in the cartilage of the vertebral centra (m, red arrow) at E13.5. *Leprel2* was expressed in forelimb bud (n, black arrowhead), hindlimb bud (r, black arrowhead), mandibular pre-cartilaginous mesenchyme (o, black arrowhead), tissues of the eye (p, black arrowhead), olfactory epithelium (q, black arrowhead), neural arch cartilages (s, black arrowheads), dorsal root ganglia (s, black arrows), and the epithelium and nephrogenic mesenchyme of the developing kidney (t, red and blue insets, respectively, black arrowheads).

FIG. 7.

A possible sequence of events leading to the expansion in the number of *Leprekan* genes and in their expression domains. Model of the mechanisms that might have led to the appearance of different *Leprekan* genes with diverse expression domains. An ancestral single-copy, tissue-specific *Leprekan* gene, such as the notochord-specific gene found in *Ciona intestinalis*, served as a precursor for the multicomponent *Leprekan* family present in vertebrates. Through rounds of gene and genome duplications, multiple *Leprekan* genes arose and diversified independently into the three genes found in extant vertebrates. The three mouse genes, *Leprekan*, *Leprekan-like1*, and *Leprekan-like2* retained the ancestral notochord expression, but through the acquisition of novel cis-regulatory elements and/or the modification of pre-existing ones became additionally expressed in paraxial mesoderm, craniofacial and branchial arch, limb, and otic capsule tissues. *Leprekan-like2** indicates that the expression depicted for *Leprekan-like2* reflects a summary of the expression patterns observed using the *Leprel2^{Sma}*, *Leprel2^{Short}*, and *Leprel2^{Fl}* probes (see Methods section).