Mammalian Genome 9, 799–806 (1998).

Mammalian Genome
Incorporating *Mouse Genome*
© Springer-Verlag New York Inc. 1998

Evolutionary conservation and tissue-specific processing of *Hoxa 11* antisense transcripts

S. Steven Potter, William W. Branford

Division of Developmental Biology, Children's Hospital Medical Center and Department of Pediatrics, 3333 Burnet Ave., University of Cincinnati College of Medicine, Cincinnati, Ohio 45229-3039, USA

Received: 31 March 1998 / Accepted: 11 June 1998

Abstract. We previously described the existence of abundant, processed, polyadenylated murine *Hoxa 11* antisense transcripts. Of particular interest, in the developing limbs the antisense transcripts were observed to be present in a pattern complementary to that of the sense transcripts, suggesting a possible regulatory function (Hsieh-Li et al. 1995). We have analyzed the human HOXA 11 genomic locus, showing strong evolutionary conservation of regions potentially encoding antisense transcripts. Human HOXA 11 fetal kidney antisense cDNAs were identified and sequenced, demonstrating the evolutionary conservation of *Hoxa 11* antisense transcription. As for the mouse, the human antisense RNAs were polyadenylated and showed several alternative processing patterns, but shared the sequences of a common 3′ exon. The evolutionary conservation of the opposite strand transcripts strongly suggests function. A significantly long open reading frame was observed, but mouse-human comparisons argued against true coding function. Murine kidney *Hoxa 11* antisense transcription and processing was also examined, revealing tissue-specific differences between limb and kidney. A novel procedure, designated Race in Circles, was devised and used to define mouse limb antisense transcription start sites. Furthermore, comparisons of human, mouse, and chicken sense transcript *Hoxa 11* homeobox nucleotide sequences and their respective encoded homeodomains indicate a very strong selective pressure in vertebrates against mutations that result in coding changes. Given the significant differences in amino acid sequences of the homeodomains of different Hox genes, this observation argues for individual homeodomain functional specificity.

Introduction

The clustered homeobox (Hox) genes encode transcription factors that are known to function as master switch genetic regulators of development in *Drosophila,* defining segment identities by initiating genetic cascades (Gehring 1987). Comparisons of *Drosophila* and mammalian Hox genes have revealed a remarkable evolutionary conservation of sequence, gene order within the clusters, and expression patterns during development (Graham et al. 1989; Duboule and Dollé 1989). Furthermore, transgenic fly experiments, in which mammalian Hox genes have been introduced into flies, have strongly indicated a surprising conservation of function (Malicki et al. 1990).

We originally cloned the *Abdominal-B* type *Hoxa 11* gene of the mouse in a screen that identified ten novel homeobox genes (Singh et al. 1991). It was shown to be expressed in the developing limbs and caudal body, including the urogenital system (Small and Potter 1993; Hsieh-Li et al. 1995). Mice homozygous for a targeted mutation in *Hoxa 11* exhibited axial skeletal homeotic transformations and sterility in both males and females (Small and Potter 1993; Hsieh-Li et al. 1995; Gendron et al. 1997). The function of *Hoxa 11* was further defined by the generation of *Hoxa 11/Hoxd 11* double mutants, where synergistic effects were evident in the development of the axial skeleton, limbs, and kidneys. The kidneys, which were normal in single mutants, were absent or severely reduced in size in the double mutants. The forelimbs, which were only mildly malformed in each single mutant (Small and Potter 1993; Davis and Capecchi 1994), showed an almost complete loss of the zeugopod in the double mutants, with ulna and radius reduced to a mere vestige of normal (Davis et al. 1995). We have also previously reported characterization of the murine *Hoxa 11* locus itself, including cDNA sequence, transcription start site, and the sequence of the genomic DNA. Surprisingly, we observed the presence of abundant natural *Hoxa 11* antisense transcripts (Hsieh-Li et al. 1995).

Natural antisense RNAs are best understood in prokaryotic systems where they are documented to regulate bacteriophage genes, transposition rates of insertion elements, and plasmid replication incompatibility and conjugation (Simons 1988). In eukaryotic systems, antisense RNAs have been most widely studied as experimental tools to regulate expression of endogenous genes. However, several examples of natural endogenous antisense transcription have now been reported in eukaryotic systems (Farnham et al. 1985; Adelman et al. 1987; Lazar et al. 1989; Krystal et al. 1990; Dolnick 1993; Rivkin et al. 1993; Campbell et al. 1994; Tasheva and Roufa 1995). In some cases the antisense RNAs have been proposed to promote the degradation of the sense mRNAs (Kimelman and Kirschner 1989; Hildebrandt and Nellen 1992), and in other cases the antisense RNA has been suggested to block sense mRNA translation (Lee et al. 1993; Wightman et al. 1993). In the majority of cases, however, the function of the natural antisense RNA, if any, remains unknown.

We previously presented an initial characterization of antisense transcripts from the mouse *Hoxa 11* locus (Hsieh-Li et al. 1995). These antisense RNAs were found in an embryonic cDNA library screen with a probe generated from the 5′ region of the sense *Hoxa 11* cDNA sequences. Surprisingly, approximately 90% of the 72 cDNAs retrieved represented antisense transcripts. Several of the antisense cDNAs were sequenced, revealing a number of interesting features. The antisense cDNAs were found to be polyadenylated and alternately processed. Their antisense orientation was confirmed by the locations of the polyA tails and the directionality of the introns (locations of splice acceptor and splice donor sequences). The first exons of the sense and antisense cDNAs sometimes showed considerable sequence overlap of more than 500 bp. The antisense cDNAs also showed significantly long open reading frames, suggesting coding potential, although conceptual translation revealed no significant homologies in the database. In addition, in situ hybridizations were used to compare embryonic expression patterns of *Hoxa 11* sense and antisense RNAs. An in-

Correspondence to: S. Steven Potter

teresting complementarity of RNA distributions was seen in the developing limb. In the early limb bud at embryonic day 9.5 (E9.5), *Hoxa 11* sense RNAs are present throughout, while antisense RNA is absent. In contrast, in the E10.5 limb bud, *Hoxa 11* sense transcripts are no longer detected in the most distal domains of the developing limb. Strikingly, at this time antisense transcripts are now extremely abundant in the distal regions where sense RNA is absent. This trend continues later in development, with the domain of sense *Hoxa 11* transcripts relatively more proximally restricted and the distal domain of antisense transcripts expanding in corresponding fashion (Hsieh-Li et al. 1995). This suggested a possible regulatory function for the antisense RNA, since the presence of antisense RNA strongly correlated with the absence of sense RNA. Alternatively, the antisense transcripts could be spurious in nature and have no function, or they could have another, nonregulatory, function.

The next step in the study of the *Hoxa 11* antisense RNAs is to determine if they are of biological importance or mere accidental transcripts. Evolutionary conservation is considered one hallmark of functional significance. In this report we describe the sequencing of the human HOXA 11 genomic locus and its evolutionary comparison with that of the mouse. Regions potentially encoding antisense RNAs were observed to be evolutionarily conserved. A particularly striking 99% nucleotide sequence identity between mouse and human was observed for an approximately 500-bp stretch that included the region of overlap between the mouse sense and antisense transcripts. Surprisingly, the frequency of nucleotide mismatch in this region of sense-antisense overlap was fourfold lower than that observed in the homeobox, even though the mouse and human homeoboxes encode identical amino acid sequences. To confirm the existence of human HOXA 11 antisense RNAs, ten human fetal kidney antisense cDNAs were isolated and sequenced. Comparisons of mouse kidney and limb and human kidney antisense transcripts revealed tissue-specific differences as well as aspects of RNA processing conserved during evolution. These results strongly suggest that the antisense RNAs from the *Hoxa 11* locus are functionally significant.

Materials and methods

RT-PCR. RNA was prepared from E11.5 embryos obtained from superovulated females. All four limbs were removed and stored in a tube on dry ice. A total of about 1 g of limbs was collected. The limbs were then ground to a powder with a liquid nitrogen-cooled mortar and pestle. RNA was prepared with RNAzol (TEL-TEST Inc.) according to recommended protocols. The RNA was then extracted further with phenol, followed by a chloroform extraction and an ethanol precipitation, to remove all enzyme contaminants. The RNA was then treated with RNase-free DNase (30 U) in 300 μl of 40 mM Tris pH 7.0, 6 mM $MgCl_2$, 2 mM $CaCl_2$ for 30 min at 37°C. The RNA was then phenol extracted twice, chloroform extracted, ethanol precipitated, pelleted, dried, and dissolved in H_2O. PolyA^+ RNA was prepared with the Oligotext kit from Qiagen. Superscript II (Gibco-BRL, Gaithersburg) was used for reverse transcription, with the recommended protocol, with 0.5 μg of polyA^+ RNA and 0.5 μg of random hexanucleotide primer. After the reverse transcription, the reaction was treated at 37°C for 30 min with 2 U RNase H and 10 U RNAce-It (Stratagene, La Jolla, Calif.). The products were size fractionated in low-gel temperature agarose, and DNA from about 300 bases to 5 kb was purified with two phenol extractions, one chloroform extraction, and ethanol precipitation. This material was then used directly for PCR reactions or for the Race In Circles procedure.

Nested PCR used first the 3′-1 primer (CTCGGACTTGGCCTTTTGGTGTGCTCTTAT) (bases 1261–1290; Hsieh-Li et al. 1995) from the most 3′ exon of the 3C cDNA and the 23-2 primer (CTCTTGGCCAGCACGTCGCCAGGCA) (bases 5039–5015) from the first exon of the 3C cDNA. The second round of PCR reactions used the 3′-2 primer (CTTCCCCAGATCCTGGTGGGCTGAAATCAA) (bases 1303–1332) from the last 3C exon and the 23-3 primer (TGCAGACAGTCTCTGTGCACGAGCTCCT) (bases 4994–4966) from the first 3C exon. Products were gel purified, subcloned, and sequenced.

Race in circles. RNA preparation and initial reverse transcription, with random hexanucleotide primers, were performed as described above. The single-stranded DNA products were then kinased with polynucleotide kinase in 20 μl T4 RNA ligase buffer. The products were diluted to 100 μl with T4 RNA ligase buffer, 60 U of T4 RNA ligase was added, and the reaction was allowed to proceed at 37°C for 4 h. The dilute reaction conditions help promote circle formation and reduce intermolecular ligations. The circle products were then used directly for nested PCR reactions, using first the 23-2 primer and the 23pp primer (CCAACGTCTACCACCACCCCAC) (bases 5045–5066; Hsieh-Li et al. 1995) and then the 23-2 primer and the 23-1 primer (CCCGCCGTCTCGTCCAATTTCTATAGCA) (bases 5067–5095). Later experiments, looking at more 5′ start sites, used the 3Cpp (ACTTCAAGTTCGGACGGCGGCGGT) (bases 5278–5301) and 3C-3 (TCGCAGCT-GCTGCTTGGGGCCCCTTCT) (bases 5233–5207) primers, followed by the 3C-2 (TGTCAGGAGGCGGCGGCGGAGGAGAAGGAG) (bases 5311–5340) and 3C-4 (GG-GTAGTCGGAGGAAGCGAGGTTTTCCGGG) (bases 5189–5159) primers. Products were subcloned and sequenced.

Mouse kidney cDNA library. The mouse kidney cDNA library was made with polyA^+ RNA prepared from K4 kidney cells grown in culture. This cell line was established from transgenic mice carrying the *Hoxa 11* promoter connected to the SV40 large T antigen gene (M.T. Valerius et al., in preparation). PolyA^+ RNA was prepared by the RNAzol (TEL-TEST Inc.) method, followed by Oligotext (Qiagen) purification. First-strand cDNA synthesis was performed with Superscript II (Gibco-BRL) with dT primer, and RNA was digested with 2 U RNase H and 10 U RNace-It (Stratagene). Second-strand synthesis was performed with Klenow, using as primers a combination of random hexamers and residual dT primers, followed by polishing of the ends with T4 DNA polymerase. Products >500 bp were size selected in low-gel temperature agarose and ligated into pBS KSII. Transformation was by electroporation, yielding approximately 5×10^5 independent clones with inserts.

Sequencing. Sequencing was done according to standard protocols, with the Applied Biosystems ABI 377. Both strands of DNA were uniformly sequenced with primer walking. Difficult regions and areas of disagreement between the two strands were subjected to additional sequence runs with more closely positioned primers.

A human HOXA 11 cosmid clone was obtained through a screen of a human genomic DNA library kindly provided by Anil Menon. The HOXA 11 region was subcloned into pBS KSII and sequenced as described above.

Results

Alternate processing of the Hoxa 11 *antisense transcripts.* Considerable variation in murine *Hoxa 11* antisense RNA processing was shown previously by the sequencing of the four antisense cDNAs labeled 40, 23A, 59, and 3C in Fig. 1A (Hsieh-Li et al. 1995). To identify possible additional splicing patterns, we prepared RNA from E11.5 limb buds, and nested RT-PCR was performed with primers specific to the first (oligos 23-2 and 23-3) and last (oligos 3′-1 and 3′-2) exons of the 3C antisense cDNA. The RT-PCR products were subcloned and sequenced. The most common splicing pattern observed was identical to that seen for the 3C cDNA, with the same intron-exon junctions. In addition, alternate patterns were found. In one case the use of an alternate splice acceptor site increased the size of the second exon by 26 bases (8-2, Fig. 1A). It was also observed that a novel exon, located between 3C exons 2 and 3 was sometimes included in the final product. This exon also had two alternate splice acceptor sites, located three bases apart, with the 8-2 exon being three bases longer than the corresponding exon for clone 1. It is interesting to note that the two alternate forms of exon 2 differ by 26 bases, which means they are frame-shifted relative to each other, while the two forms of the novel exon differ by three bases.

Mouse kidney cDNAs. We previously used in situ hybridization to observe the presence of antisense *Hoxa 11* transcripts in the developing kidneys and reproductive tracts as well as limbs. To

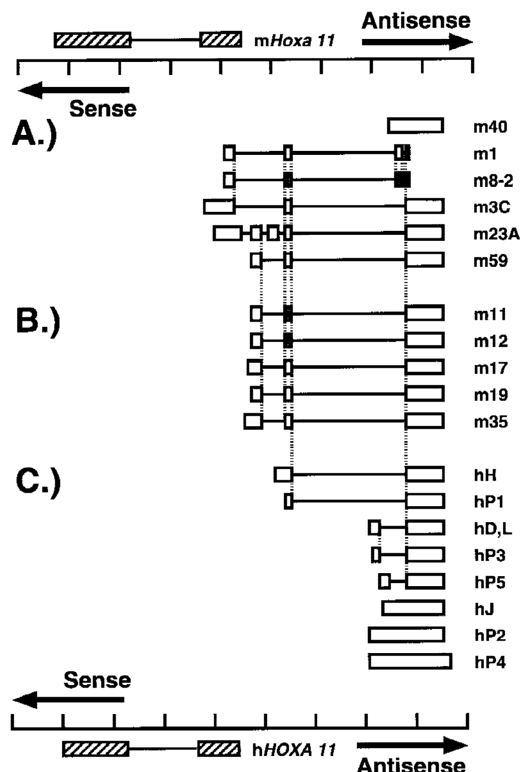

Fig. 1. Alternate splicing and evolutionary conservation of *Hoxa 11* antisense RNA. Rectangles represent exons, and connecting lines represent introns. The mouse *Hoxa 11* sense exons are shown at the top, and the vertically aligned human HOXA 11 sense exons are at the bottom. As illustrated by the arrows, sense RNAs are transcribed to the left, and antisense RNAs are transcribed to the right. **(A)** Clones m3C and m23A were isolated from a mouse E12 limb cDNA library. Clones m40 and m59 were from an E12.5 total embryo mouse cDNA library. Clones m1 and m8-2 were generated by RT-PCR with mouse limb bud RNA (see text for details). **(B)** Mouse kidney *Hoxa 11* antisense cDNAs, generated from a cloned kidney cell line. **(C)** Human fetal HOXA 11 kidney antisense cDNAs. Shaded rectangles represent exons with distinct splice acceptor sites. Vertical dashed lines mark common splice sites, except that the shaded second and third exons have alternate splice acceptor sites that increase exon sizes by 26 and 3 bases, respectively. The h and m prefixes designate human and mouse clones.

address the issue of tissue specificity of *Hoxa 11* antisense transcription and processing, a mouse kidney cDNA library was generated from K4 cells, a mouse kidney cell line (M.T. Valerius, et al. in preparation; see Materials and methods). Five mouse kidney antisense cDNAs were recovered by probing with 3C antisense cDNA sequences. Fig. 1B shows the exon organization of these cDNAs. Each cDNA contains the sequence of the common final exon of the 3C, 23A, and 59 cDNAs, and the second exons are either the same as the 3C second exon or the alternatively spliced form of this exon in 8-2. For each kidney cDNA the splice donor site of the first exon fell at the boundary of bp 4225/4226 of the published sequence (Hsieh-Li et al. 1995), although each of the 5′ ends of these cDNAs varied somewhat, owing either to incompleteness of the cDNAs and/or distinct transcription start sites. Of particular interest, none of these ‘‘antisense’’ transcripts actually overlapped the region encoding the sense transcript, in contrast with the extensive overlap often observed for the limb cDNAs. For the mouse kidney, therefore, these transcripts are more accurately referred to as opposite strand rather than antisense.

Evolutionary conservation of the Hoxa 11 locus. The *Hoxa 11* antisense transcripts could have biological function, or they could represent spurious transcripts with no functional significance. To begin to address this issue, the nucleotide sequence of the human HOXA 11 locus was determined. One common feature of functional significance is evolutionary conservation. It was, therefore, of interest to look for conservation of the sequences encoding the *Hoxa 11* antisense cDNAs.

A human genomic DNA cosmid library was screened with a mouse *Hoxa 11* cDNA probe, and the HOXA 11 region of a positive clone was subcloned and sequenced. The sequence of the human HOXA 11 locus, including sufficient flanking DNA to total 8,624 bp, was determined. Much of the human HOXA 11 sense cDNA sequence can be inferred from comparison with the previously described mouse *Hoxa 11* genomic and cDNA sequences (Hsieh-Li et al. 1995). In addition, two HOXA 11 sense cDNAs were isolated from a human fetal kidney cDNA library and sequenced at the 5′ ends. One initiated at a major transcription start site previously identified in the mouse (Hsieh-Li et al. 1995). The other cDNA was incomplete at the 5′ end.

A comparison of the human and mouse sense *Hoxa 11* 5′ UTR and coding sequences is shown in Fig. 2. A striking evolutionary conservation of sequence is revealed for the 5′ regions of the cDNAs. For the first 500 bases of sequence there are only seven bases of mismatch. This approximately 99% sequence identity at the 5′ end is particularly striking when compared with the level of homology observed for the homeobox. In comparing orthologous homeobox genes from very distantly related organisms, such as Drosophila and mouse, it is not uncommon to observe that significant homology is restricted to the homeobox itself. In general, there appears to be evolutionary pressure to maintain the amino acid sequence of the encoded homeodomain, perhaps consistent with its important DNA binding function. In comparing the mouse and human *Hoxa 11* genes, it was observed that there are 10 base differences in the 180-bp homeobox. All ten of these differences occur at codon third nucleotide positions, and none alter the amino acid encoded. This is consistent with a rigid evolutionary requirement to maintain the precise amino acid sequence encoded by the homeobox. Nevertheless, the frequency of base mismatches observed in the homeobox (10/180) is four times higher than that seen for the first 500 bases of the cDNA sequence (7/500). That is, the mismatch frequency for the first 500 bases is fourfold lower than that observed for the homeobox, even though the two homeoboxes encode identical amino acid sequences. It is interesting to note that the first five hundred bases of the human HOXA 11 cDNA includes 70 bases of 5′ UTR, which do not contribute to the coding of the Hoxa 11 protein. There is an in-frame translation termination codon at base 32 of the 5′ UTR, confirming that it is genuine UTR and not coding sequence mis-identified as UTR. This high-level conservation of the first 500 bases of the *Hoxa 11* cDNA strongly argues that it is under more selective pressure than simple maintenance of *Hoxa 11* coding function. This additional selective pressure could be related to function associated with antisense transcripts, at least some of which overlap the sense transcript for most of the first 500 bases.

Evidence for surprising conservation of the 5′ end of the HOXA 11 transcript is also seen in comparing the human and chicken sequences. The human and chicken HOXA 11 180-bp homeobox sequences showed 29 nucleotide differences. Of interest, all of the base changes were silent, not changing amino acids encoded, with 26 of 29 found in codon third nucleotide positions. In contrast, the first 200 bases of the 5′ end of the known chicken cDNA differ from the human at only 14 positions, or less than half the mis-match frequency of the homeobox. It is interesting to note that the 5′ region of extreme conservation with the human sequence appears more restricted in the chicken than that observed in mouse, with only approximately 200 bases instead of 500. Also, the mis-match frequency for the 5′ end in chicken is only about one half that observed for the homeobox, while in the mouse comparison a fourfold difference was seen. Nevertheless, once

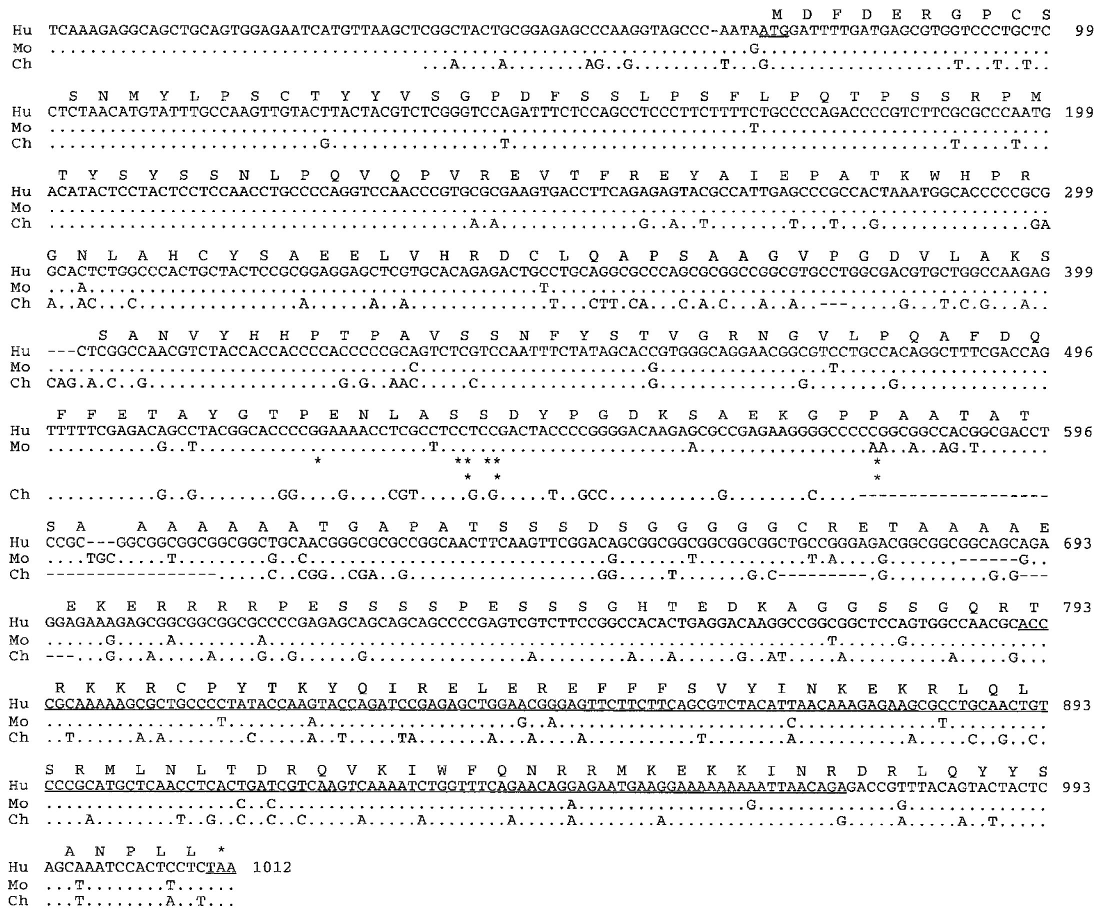

Fig. 2. Comparison of human, mouse, and chicken 5′ UTR and sense coding HOXA 11 sequences. Hu is human, Mo is mouse, and Ch is chicken. Dots represent sequence identity with the human sequence, and dashes designate a deletion. The ATG translation start site, the homeobox, and the TAA translation termination codon are underlined. Antisense transcription start sites observed in the mouse limb are marked with an asterisk. The 5′ sequences are, surprisingly, even better conserved than the homeobox.

again, the greater conservation of the chicken *Hoxa 11* 5′ cDNA end, as compared with the homeobox, suggests the presence of more than just *Hoxa 11* coding function.

The human HOXA 11 genomic locus, including the single intron and 5′ flanking sequence, which includes the potential antisense exons, showed strong sequence homology with mouse *Hoxa 11.* A Pustell DNA matrix comparison (MacVector; minimum % score of 65 and remaining parameters at default) generated an almost continuous line from one end of the sequence to the next (data not shown). Increasing the stringency of the comparison to a minimum % score of 85 revealed extensive blocks of high homology (Fig. 3). This homology includes, but is not restricted to, sequences previously observed in mouse antisense cDNAs. Of interest, both the human and mouse *Hoxa 11*/HOXA 11 genomic DNA regions appear devoid of transposable repetitive DNA. A screen for all repeat sequences, with the program of the Genetic Information Research Institute, revealed only a few scattered di- and trinucleotide repeats. As transposable element repeat DNA is quite common, in particular the *Alu* sequences in human DNA and the *B1* and *B2* repeats in the mouse, this significant absence of such repeats in a total of about 17 kb of DNA in the combined mouse *Hoxa 11* and human HOXA 11 sequences suggests that mobile genetic element insertions into these regions are not well tolerated.

Transcription start sites for the mouse Hoxa 11 antisense RNAs. The definition of the *Hoxa 11* sense-antisense transcript overlap region relied on the antisense transcription start sites inferred from cDNAs. To better describe the most commonly used start sites for mouse *Hoxa 11* antisense transcription, a new technique, designated Race in Circles (RIC), was devised. In principle, RIC resembles standard primer extension, except that the reverse transcription products are circularized, instead of run on a gel, allowing the 5′ ends to be PCR amplified and sequenced. In outline form, RNA is first reverse transcribed with random hexamer primers. This results in random reverse transcription start sites and termination at the 5′ ends of the RNAs. Under dilute conditions, to reduce intermolecular ligations, the resulting single-stranded DNA is circularized with T4 RNA ligase. The junction regions, where the 5′ ends of the reverse transcripts are ligated to the 3′ ends, are then amplified by an inverse PCR approach with primers in both directions from near the 5′ end of the transcript. The PCR products are then cloned and sequenced (see Materials and methods for details). The junction points mark the 5′ ends of the RNAs.

As is often observed for promoters lacking a TATA box, there appears to be a preferred region of antisense transcription initiation instead of a single precise start site (Farnham et al. 1985). Sequencing of nine PCR products showed seven start sites within a five base region centered at base 5180 of the published *Hoxa 11* sequence (Hsieh-Li et al. 1995), which corresponds to base 540 of the human-mouse *Hoxa 11* cDNA comparison of Fig. 2. Antisense transcription start sites were located within the first exon of the sense *Hoxa 11* gene, clustered around base 540, as shown in Fig. 2. More 5′ start sites appear to be used rarely, since none of these ten RIC defined start sites were positioned more 5′ than bp 579. Nevertheless, the 3C antisense cDNA starts at base 705, demon-

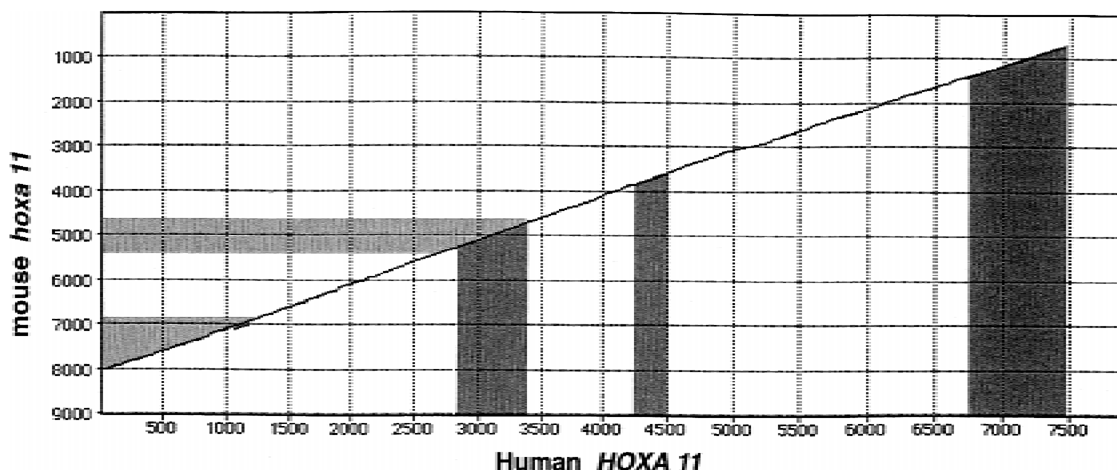

Fig. 3. Pustell DNA matrix comparison of the mouse *Hoxa 11* and human HOXA 11 genes. The matrix was generated with MacVector with a window size of 30, minimum % score of 85, and remaining values at default. Light gray marks the regions corresponding to sense exons, and dark gray shows the regions corresponding to the mouse 3C antisense cDNA. The two genomic DNAs show extensive homology, which is not restricted to sense and antisense exons.

strating that more 5′ sites can sometimes be used. To further examine the transcription starts at more 5′ positions, a second set of PCR primers located 5′ of the most common transcription start sites were used (see Material and methods for precise primer positions). These primers selectively amplified RNAs that start 5′ of position 700. Fifteen RIC products were sequenced, revealing potential start sites at bases 709 (three), 712 (two), 713, 727, 735, 740, 745, 755, and 765. The three remaining RIC products extended more 5′, but all three spliced at 776, where the intron for the sense transcript is located, indicating that they represent sense and not antisense products. In summary the results indicate that the most abundant antisense transcription start sites cluster around base 540 of Fig. 2, but that in addition there are multiple, less frequently used start sites that are scattered to almost the 3′ end of the sense Hoxa 11 first exon.

The 3′ region of the sense Hoxa 11 gene first exon, which serves as the promoter region for antisense transcription, has no TATA box. It is, however very GC rich, with 75% of the last 230 bp being GC. This region carries numerous GC boxes that represent potential SP1 binding sites, which are often associated with TATAless promoters.

Human HOXA 11 antisense cDNAs. To search for human HOXA 11 antisense transcripts, we screened a human fetal kidney cDNA library with a mouse 3C third exon antisense cDNA probe. Ten antisense cDNA clones were recovered and sequenced. The results are summarized in diagrammatic form in Fig. 1C. Although these cDNAs represent transcripts oriented in the opposite direction of the standard sense HOXA 11 transcript, it is interesting to note that, as observed for the mouse kidney *Hoxa 11* ‘‘antisense’’ cDNAs, they do not overlap the sense cDNAs in sequence. In addition, four of the cDNAs have no introns. One clone, not shown, was a duplicate of J. The P2 and P4 cDNAs share a common 5′ start site but show distinct 3′ ends. The six spliced human kidney antisense cDNAs share a common 3′ exon, with the same splice acceptor site. This exon is also common to all examined spliced mouse *Hoxa 11* antisense cDNAs. In contrast to the mouse antisense cDNAs, these human kidney antisense cDNAs had only two exons. The first exon of the P1 and H cDNAs overlap the second exon of the mouse 3C cDNA. These exons all share the same 3′ splice donor site. The human P1 clone sequence starts within the sequence of the second 3C exon, while the sequence of the human H clone starts upstream of the second exon of the 3C clone and extends directly into it. Three frame shift mutations in the human versus mouse versions of this exon were found. The D and L human kidney antisense cDNAs are duplicates, except that they differ by one base at their 3′ ends. These clones and the P3 cDNA share common splice sites, while the P5 clone uses a unique splice donor site.

Two polymorphisms were found in comparing the human cDNA and genomic DNA sequences. One polymorphism was located 116 bases into the common 3′ exon shared by all of the spliced human HOXA 11 antisense RNAs. For the genomic sequence this base was a G, yet on two cDNAs, clones J and H, this base was an A. The second polymorphism was 333 bases further into this same 3′ exon. This base was a C in the genomic DNA and was absent in the P4 cDNA. This frame-shifting polymorphism suggests an absence of coding function for this region of the antisense RNA.

There are open reading frames in the human HOXA 11 antisense RNAs that could represent coding function. As shown in Fig. 4, almost the entire 3′ exon is open reading frame, from the beginning to the TAA termination codon within the AATAAA polyadenylation signal. The first methionine codon in this reading frame occurs 268 bases into the exon, giving the potential to encode 135 amino acids. The unspliced forms of the antisense cDNAs (J, P2, and P4) do not extend this coding potential, with in-frame stop codons slightly upstream of the start of the 3′ exon. The D and P3 spliced forms, however, do extend the coding capacity. These two cDNAs connect an upstream methionine codon in frame to the 3′ shared exon, generating a 705-base open reading frame potentially encoding 235 amino acids, as shown in Fig. 4. The comparison of the 3′ antisense exons of mouse and human, however, argues against coding function. In the first 500 bases of this exon, there are six frame-shifting single base pair insertions/deletions located at positions 10, 187, 233, 273, 394, and 402. Also consistent with no coding function, there is not a strong tendency for missense mutations to be located at third nucleotide positions of codons in order to preserve amino acid sequence encoded.

Discussion

Evolutionary conservation of Hoxa 11 antisense transcripts. The characterization of the human HOXA 11 locus revealed a striking evolutionary conservation of sequence, antisense transcription, and processing that strongly suggests functional significance. This also defined the human HOXA 11 sense cDNA sequence and genomic

```
    M  P  G  H  L  P  Q  R  H  A  A  P  L  L  I  R  S  V  F  P  W  C  P  G  E  R  S  R  Q  K
Hu ATGCCAGGCCACCTCCCCCAGCGGCACGCAGCCCCTCTCTTAATTAGATCGGTTTTCCCCTGGTGTCCGGGAGAGCGGTCCCGGCAGAAAG//

   G  G  P  T  A  E  L  S  P  P  G  S  G  K  D  P  T  E  A  K  A  A  P  E  G  Q  V  R  V  P  F  L  E  E
Hu GAGGTCCAACAGCCGAGCTTAGCCCACCGGGCTCTGGGAAAGACCCGACTGAGGCTAAAGCCGCCCCGGAAGGCCAAGTCCGAGTTCCATTTCTTGAAGA 100
Mo ....C....-G.TT..TT.C........A..A......G.......A...AG..A...GAG.A.A..AA........................A.....

    A  G  A  R  K  A  V  T  L  A  L  A  T  G  F  P  G  A  V  L  S  Q  E  L  H  S  A  G  H   L  Q  K  T
Hu GGCCGGCGCGCGTAAGGCTGTGACATTGGCCCTGGCGACTGGCTTCCCAGGAGCTGTTCTTTCTCAGGAGCTCCACAGCGCGGGCCA-TCTCCAGAAAAC 199
Mo ...G.CTT.CG.C......A..............AC.T...T............C.T......A...........A....G.TG...........

    V  F  R  V  Y  F  L  L  S  S  T  Q  S  P  T  A  A  N  A  R  G  Q  K  M  F  G  G  R  K  T  K  A  G
Hu TGTCTTCAGAGTGTATTTCCTTTTATCGTCAACCCAGAGCCCCACCGCGGCTAATGCAAGAGGCCAAAAAATGTTTGGAGGAAGAAAAACAAAGGCAGGA 299
Mo GC......GC..T......T...A......-G...G......T.AG.......................-...........................

    S  G  G  G  L  T  V  R  V  C  L  Q  R  R  E  G  A  G  S  V  S  S  C  F  S  K  L  Q  G  P  G  S   P
Hu AGTGGCGGCGGCCTGACGGTGCGTGTGTGTCTGCAGAGAAGGGAGGGAGCCGGCTCAGTCTCTTCTTGTTTTTCCAAACTTCAAGGTCCAGGCAG-CCCT 398
Mo ......C.........................T.A.................T...AC...CC..C.....C..G...........T......A...C

    L  Q  G  R  A  P  L  L  P  A  R  H  W  R  W  P  P  G  E  E  K  A  N  A  C  A  R  L  V  R  R  K  R
Hu CTGCAGGGCCGGGCCCCATTGCTCCCCGCGCGGCATTGGAGGTGGCCGCCCGGAGAGGAGAAGGCCAACGCCTGCGCCAGGCTTGTCAGGCGGAAACGGC 498
Mo ..T-......TT...GAGGCT.G....CACACCC.CA.CG.CGCAG.ATTT....GT.GCC.ACGATTTAAGCCTCGGTC.GGCTGA.A.GA..TTT.AT

   L  T  R  R  F  G  Q  Q  N  R  P  S  L  S  E  A  S  S  D  L  A  R  K  V  G  E  G  G  L  A  Q  S  L  R
Hu TAACAAGGAGATTTGGTCAGCAAAACAGACCCAGCCTTTCCGAGGCTTCGTCTGACTTGGCCCGAAAGGTTGGGGAGGGGGGGCTTGCGCAGAGCCTCAG 598
Mo CGG..GAACA.ACCAAC.CTTTTCGG..GTTTCTTT.GATTTG.T.C.AAAGG.TA.AT..TA.TGTCCACA.C.GCT....TGGCTGCTGTTTT.CTCC

    D  P  P  L  W  G  L  P  S  L  S  L  T  L  L  S  T  A  F  A  G  G  I  S  E  *
Hu GGACCCTCCTCTCTGGGGACTACCATCCCTGAGCCTTACGCTTCTTTCCACAGCCTTTGCAGGCGGAATATCGGAATAAA
Mo C.C.GGG.TAAAAGTACC.AG.AGGGAGGGAG.GAGAGA.A...AGG.AC.TTG.GC..GCT..ACTC.CCTTCTGAG.TAGAATACCAGAATAAA
```

Fig. 4. Open reading frame for the HOXA 11 antisense transcript. Predicted amino acid sequence is shown above the nucleotide sequence. Hu is human and Mo is mouse. Dots represent identities, and dashes deletions. The top nucleotide sequence line shows the 3′ end of the first exons of the human D and P3 cDNAs (Fig. 1). No corresponding exon for mouse has been detected. The remaining lines represent the most 3′ antisense exon. Although the open reading frame is significantly long, the multiple frame shift differences in the mouse-human comparison, and the absence of codon third nucleotide mutation preference, argue against coding function.

organization. Comparison of the sequences including and surrounding the mouse and human *Hoxa 11*/HOXA 11 genes revealed strong conservation and an absence of transposable repetitive DNA.

To address the issue of possible organ specificity of transcription and processing of *Hoxa 11* antisense transcripts, mouse kidney antisense cDNAs were isolated and sequenced. The relative uniformity of the mouse kidney antisense cDNAs is worthy of note. These cDNAs all shared the same three exons, but with alternate usage of the splice acceptor site for the second exon and different apparent transcription start sites in the first exon. Such uniformity could reflect the source of mRNA used to make this cDNA library. While the mouse limb cDNA library was made from tissue, the mouse kidney library was made from a clonal kidney cell line. These observations suggest, but do not prove, that part of the heterogeneity of antisense RNA transcripts previously seen could be the result of heterogeneity of cell types present in the tissue examined.

The identification and sequencing of ten human fetal kidney HOXA 11 antisense cDNAs demonstrated that the antisense transcripts are indeed evolutionarily conserved. Furthermore, the sequences defined appear to restrict possible mechanisms of action, at least in the kidney. As observed in the mouse, these kidney cDNAs were transcribed from the opposite DNA strand, but did not overlap the sense HOXA 11 transcripts in sequence. This argues that in the developing kidney it is unlikely that the opposite strand RNAs function by duplexing to the sense HOXA 11 mRNA.

There were distinct similarities in exon usage between the human and mouse antisense transcripts. In particular, the last exon of human clones H, D, L, P5, P3, and P1 corresponded to the last exon of all of the spliced forms of the mouse antisense cDNAs, with the same splice acceptor site. In addition, the first exon of human clones H and P1 partially corresponded to the second exon of the 3C mouse cDNA, with the same splice donor site.

Comparisons of the sequences of the human and mouse antisense cDNAs do not provide significant support for coding function. The second exon of the 3C cDNA, for example, has multiple frame shift differences when compared with the corresponding first exons of human clones H and P1, arguing against coding function. Likewise, the 3′ exon region, common to all of the antisense RNAs, also showed multiple scattered human-mouse frameshifts. Furthermore, the base mismatch distribution did not show codon third nucleotide position preference, again arguing against coding function. Comparisons revealed some conserved open reading frames, but they were not dramatically long and did not encode proteins that showed significant homologies to other proteins in the database. It is difficult to disprove any coding function. The data presented in this report, however, argues that any protein encoded is likely small.

It is also interesting to note that the antisense cDNAs with sense sequence overlap occur in the limb, where we previously observed a dramatic complementarity in the domains of sense and antisense RNAs (Hsieh-Li et al. 1995). In the early (E9.5) limb bud the sense transcript is abundant throughout and the antisense transcript is absent. One day later, at E10.5, antisense RNAs are very abundant in the distal bud, immediately flanking and indeed somewhat overlapping the expression domain of the now more proximally restricted *Hoxa 11* sense RNA. This complementarity in transcript domains suggests that the limb bud antisense RNA could have a regulatory role, perhaps duplexing with and thereby driving the degradation of the sense RNAs. In the kidneys, however, where we have observed no complementarity in expression domains, there is no apparent sequence overlap between sense and antisense RNAs.

The very strong evolutionary conservation of the 5′ ends of the *Hoxa 11* sense cDNAs, which in the mouse limb overlap in sequence with the 5′ ends of antisense cDNAs, suggest more than just *Hoxa 11* coding function for these regions. This hyper-

conservation of nucleotide sequence, beyond that necessary to preserve identical coding capacity, suggests that in chicken and human there may also be overlapping antisense transcripts that remain to be found. As observed in the mouse, these overlapping antisense RNAs may be tissue specific, and present, for example, in the limb but not in the kidney. The 5′ sequence hyperconservation could reflect the presence of essential cis-regulatory elements within the first exon, the requirement to preserve coding function on both strands, or a requirement for perfect sense-antisense duplex formation, as discussed further below.

A brief consideration of other examples of vertebrate antisense RNAs places the *Hoxa 11*/HOXA 11 results reported here in perspective and suggests possible functions. Farnham and coworkers (1985) reported the presence of opposite strand RNAs from the dihydrofolate reductase (*dhfr*) gene. The *dhfr* antisense transcripts, however, were without polyA tails, did not appear to be processed, and were small (180–240 nucleotides) and limited to the nucleus. In many respects, therefore, the opposite strand transcripts of *dhfr* are distinct from those of *Hoxa 11*/HOXA 11.

Several other examples of endogenous antisense transcripts have been reported in vertebrate systems. Krystal and associates (1990) described antisense transcripts for the *N-myc* locus. Again there was the presence of a promoter without a TATA box, leading to a range of initiation sites for both sense and antisense transcripts. Of particular interest, this study showed the presence of in vivo sense-antisense RNA duplexes. It is also interesting to note that both Lee and colleagues (1993) and Wightman and coworkers (1993) have shown that in *C. elegans* small noncoding RNAs from one locus can have antisense complementarity to coding RNAs from another locus and through this complementarity can exert regulatory function, apparently at the translational level. The work of Kimelman and Kirschner (1989) in their study of the *bFGF* gene in *Xenopus* provides further precedent for possible antisense regulatory function. They showed the presence of a processed, polyadenylated antisense transcript with significant sequence overlap with the *bFGF* gene. Their data indicated duplex formation between sense and antisense RNAs, followed by sequence modification and RNA degradation. The *bFGF* antisense RNA has also been shown to be conserved during evolution (Volk et al. 1989).

Perhaps the *Hoxa 11*/HOXA 11 antisense RNA, as proposed for *bFGF*, regulates the stability of the sense RNA. Consistent with this, the enzyme that modifies dsRNA molecules, double-stranded RNA-specific adenosine deaminase, is ubiquitous (Wagner et al. 1990). Moreover, the striking (~99% nucleotide sequence identity) evolutionary conservation of the first 500 bases of the 5′ end of the mouse/human *Hoxa 11*/HOXA 11 sense cDNAs, representing the bulk of the region of overlap with the 3C antisense transcript, suggests more than simple *Hoxa 11* coding function. Lipman (1997) has pointed out that extreme conservation of noncoding sequences and at silent base positions in coding regions could reflect a requirement for long, perfect sense-antisense duplexes to properly regulate mRNA stability. In heterozygotes a single base change would result in a 50% frequency of imperfect heteroduplex formation and subsequent dominant negative effects on gene regulation.

Nevertheless, in the kidney no sequence overlap with *Hoxa 11*/HOXA 11 sense transcripts was detected, arguing for another function besides sense-antisense duplex formation. Perhaps the *Hoxa 11*/HOXA 11 opposite strand transcript function is similar to the roles of the *H19* and *Xist* noncoding RNAs, believed to be important in genomic imprinting and X Chromosome (Chr) inactivation, respectively (Pachnis et al. 1988; Brannan et al. 1990; Ripoche et al. 1997; Clemson et al. 1996; Penny et al. 1996; Lee et al. 1996; Lee and Jaenisch 1997). Although there is no evidence of genomic imprinting for the Hox clusters, it remains possible that *Hoxa 11*/HOXA 11 antisense RNA functions in *cis* to regulate gene expression patterns.

Homeodomain functional specificity. It has been proposed that the many different mammalian Hox genes may be largely functionally equivalent (Duboule 1995). Unlike *Drosophila* homeobox genes, which appear capable of initiating distinct developmental programs, it has been suggested that the vertebrate Hox genes function primarily in controlling cell proliferation rates. This interesting model is consistent with many experimental observations. For example, the often observed alterations in contour shapes of bones in vertebrate Hox mutants are perhaps more easily interpreted as the result of cell proliferation changes rather than homeotic transformations of segment identity. In extreme form this model states that the various vertebrate Hox genes ‘‘may not even engage the regulation of qualitatively different target genes’’ (Duboule 1995). That is, the homeodomains of different Hox proteins may all bind identical or functionally equivalent gene targets involved in regulating cell proliferation.

In this report we describe the results of an extensive mutagenesis analysis conducted by nature, which argue that the Hox-encoded homeodomains are not functionally equivalent. In comparing the human HOXA 11 homeobox 180 bp nucleotide sequence with that from mouse, we observed ten base differences. A similar comparison of orthologous human HOXA 11 and chicken homeoboxes showed 29 nucleotide differences. The striking result, however, was that all 39 of the combined nucleotide differences were silent, maintaining identical homeodomain sequences. Mutations in nucleotide sequence can be presumed to occur in random fashion, and despite the redundant nature of the genetic code, a majority of nucleotide sequences changes will alter the amino acid encoded. Yet chicken, mouse, and human encode identical Hoxa 11 homeodomains. Only silent mutations survived the stringent test of natural selection. This argues that there is very tight selective pressure maintaining the Hoxa 11 homeodomain amino acid sequence. This is particularly interesting in light of the considerable variation observed in the amino acid sequences of different homeodomains. For example, the *Hoxa 10* gene is very closely related to *Hoxa 11. Hoxa 10* and *Hoxa 11* are both *Abd-B* type genes that are adjacent to each other in the A cluster of Hox genes. Nevertheless, their encoded homeodomains differ at 19 of 60 amino acids. Other, less closely related Hox genes show even greater amino acid sequence divergence with *Hoxa 11.* Yet the Hoxa 11 homeodomain appears rather tightly frozen in sequence during evolution, with the human-chicken and human-mouse comparisons showing 39 silent versus zero missense mutations surviving. There is no detected freedom to drift in sequence at the multiple variable amino acids of different homeodomains. This suggests that a change of Hoxa 11 homeodomain amino acid sequence, even towards sequence found in other homeodomains, perturbs function.

Acknowledgments. We thank Ryan Walsh for assistance in the early phases of this work. We thank Anil Menon for providing the human genomic cosmid library. We thank members of the Potter laboratory for helpful reading of the manuscript. We thank Jan Hagedorn for assistance in preparing the manuscript. This work was supported by National Institutes of Health grant NIH HD32061.

References

Adelman JP, Bond CT, Douglass J, Herbert E (1987) Two mammalian genes transcribed from opposite strands of the same DNA locus. Science 235, 1514–1517

Brannan CI, Dees EC, Ingram RS, Tilghman SM (1990) The product of the *H19* gene may function as an RNA. Mol Cell Biol 10, 28–36

Campbell CE, Huang A, Gurney AL, Kessler PM, Hewitt JA, et al. (1994) Antisense transcripts and protein binding motifs within the *Wilms tumor (WT1)* locus. Oncogene 9, 583–595

Clemson CM, McNeil JA, Willard H, Lawrence JB (1996) *XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear chromosome structure. J Cell Biol 132, 1–17

Davis AP, Capecchi MR (1994) Axial homeosis and appendicular skeleton defects in mice with a targeted disruption of *hoxd-11.* Development 120, 2187–2198

Davis AP, Witte DP, Hsieh-Li HM, Potter SS, Capecchi MR (1995) Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. Nature 375, 791–795

Dolnick BJ (1993) Cloning and characterization of a naturally occurring antisense RNA to human *thymidylate synthase* mRNA. Nucleic Acids Res 21, 1747–1752

Duboule D (1995) Vertebrate Hox genes and proliferation: an alternative pathway to homeosis? Curr Opin Genet Dev 5, 525–528

Duboule D, Dollé P (1989) The structural and functional organization of the murine *Hox* gene family resembles that of *Drosophila* homeotic genes. EMBO J 8, 1497–1505

Farnham PJ, Abrams JM, Schimke RT (1985) Opposite-strand RNAs from the 5′ flanking region of the mouse *dihydrofolate reductase* gene. Proc Natl Acad Sci USA 82, 3978–3982

Gehring WJ (1987) Homeo boxes in the study of development. Science 236, 1245–1252

Gendron RL, Paradis H, Hsieh-Li HM, Lee DW, Potter SS, et al. (1997) Abnormal uterine stromal and glandular function associated with maternal reproductive defects in *Hoxa-11* null mice. Biol Reprod 56, 1097–1105.

Graham A, Papalopulu N, Krumlauf R (1989) The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. Cell 57, 367–378

Hildebrandt M, Nellen W (1992) Differential antisense transcription from the *Dictyostelium EB4* gene locus: implications on antisense-mediated regulation of mRNA stability. Cell 69, 197–204

Hsieh-Li HM, Witte DW, Weinstein M, Branford W, Li H, et al. (1995) *Hoxa 11* structure, extensive antisense transcription, and function in male and female fertility. Development 121, 1373–1385

Kimelman D, Kirschner MW (1989) An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in *Xenopus* oocytes. Cell 59, 687–696

Krystal GW, Armstrong BC, Battey JF (1990) N-*myc* mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. Mol Cell Biol 10, 4180–4191

Lazar MA, Hodin RA, Darling DS, Chin WW (1989) A novel member of the thyroid/steroid hormone receptor family is encoded by the opposite strand of the rat c-*erbA*α transcriptional unit. Mol Cell Biol 9, 1128–1136

Lee JT, Jaenisch R (1997) Long-range *cis* effects of ectopic X-inactivation centres on a mouse autosome. Nature 386, 275–279

Lee JT, Strauss WM, Dausman JA, Jaenisch R (1996) A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell 86, 83–94

Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14.* Cell 75, 843–854

Lipman DJ (1997) Making (anti)sense of non-coding sequence conservation. Nucleic Acids Res 25, 3580–3583

Malicki J, Schughart K, McGinnis W (1990) Mouse *Hox-2.2* specifies thoracic segmental identity in *Drosophila* embryos and larvae. Cell 63, 961–967

Pachnis V, Brannan CI, Tilghman SM (1988) The structure and expression of a novel gene activated in early mouse embryogenesis. EMBO J 7, 673–681

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996) Requirement for *Xist* in X chromosome inactivation. Nature 379, 131–137

Ripoche M-A, Kress C, Poirier F, Dandolo L (1997) Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. Genes Dev 11, 1596–1604

Rivkin M, Rosen KM, Villa-Komaroff L (1993) Identification of an antisense transcript from the *IGF-II* locus in mouse. Mol Reprod Dev 35, 394–397

Simons RW (1988) Naturally occurring antisense RNA control—a brief review. Gene 72, 35–44

Singh G, Kaur S, Stock JL, Jenkins NA, Gilbert DJ, et al. (1991) Identification of 10 murine homeobox genes. Proc Natl Acad Sci USA 88, 10706–10710

Small KM, Potter SS (1993) Homeotic transformations and limb defects in *Hox A11* mutant mice. Genes Dev 7, 2318–2328

Tasheva ES, Roufa DJ (1995) Regulation of human *RPS14* transcription by intronic antisense RNAs and ribosomal protein S14. Genes Dev 9, 304–316

Volk R, Köster M, Pöting A, Hartmann L, Knöchel W (1989) An antisense transcript from the *Xenopus laevis bFGF* gene coding for an evolutionarily conserved 24 kd protein. EMBO J 8, 2983–2988

Wagner RW, Yoo C, Wrabetz L, Kamholz J, Buchhalter J, et al. (1990) Double-stranded RNA unwinding and modifying activity is detected ubiquitously in primary tissues and cell lines. Mol Cell Biol 10, 5586–5590

Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans.* Cell 75, 855–862