Elsevier Editorial System(tm) for BBA - Gene Regulatory Mechanisms
Manuscript Draft

Manuscript Number: BBAGRM-15-28R2

Title: MicroRNA-196b is transcribed from an autonomous promoter and is directly regulated by Cdx2 and by posterior Hox proteins during embryogenesis

Article Type: Regular Paper

Keywords: transcriptional regulation;
Hox genes;
Cdx genes;
post-transcriptional control mechanisms;
developmental biology

Corresponding Author: Prof. Vincenzo Zappavigna,

Corresponding Author's Institution: University of Modena and Reggio Emilia

First Author: Sebastian Fantini, PhD

Order of Authors: Sebastian Fantini, PhD; Valentina Salsi, PhD; Antonio Vitobello, PhD; Filippo M Rijli, PhD; Vincenzo Zappavigna

Abstract: The miR-196 miRNA gene family located within the Hox gene clusters has been shown to function during embryogenesis and to be aberrantly expressed in various malignancies, including leukemia, melanoma, and colorectal cancer. Despite its involvement in numerous biological processes, the control of miR-196 expression is still poorly defined. We identified the miR-196b promoter and found that the mature miR-196b originates from a large, non-coding primary transcript, which starts within an autonomous TATA box promoter and is not in physical continuity with either the Hoxa10 or Hoxa9 main primary transcripts. A ~300 bp genomic fragment, spanning the pri-miR-196b transcription start site, is sufficient to recapitulate the neural tube expression pattern of miR-196 during embryogenesis. This region contains potential binding sites for Cdx and 5'Hox transcription factors. Two of these sites revealed to be necessary for neural tube expression and were bound in vivo by Cdx2 and Hoxd13. We show that Cdx2 is required for miR-196 expression and that both Cdx2 and 5'Hox, but not 3'Hox, are able to activate the miR-196b promoter. The possible role of Cdx2- and 5'Hox-mediated regulation of miR-196 expression in vertebrate anterior-posterior (AP) axis formation during embryogenesis is discussed.

Response to Reviewers:

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
DEPARTMENT OF LIFE SCIENCES
Via Campi 213/D, 41100 MODENA, Italy
☎ **059 205.5544 – Fax 059 205.5548**

Dr. Joseph Reese
Executive Editor
BBA - Gene Regulatory Mechanisms

Dear Dr. Reese,

enclosed please find a re-revised version of manuscript BBAGRM-15-28 titled: "MicroRNA-196b is transcribed from an autonomous promoter and is directly regulated by Cdx2 and by posterior Hox proteins during embryogenesis" by Sebastian Fantini, Valentina Salsi, Antonio Vitobello, Filippo M. Rijli, and myself.

We were happy to realize that the reviewers found our revised version of the manuscript suitable for publication with very minor corrections and that you have appreciated our efforts to produce a thoroughly revised manuscript. We would therefore like to thank you for your expert handling of the manuscript. We have now produced re-revised manuscript, which contains the minor changes that you requested (details on the cat. numbers of the antibodies used and their concentrations, see marked manuscript). We have also uploaded a corrected version of Fig. 6 according to the comment of reviewer#2.

Hoping that our re-revised manuscript is suitable for publication in BBA - Gene Regulatory Mechanisms, we look forward to hearing from You at Your earliest convenience.

Sincerely,

Prof. Vincenzo Zappavigna
Department of Life Sciences
University of Modena and Reggio Emilia
Via Campi 213/D
41100 Modena
ITALY
Tel. 0039 059 2055537
Fax 0039 059 2055548
E-mail zappavigna.vincenzo@unimore.it

HIGHLIGHTS

- We identified and characterized the promoter of the Hox-embedded miR-196b.
- miR-196b is originated from an autonomous transcription unit.
- A conserved genomic region recapitulates miR-196b developmental expression pattern.
- Cdx2 and 5'Hox binding sites are required for neural tube expression.
- Cdx2 and 5'Hox bind in vivo the miR-196b promoter and regulate its activity.

MicroRNA-196b is transcribed from an autonomous promoter and is directly regulated by Cdx2 and by posterior Hox proteins during embryogenesis

Sebastian Fantini°, Valentina Salsi°, Antonio Vitobello[#,*], Filippo M. Rijli[#] and Vincenzo Zappavigna°

°Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 213/d, Modena 41125, Italy; [#] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland; [*] Present address: Preclinical Safety, Novartis Institutes for Biomedical Research, CH-4057 Basel, Switzerland

Address correspondence to: Vincenzo Zappavigna, Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 213/d, Modena 41125, Italy, Tel. 0039 059 2055537, Fax 0039 059 2055548, E-Mail: zappavigna.vincenzo@unimore.it

ABSTRACT

The *miR-196* miRNA gene family located within the Hox gene clusters has been shown to function during embryogenesis and to be aberrantly expressed in various malignancies, including leukemia, melanoma, and colorectal cancer. Despite its involvement in numerous biological processes, the control of *miR-196* expression is still poorly defined. We identified the *miR-196b* promoter and found that the mature *miR-196b* originates from a large, non-coding primary transcript, which starts within an autonomous TATA box promoter and is not in physical continuity with either the *Hoxa10* or *Hoxa9* main primary transcripts. A ~680 bp genomic fragment, spanning the *pri-miR-196b* transcription start site, is sufficient to recapitulate the neural tube expression pattern of *miR-196* during embryogenesis. This region contains potential binding sites for Cdx and 5'Hox transcription factors. Two of these sites revealed to be necessary for neural tube expression and were bound in vivo by Cdx2 and Hoxd13. We show that Cdx2 is required for *miR-196* expression and that both Cdx2 and 5'Hox, but not 3'Hox, are able to activate the *miR-196b* promoter. The possible role of Cdx2- and 5'Hox-mediated regulation of *miR-196* expression in vertebrate anterior-posterior (AP) axis formation during embryogenesis is discussed.

1. INTRODUCTION

MicroRNAs (miRNAs) are evolutionarily conserved, small non-coding RNAs of ~20-24 nucleotides (nt) in length, which negatively regulate gene expression at the post-transcriptional level. They typically bind to the 3' UTR of their target mRNAs, and either inhibit translation or induce mRNA degradation (1, 2). miRNAs derive from larger primary miRNAs (pri-miRNAs), in many cases non-coding, which are processed by endonucleases, to precursors (pre-miRNAs) in the nucleus and eventually to their mature form in the cytoplasm. miRNAs participate in the regulation of a vast array of biological processes, ranging from cell cycle regulation to differentiation and development. Consequently, the deregulation of miRNA expression is implicated in several human diseases including cancer (3). The mechanisms controlling miRNA gene expression, especially those regarding the transcription of their pri-miRNA precursors, have been only partially elucidated. These studies reveal a primary role for RNA polymerase II (RNA polII) in miRNA transcription (4), and that many miRNA genes use their own promoter and transcriptional regulatory elements, even if located within gene clusters (intergenic) or within introns of coding genes (intragenic) (5). Still less known are the transcription factors directing miRNA gene transcription. Their knowledge, however, would greatly increase our understanding of miRNA gene deregulation in disease, and would allow to define novel gene regulatory networks, comprising miRNAs, their transcriptional regulators, and their target mRNAs, thus connecting transcriptional and post-transcriptional controls in single pathways.

The *miR-196* gene family is evolutionarily conserved, having arisen in a common ancestor of chordates and urochordates, and includes three genes, all located within the *HOX* gene clusters (6). In the human genome, the *miR-196a-1* gene is located in the intergenic region between *HOXB9* and *HOXB10* on chromosome 17, *miR-196a-2* between *HOXC9* and *HOXC10* on chromosome 12 (7). *miR-196b* is located between *HOXA9* and *HOXA10* on chromosome 7 in human, and on chromosome 6 in mice, in a region evolutionarily conserved among vertebrates (7). *miR-196a-1* and *miR-196a-2* share the same mature, functional miRNA sequence, while *miR-196b* differs from the *miR-196a* sequence by one nucleotide (8). *miR-196* family members have been reported to participate in relevant biological processes such as embryonic development (reviewed in 9) and tumorigenesis where they have been found to be aberrantly expressed in various malignancies, including glioblastoma, melanoma, leukemia, and colorectal cancer (reviewed in 7).

Among all putative miR-196 targets, found by in silico prediction, *Hox* mRNAs appear to be the class of transcripts that are preferentially targeted by the miR-196 gene family (10). Indeed, if the percentage of conserved miRNA target sites within the 3' UTRs of *Hox* transcripts is taken into account, the *miR-196* gene family ranks first among all conserved miRNA families (10). The

"central" *Hoxa7*, *Hoxb8*, *Hoxc8*, and *Hoxd8* genes, expressed in the developing trunk during embryogenesis, have been demonstrated experimentally to be conserved targets of miR-196 (9,11,12).

Hox homeodomain-containing transcription factors have been shown to play crucial roles in embryonic development by controlling determination of cell fate, motility, adhesion, proliferation, and apoptosis, eventually orchestrating the generation of different morphological identities along the primary body and limb axes (reviewed in 13,14). Alterations of HOX gene expression and/or structure have been moreover found to be associated with tumorigenesis (reviewed in 15,16). In addition to *Hox* genes, the *ParaHox* family *Cdx* genes (17) are relevant to anterior-posterior axis patterning and elongation during embryogenesis, in part also as upstream regulators of *Hox* genes (18-21).

The remarkably close proximity of miR-196 and its most relevant target genes has prompted speculations about its biological significance. It has been proposed that HOX-embedded miRNAs (*miR-10* and *miR-196*) may play a role in the phenomenon termed "posterior prevalence". Posterior prevalence, represents one of the characteristic features of the *Hox* gene system (reviewed in 13). It accounts for the fact that more posteriorly-expressed *Hox* genes, such as the *AbdB*-related, 5' located Hox (5'Hox), are functionally dominant over anteriorly-expressed genes, as single loss-of-function *Hox* mutants display phenotypes only in the most anterior regions of their expression domains (22). The putative role of Hox cluster-linked miRNAs in posterior prevalence rests on the observation that their *Hox* target mRNAs are asymmetrically distributed with respect to their location within the HOX clusters. For instance, the majority of *Hox* target mRNAs of *miR-196* are located 3' to the *miR-196* loci, thus the activity of *miR-196* would repress expression of *Hox* genes located 3' to its locus, while sparing 5' located *Hox* (reviewed in 10).

Both loss- and gain-of-function experiments have shown that the miR-196 gene family participate in developmental processes such as limb A-P, nervous system, and axial skeletal patterning (reviewed in 6). The majority of these studies have focused on the capability of *miR-196* to negatively regulate the expression of *Hoxb8* (23-26), a key representative of "central" Hox genes, expressed in developing trunk regions of the embryo. *miR-196* has been shown to reinforce the repression and to contribute to the clearance of the *Hoxb8* gene product in caudal regions of the embryo, to define the posterior boundary of *Hoxb8* function (23,24). In more anterior regions where it is co-expressed with *Hoxb8*, *miR-196* has been shown to increase the robustness of the *Hoxb8*-controlled genetic program by eliminating fluctuations in its expression (25). Finally, overexpression of *miR-196* in zebrafish causes transformations of anterior vertebrae identities and the failure of pectoral fin bud initiation (27).

Despite its roles in development and cancer (reviewed in 7) the regulation of *miR-196* gene expression has never been thoroughly characterised (see 6). Still unknown are its promoters, their upstream regulators as well as the regulatory elements to which these bind. We thus set out to characterize the miR-196b promoter, and to explore the possibility that it might be controlled by upstream regulators of the Hox gene clusters and/or by Hox gene products themselves possibly as part of the mechanism(s) underlying posterior prevalence. We found that the mature *miR-196b* originates from a large, 2671 bp, primary, non-coding transcript (*pri-miR-196b*), which is, unlike previously assumed, not in physical continuity with either the *Hoxa10* or *Hoxa9* main primary transcripts, and starts within an autonomous TATA box promoter. The evolutionarily conserved genomic region upstream to the *pri-miR-196b* transcriptional start site (TSS) was found to contain several potential binding sites for Cdx and 5'Hox transcription factors. We establish that a ~680 bp genomic fragment spanning the *pri-miR-196b* TSS is sufficient to recapitulate the neural tube expression pattern of *miR-196* and that two of the analysed sites are necessary in vivo for *pri-miR-196b* promoter function. We show that these sites are bound in vivo by the Cdx2 and Hoxd13 proteins in posterior regions of E10 mouse embryos. Cdx2 and 5'Hox, but not 3'Hox, were found to transactivate the *pri-miR-196b* promoter and Cdx2 to be necessary for the expression of *miR-196* in human embryonal carcinoma (EC) cells. We discuss the possible significance of the regulation of *miR-196* expression by Cdx2 and/or 5'Hox in the context of vertebrate anterior-posterior (AP) axis formation.

2. MATERIALS AND METHODS

2.1. Plasmid constructs

The *pri-miR-196b* murine promoter sequence was obtained by PCR amplification from genomic DNA of P19 cells using the following primers: 196b(-249)for: 5'-AAAAGCTGTGGAATGAAGC-3'; 196b(+624)rev: 5'-CTCCGGAAAGGTACGCCTAG-3'. The amplified fragment was sequence-verified and cloned into the HindIII site of the pXP2-luciferase vector (28) to obtain the p196b(-249;+624) reporter construct. A 680-bp fragment, encompassing sites III and IV, and V, was PCR-amplified using the following primers: 196b(-171)for: 5'-GAGGGCTAGCCCGAGCCC-3'; 196b(+508)rev: 5'-CGGTGTGAGGCGGGTTCC-3'. The fragment was verified by sequencing and cloned into the HindIII site of pXP2 to generate p196b(-171;+508). The p196b(-249;+624)M3 mutant was generated via splicing by overlapping extension (SOE), using the wild type promoter as a template and mutating site IIIab using the following primers: 196b(M3)for 5'-TGCCTTTAGGCATCTATTTAAATATGCCCTAGACGGTCGGCATT-TGTCTGGGCCCTATAG-3'; 196b(M3)rev: 5'-ATGCCGACCGTCTAGGGCATATTTAAATA-GATGCCTAAAGGCATTGCAGGGGGCTCGGG-3'. Starting from the wild type promoter and from the p196b(-249;+624)M3 constructs, respectively, p196b(-249;+624)M4 and p196b(-249;+624)M3/4 were generated using the following primers: 196b(M4)for: 5'-TTTGTTCAGGGGG-CGTCATTGCTAGTATATCTGGATTG-3'; 196b(M4)rev: 5'-TACTAGCAATGACGCCCCCTG-AACAAAAGCCGACAAACC-3'. All mutant fragments were cloned into the HindIII site of pXP2.

The promoter sequences described above were also cloned into the pGL3basic luciferase reporter vector (Promega) using the same strategies. The reporter constructs used in electroporation experiments were generated using the pNASSβ reporter vector (Clontech). They were generated by cloning the PCR fragments described above into the EcoRI-blunted site of pNASSβ. Following the same strategy, the wild-type promoter was cloned in addition into pNASSβ in antisense orientation.

Expression vectors for PBX1a, HOXB1, HOXD8, HOXD10, and HOXD13 were described previously (29-31). The pSG-FHOXA9 expression construct was obtained by PCR amplification of the human HOXA9 coding sequence, which was cloned into the BamHI site of the pSG5-FLAG vector.

2.2. Whole mount in situ hybridization (WISH)

WISH was performed using the method described in (32) with some minor modifications. Embryos at 17-25 Hamburger-Hamilton (HH) (33) stages were dissected in PBS and fixed

overnight in 4% paraformaldehyde. Embryos were then permeabilized with proteinase K and hybridized with 5nM miR-196 digoxygenin-labeled LNA probe (Exiqon) in hybridization buffer (50% formamide, 5X SSC pH 7, 0,1% Tween-20, 50 µg/mL total yeast RNA, 50 µg/mL heparin, 0,1% CHAPS, 5mM EDTA and 2% blocking powder) overnight at 44°C.

2.3. Chromatin immunoprecipitation (ChIP)

ChIP assays were performed as described in (34). Briefly, formaldehyde-crosslinked chromatin was extracted from caudal trunk and forebrain dissected from ~200 E10 mouse embryos, and subjected to immunoprecipitation using 1 µg anti-me3K4 (Abcam, cat#ab8580), 5 µg anti-me3K27 (Abcam, cat#ab6002), 5 µg anti-AcK27 (Abcam, cat#ab4729), 5 µg anti-HOXD11 (Abcam, cat#ab60715) (35), 5 µg anti-HOXD13 (Abcam, cat#ab19866), 5 µl anti-CDX2 (Biogenex, cat#MU392A-UC clone Cdx2-88), 5 µg anti-HOXA9 (Upstate, cat#07-178) (36) and 5 µl anti-HOXD10 (37), 5 µg anti-FLAG (F3165, Sigma, clone M2) was used as control antibody. The enrichment of amplified DNA sequences in immunoprecipitates was calculated as a percentage of total input chromatin. Mean enrichments for each ChIP were assessed using at least three independent ChIP quantitative real-time PCR (qPCR) results. Primers used in qPCR analysis were: miR196b F_: 5'-GAGCCCCCTGCAATTACTTT-3'; miR196b R_: 5'-CAAGACAGCCACAATCCAGA-3'; CTRL1 F_: 5'-CTCCGCCGCTCTCATTCTC-3'; CTRL1 R_: 5'- ACGCTTGACACT-CACACTTTG-3'; CTRL2 F_: 5'-GGAGGGATTCAAAGTCAGTGTC-3'; CTRL2 R_: 5'-CTGCCTT-ATCTTGGAGCCATC-3'.

2.4. RRL, nuclear extracts, and EMSAs

CDX2 and human 5'HOX proteins were synthesized *in vitro* using the TNT-coupled transcription/translation system (Promega). In EMSAs the total amount of reticulocyte lysate utilized was adjusted to normalize for translated protein content, based on the amount of ^{35}S-Met incorporated and on the number of Met residues for each protein. The amount of the control, "empty" RRL used corresponded to the highest amount of total RRL used in the other binding reactions for each experiment. The binding reactions were performed as described previously by (38). Briefly, proteins were pre-incubated with 100 ng poly-(dI-dC) in 1x binding buffer (0.1 M KCl, 2 mM MgCl2, 4 mM spermidine, 0.1 mg/ml BSA) for 15 minutes on ice. 20,000-50,000 cpm of ^{32}P-labelled probe were then added and samples were incubated for 30 minutes on ice. For supershift experiments, 2 µg of the indicated antibodies, or 2 µL αCDX2 ascite (Biogenex), were added to the binding reactions. In competition assays, a 500-fold excess of cold probe was used.

Reactions were separated on 6% polyacrylamide gels in 0.5x TBE at 4°C. The oligonucleotides probes used are:

SITE I: 5'-GAAGCTTTTTAATTTCACCTCTTTTTTGACTTCTTATCTTCACCTTC-3';

SITE IIab: 5'-CACCTTCCATTTTATCCTGTCCACCACTTTTACAACAGGAGGGC-3';

SITE IIIab: 5'-CTGCAATTACTTTAGGCATCTATTTAAATATTACCTAGACGGTCGT-AATTTGTCTGG-3';

SITE IV: 5'-CGGCTTTTGTTCAGTTTTATGACTTGCTAGTATATCTGG-3';

SITE V: 5'-CGCAGGAGCCAGTTATTTTGCGGGTATCCAGGTCCC-3'.

SITE IIa MUT: 5'-CACCTTCCAT**GGGCGA**CTGTCCACCACTTTTACAACAGGAGGGC-3'

SITE IIb MUT: 5'-CACCTTCCATTTTATCCTGTCCACCAC**GGGGCA**AACAGGAGGGC-3'

SITE IIab MUT: 5'-CACCTTCCAT**GGGCGA**CTGTCCACCAC**GGGGCA**AACAGGAGGGC-3'

SITE IIIa MUT:

5'CTGCAAT**GCA**TTTAGGCATCTATTTAAATATTACCTAGACGGTCGTAATTTGTCTGG-3'

SITE IIIb MUT: 5'-

CTGCAATTACTTTAGGCATCTATTTAAATAT**GCA**CTAGACGGTCGTAATTTGTCTGG-3'

SITE IIIab MUT: 5'-

CTGCAAT**GCA**TTTAGGCATCTATTTAAATAT**GCA**CTAGACGGTCGTAATTTGTCTGG-3'

SITE IV MUT: 5'-CGGCTTTTGTTCAG**GGGGC**GTCATTGCTAGTATATCTGG-3'

Nuclear extracts from HEK293 cells were prepared by lysing them in cytoplasmic buffer (10mM HEPES pH 7,9, 1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT) and subsequent extraction in nuclear buffer (20mM HEPES pH 7.9, 420mM NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM PMSF and 0.5mM DTT) (39).

2.5. 5' and 3' rapid amplification of cDNA ends (RACE)

For 5' RACE experiments, total RNA was isolated from posterior trunk of E10.5 CD1 mouse embryos by using TRIzol (Invitrogen) and treated with RQ1 DNAse (Promega) before the reverse-transcription step with 196b-5'-GSP1. A tail of dCTP was then added to the cDNA by using terminal deoxynucleotidyl transferase (TdT, Promega). The tailed DNA was initially amplified with a dG-adapter primer and 196b-5'-GSP2 and finally nested-amplified with an abridged primer (AUAP) anchoring to the adapter and 196b-5'-GSP3 primer. The human *miR-196b* promoter was investigated with the same strategy using three human gene-specific primers and RNA from HEK293 cells.

In 3'-RACE experiments, cDNA was retrotranscribed with dT-anchor primer annealing on the poly-A tail of the pri-miR-196b, initially amplified with AUAP and 196b-3'-GSP1 and finally nested-amplified with the AUAP and 196b-3'-GSP2 primers. Both 5'- and 3'-RACE products were visualized by gel electrophoresis , cloned into the pBluescipt SK+ vector and sequence verified.
The primers used in RACE experiments are listed below:
196b-5'-GSP1: 5'-TGATGCTGCGGGCGTAGGG-3';
dG-adapter: 5'-GGCCACGCGTCGACTAGTACGGGGGGGGGGGGGGG-3';
196b-5'-GSP2: 5'-GGAAACGCGTGGCCTAGCGG-3';
AUAP: 5'-GGCCACGCGTCGACTAGTAC-3';
196b-5'-GSP3: 5'-CCCTCCTCCGCGGATTTCGGG-3'.
hsa 196b-5'-GSP1: 5'-CAGTAAGGTATTCCTGGGAGG-3';
hsa 196b-5'-GSP2: 5'-ACTCTGGCAAGCAGGAAACG-3';
hsa 196b-5'-GSP3: 5'-ATGCTGCAGAGAGATGCGCTGTGGGCCAG-3';
dT-anchor primer: 5'-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTT-3';
196b-3'-GSP1: 5'-GTCGCCCGACTACTGAGAAC-3';
196b-3'-GSP2: 5'-ATGCTGCAGTGGATGCAACCACACAACAC-3'.

2.6. Cell culture, transfection and luciferase assays

P19 cells (ATCC# CRL-1825) mouse embryonal carcinoma cells were cultured in Mininum Essential Medium Alpha (αMEM, Gibco) with nuleosides. HEK293 (ATCC# CRL-1573) and NTERA-2 (ATCC# CRL-1973) human cells were cultured in Dulbecco modified Eagle medium. All the media were supplemented with 10% fetal bovine serum, 2mM L-glutamine, 100U/mL penicillin and 100µg/mL streptomycin. NTERA-2 neuronal differentiation was induced by treatment with 10^{-5} all-trans retinoic acid. Transfections were carried out by $CaPO_4$ precipitation. In a typical experiment, P19 cells were transfected with 2,5-6 µg of reporter plasmid, 0,12-0,25 µg of expression construct and 0,1 µg of pCMV-β-Gal (Clontech) per 6cm dish. Cells were washed and 48 hours after transfection lysed and assayed for luciferase and β-galactosidase as described in (30).

2.7. In ovo electroporation

Chicken eggs were incubated in a humidified incubator at 37°C. Embryos at stage HH10-12 were injected and electroporated into the neural tube as described by (40) Construct concentrations were as follows: 1.0 µg/µL for pNASSβ::lacZ reporter construct and 0.5 µg/µL

for pCMV::eGFP, coinjected as positive control of electroporated tissues. Embryos were collected 24 hrs after electroporation and subjected to β-galactosidase staining.

2.8. siRNA depletion of CDX2

Three different small interfering RNA (siRNA) duplexes were designed targeting the human CDX2 mRNA (Stealth siRNA; Life Technologies). siRNA and CDX2 transfection were performed in induced NTERA-2(RA14) with 100nM RNAi duplexes using Lipofectamine 2000 (Invitrogen) according to the manufacturer instructions. CDX2 depletion and overexpression were verified by immunoblotting 48 h after siRNA transfection. Whole cell extracts were prepared through repeated freeze and thaw cycles and extraction in 10mM Hepes PH7.9, 100mM NaCl, 0,1mM EGTA, 5% glycerol and proteases and phosphatases inhibitors. Immunoblotting analyses were performed with anti-CDX2 antibody (Biogenex, cat#MU392A-UC clone Cdx2-88) and anti-α-Tubulin (Sigma, T6074, clone B-5-1-2). Protein expression in immunoblots was quantified using the ImageJ software.

2.9. RNA extraction and miRNA RT-qPCR detection

Total RNA from dissected embryonic tissues or cultured cells was isolated by TRIzol (Invitrogen) according to the manufacturer instructions. Mature miR-196b and the endogenous control RNU6B (Suppl. Fig. 1B) were reverse transcribed and detected in qPCR with TaqMan® microRNA assays (Applied Biosystems). Real-time (qPCR) relative quantification was performed according to the ΔΔCt method (41). Synthesis of cDNA was done starting from 3 µg of total RNA using M-MLV reverse transcriptase (Promega). Quantitative PCR was performed with the following oligonucleotides: hCDX2 F: 5'-GTGCTAAACCCCACCGTCAC-3'; hCDX2 R: 5'-CTGAGGAGTCTAGCAGAGTC-3'; hACTB F: 5'-AGGCACCAGGGCGTGAT-3'; hACTB R: 5'-GCCCACATAGGAATCCTTCTGAC-3'.

Statistical analysis

Results are expressed as mean ± SEM from at least triplicate experiments. Statistical analyses were performed with GraphPad Prism version 5.0. In transfection experiments the non-parametric Kruskal-Wallis with by Dunn's post test were used. ChIP statistical analyses were performed with the Student's t test.

3. RESULTS

3.1. miR-196 displays a dynamic expression pattern along the anterior-posterior axis during chick embryogenesis

miR-196 expression during chick embryonic development was analysed by whole-mount in situ hybridisation (WISH) using a locked nucleic acid (LNA)-modified oligonucleotide probe (Exiqon). At Hamburger-Hamilton (HH) (33) stage HH17 *miR-196* expression was relatively weak and confined to the caudal-most regions of the embryo within the neural tube and flanking tissue (Fig. 1 panels A and B). At HH stage 18, the expression of *miR-196* was consistently increased and extended more anteriorly, with an expression boundary located posterior to the hindlimb buds (Fig. 1 panels C and D). The maximal *miR-196* expression levels were reached at HH stage 19, where *miR-196* was uniformly expressed in the posterior chick embryo including the hind limbs (Fig. 1E and F). At this stage the anterior-most boundary of expression was found to lie within the posterior trunk, approximately at the level of somites 27-28. At later stages (HH 20-24) *miR-196* expression remained high with an anterior boundary within the lumbar region (somites 27-28, at HH 24). At HH24, within the neural tube, the anterior expression boundary was consistently more rostral (by approx. 3-4 somites) than in the flanking tissues (Fig. 1 I and J). At this stage, a possible additional, weaker expression domain could be detected at the posterior margin of the forelimbs (Fig. 1 I and J). At HH stage 25 *miR-196* expression was still elevated in the whole caudal portion of the embryo (Fig. 1 K and L). In conclusion, *miR-196* expression revealed to be dynamic, both in time and spatial restriction, gradually increasing and extending towards more anterior territories from HH stage 17 onwards, to reach a peak of expression at stages HH 19-20. The posteriorly restricted expression of *miR-196* overlaps the previously reported embryonic expression of 5'Hox genes in the neural tube and flanking tissues (reviewed in 10,13,42), and of Cdx2 (43).

miR-196b expression during mouse embryogenesis was analyzed using real-time reverse transcriptase PCR (qRT-PCR) on dissected embryos at different developmental stages (Suppl. Fig. 1). At E8.5, the earliest stage analysed, expression was relatively low and confined to the posterior portion of the trunk. At later stages, *miR-196b* expression gradually increased, reaching its maximal expression at E10.5 in the posterior trunk (Suppl. Fig. 1). A low level of expression was observed also in the anterior trunk at E10.5 and at later stages (Suppl. Fig. 1). From E11.5 onwards the expression in the posterior trunk remained approximately constant up to stage E14.5 (Suppl. Fig. 1). These results parallel those obtained in chick embryos, confirming a dynamic, posteriorly-restricted expression of *miR-196b*.

3.2. The pri-miRNA giving rise to miR-196b is transcribed from an independent promoter located in the Hoxa10-Hoxa9 intergenic region

To characterise the murine primary miRNA (pri-miRNA) originating the mature *miR-196b*, we mapped its boundaries within the intergenic region between the *Hoxa10* and *Hoxa9* genes, located on chromosome 6, by 5' and 3' RACE (Fig. 2). Three different primers (5'GSP1, 5'GSP2, 5'GSP3, Fig. 2A) allowed us to identify the *pri-miR-196b* transcription start site (TSS) at a distance of 260 bp from the polyadenylation signal of the *Hoxa10* gene, and of ~780 bp from the mature *miR-196b* sequence (Fig. 2 and Suppl. Fig. 2). A putative TATA box (5'-TATATC-3', see Fig. 3B) was found to be located at 35 bp 5' from the mapped TSS, which itself is located within an Initiator sequence (5'-TCAGTT-3', Fig. 3B) (44). Finally, a CpG island was identified at 118 bp 3' to the TSS.

Similarly, the 3' end of the *pri-miR-196b* transcript was mapped, using two different primers (3'GSP1, 3'GSP2, Fig. 2 and Suppl. Fig. 2), and found to be located at a distance of ~920 bp from the *Hoxa9* gene transcription start site (TSS). A canonical polyadenylation signal (5'-AATAAA-3') was identified in close proximity (19 bp) of the mapped transcript 3' end (Fig. 2 and Suppl. Fig. 2), hinting at a possible polyadenylation of the *pri-miR-196b*. Interestingly, a RIKEN cDNA clone (45), deposited in GeneBank (AK054058) (Fig. 3A), covers most of the length of the mapped transcript further supporting the existence of a long non-coding transcript representing the *pri-miR-196b.*

These results suggest that *miR-196b* originates from a large 2671 bp transcript, starting within the Hoxa10-Hoxa9 intergenic region, which is not in physical continuity with either the *Hoxa10* or *Hoxa9* main primary transcripts (Fig. 3A). Our data also indicate that the *pri-miR-196b* is transcribed from an autonomous promoter, independent of the known *Hoxa10* or *Hoxa9* promoters.

3.3. The miR-196b promoter region contains potential responsive elements for Hox and/or Cdx proteins

To characterise the *miR-196b* promoter, and to identify possible relevant regulatory proteins controlling its activity, we performed an in silico analysis of 873 bp of the *Hoxa10-Hoxa9* intergenic region, spanning from the polyadenylation signal of *Hoxa10* to the TSS of *Hoxa9*. Binding sites for 5' (posterior) Hox and Cdx proteins, as defined in (38) (5'-TTTTATTGG-3' and 5'-TTT(T/A)ACGAG-3') and (46) (5'-A/CTTTATA/G-3'), were searched using the Matinspector software (47). We focused on 5' (posterior) Hox and Cdx proteins, as we hypothesized that these would negatively control the function of 3' located Hox gene products, via an indirect, post-

transcriptional mechanism based on the activation of *miR-196b* expression. In parallel, we performed an interspecies comparison using the UCSC genome browser (48), as we reasoned that evolutionarily sequence conservation might highlight relevant regulatory elements. We could thus identify seven evolutionarily conserved putative binding sites for 5' (posterior) Hox proteins and/or for Cdx proteins (sites I to V, Fig. 3B) within a region of ~330 bp encompassing the *pri-miR-196b* TSS (Fig. 3B). Of these sites IIb, IIIb, and IV showed full sequence conservation in all species analysed (Fig. 3B) whereas sites I, IIa, IIIa, and V showed a more limited conservation (Fig. 3B).

To test whether these sites could indeed bind 5' HOX proteins, we initially performed electrophoretic mobility shift assays (EMSAs) using the HOXD10, HOXD11, and HOXD13 5'HOX proteins produced in rabbit reticulocyte lysates (RRLs) (Fig. 4A). Of the five potential binding sites tested, sites IIIab and IV were bound strongly by HOXD10 and HOXD11, and to a lesser extent by HOXD13, whereas site IIab was bound weakly by all the three HOX proteins tested. Sites I and V showed no or only barely detectable binding (Fig. 4A). We next tested for binding of the CDX2 protein, a member of the CDX ParaHox family of transcription factors. CDX2 also showed a distinct preference for site IV (Fig. 4B) and a weak binding to site IIab (Fig. 4B). The binding of CDX2 to site IV was also verified in EMSAs using nuclear extracts from HEK293 cells (Fig. 4), a cell line which was found to endogenously express CDX2 (data not shown) and *miR-196b* (Fig. 4D). A retarded band, comigrating with the CDX2 shifted complex observed using RRL, was detected in HEK293 cells nuclear extract (Fig. 4C, lane 3). This complex was outcompeted by a molar excess of unlabelled site IV oligonucleotide probe (Fig. 4C, lane 5), but was not affected by a molar excess of unlabelled mutated site IV oligonucleotide (Fig. 4C, lane 6). A specific anti-CDX2 antibody (Biogenex) was able to supershift the CDX2 retarded band (Fig. 4C, lane 4), whereas the addition of the available anti-HOXA9, anti-HOXD10, or anti-HOXD11 antibodies did not alter CDX2 binding (Fig. 4C, lanes 7-9). Anti-HOXA9 addition led to the disruption of a faster migrating complex (Fig. 4C, lane 7), indicating a weak binding of HOXA9 and/or of a paralogy group 9 HOX protein to site IV.

To verify whether the sites identified via in silico analysis are indeed required for the observed binding by the HOXD10, HOXD11, HOXD13, and CDX2 proteins, we tested in EMSAs mutant derivatives (see Materials and Methods) of the oligonucleotide probes representing sites II to IV (Suppl. Figs. 4, 5, and 6). While the mutation of site IIa did not affect the binding by HOXD10, HOXD11, HOXD13, and CDX2 (Suppl. Fig. 4A, B, C, and D, lanes 3), the mutation of site IIb abolished the binding by all proteins tested (Suppl. Fig. 4A, B, C, and D, lanes 6). The mutation of site IIIa led to a substantial reduction of the binding by HOXD10, HOXD11, and HOXD13 (Suppl. Fig. 5A, B, C, and D, lanes 3). The mutation of site IIIb, conversely, did not

significantly affect the binding by HOXD10 and HOXD13, while the binding by HOXD11 was reduced (Suppl. Fig. 5A, B, C, and D, lanes 6). Finally, the mutation of site IV completely abolished the binding by HOXD10, HOXD11, HOXD13, and CDX2 (Suppl. Fig. 6A, B, C, and D, lanes 3).

These results show that the 330 bp region encompassing the *pri-miR-196b* TSS contains several potential 5'HOX and CDX binding sites of which only one, site IV, is strongly bound in vitro by both 5' HOX proteins and CDX2.

A sequence alignment between the mouse *pri-miR-196b* promoter region and the corresponding region of the paralogous *miR-196a-2* locus within the HOXC cluster on mouse chromosome 15 revealed a conservation of binding sites IIIa, IIIb, and IV and of the putative TATA box (Suppl. Fig. 3) together with their relative spacing, suggesting the evolutionary conservation of these relevant regulatory elements following the major genome duplication events that led to the formation of the HOXA and HOXC clusters (49). In addition, this finding suggests that the expression of *miR-196b* and its paralog *miR-196a-2* might be regulated in concert by the same upstream transcription factors. Interestingly, no obvious sequence conservation was detected analyzing the genomic region upstream to the paralogous *miR-196a-1* locus located within the HOXB cluster on chromosome 11, leaving its transcriptional layout and regulation still an open question.

3.4. Cdx2 and Hoxd13 bind the pri-miR-196b promoter region in vivo, during embryogenesis

We next tested, using in vivo chromatin crosslinking and immunoprecipitation (ChIP), whether we could identify, during embryogenesis when *miR-196b* is maximally expressed, the canonical epigenetic hallmarks of promoter activity and the binding of 5' HOX and/or CDX within the genomic region encompassing the TSS. To this end, we prepared crosslinked chromatin from E10 mouse embryos, a stage comparable to HH stage 21 in chick (50) when *miR-196b* is highly expressed (see also 51). E10 mouse embryos were dissected to obtain chromatin from anterior regions (including hind-, mid-, and fore-brain) and from posterior ones (including posterior trunk, hindlimbs and tail) (Fig. 5A). Immunoprecipitations were performed using specific antibodies against trimethylated lysine 4 of histone H3 (me3K4), acetylated lysine 27 of histone H3 (AcK27), and trimethylated lysine 27 of histone H3 (me3K27) (Abcam). The first two modifications represent hallmarks of transcriptionally active promoters, with me3K4 usually found at transcriptional start sites, whereas the latter modification is characteristic of transcriptionally silent regions (52-54). The immunoprecipitated chromatin associated DNA was assayed for enrichment of specific sequences by real-time PCR (qPCR) analysis. Three primer sets were used, one amplifying

a fragment located close to the TSS, another amplifying a control region located within the first intron of the Hoxa9 gene, and a third one amplifying a fragment located within a chromosome 6 gene desert ~75 Mb 3' from the TSS of *miR-196b* (Fig. 5A). Substantial enrichments of the *miR-196b* promoter region were observed with the anti-me3K4 and anti-AcK27 in chromatin extracted from the tail region (Fig. 5B, left panel), whereas considerably lower enrichments were observed in chromatin derived from the anterior portion of the embryo (Fig. 5B, right panel). Conversely, a significant enrichment was observed using anti-me3K27 only in chromatin from the anterior E10 embryo (Fig. 5B, right panel). No, or only comparably low enrichments of the gene desert fragment were observed with all three antibodies (Fig. 5B). These data confirm that the identified *pri-miR-196b* promoter region carries the hallmarks of a transcriptionally active promoter region in the posterior part of E10 mouse embryos, i. e. where and when *miR-196b* is maximally expressed.

We then analysed by ChIP the binding in vivo, in mouse embryos, of select 5' Hox proteins and of the Cdx2 transcription factor. Out of four different 5' Hox proteins (Hoxa9, Hoxd10, Hoxd11, Hoxd13), for which specific, ChIP-grade antibodies were available, only Hoxd13 showed an appreciable binding to the *pri-miR-196b* promoter region in E10 embryo trunk chromatin (Fig. 5C, left panel). Similarly, Cdx2 was found to bind the *pri-miR-196b* promoter region selectively in E10 embryonic posterior trunk chromatin (Fig. 5C, left panel). No significant binding by 5' Hox or Cdx2 was observed at the *pri-miR-196b* promoter region in E10 head chromatin (Fig. 5C, right panel), and both in head and trunk E10 chromatin at the control genomic region (Fig. 5C right and left panels, respectively). At a later stage, in E13.5 embryonic trunk chromatin, Cdx2 binding to the *pri-miR-196b* promoter region was considerably reduced, whereas Hoxd13 binding was unchanged (Fig. 5D).

Thus, among the homeodomain proteins that were found in EMSAs to be potentially capable of binding individual sites within the *pri-miR-196b* putative promoter, Cdx2 and Hoxd13 were found to actually bind it in vivo, during mouse development. These data hence establish Cdx2 and Hoxd13 as possible regulators of *miR-196b* expression during embryogenesis.

3.5. The miR-196b promoter region is capable of driving expression in the chick embryo neural tube

To verify whether the identified *pri-miR-196b* promoter regulatory elements were sufficient and necessary to recapitulate the embryonic expression pattern of *miR-196b*, we carried out in ovo electroporation of chick embryos (40) as this represents a profitable system to rapidly characterise conserved transcriptional regulatory regions. Four different reporter constructs were generated driving the lacZ reporter gene (Fig. 6). The p196b(-249;+624)WT reporter contains a ~870 bp

genomic region spanning the TSS of *pri-miR-196b* (see also Fig. 3B) and comprises all five analysed binding sites (I-V), the p196b(-171;+508) reporter contains a deletion of 77 bp 5' to the TSS and of 116 bp 3' to it, the p196b(-249;+624)MUT3/4 reporter represents the entire 870 bp promoter region carrying DNA-binding disrupting mutations within sites IIIab and IV, and finally, the p196b(-249;+624)INV reporter represents a control construct containing the whole promoter region in inverted orientation (Fig. 6). Stage HH 10 chick embryos were co-electroporated in ovo with the indicated reporter constructs and with a GFP control construct (pCMV::eGFP Fig. 6A). The expression of the reporter gene was monitored 24 hours after electroporation (at stage HH 17-19).

Expression of the p196b(-249;+624)WT lacZ reporter was consistently detected in GFP positive stage HH19 chick embryos in neuroepithelial cells of the neural tube, and showed reproducibly an anterior expression front reaching approximately somites 23-24 (Fig. 6, panels B and C). The p196b(-171;+508) reporter also displayed with high frequency an expression pattern within the neural tube that was superimposable to that of p196b(-249;+624)WT. Conversely, both the p196b(-249;+624)MUT3/4 and the p196b(-249;+624)INV reporters showed little or no expression in the neural tube, respectively (Fig. 6, panels E and F). The results are summarized as a graph in Fig. 6 (panel G).

Overall, these data indicate that a ~680 bp genomic region spanning the TSS of *pri-miR-196b* is sufficient to recapitulate the neural tube expression pattern of *miR-196* and that either 5'HOX/CDX binding site IIIab or site IV, or both, are necessary for driving *miR-196b* transcription in the neural tube.

3.6. CDX2 and 5' HOX proteins but not 3' HOX proteins activate transcription from the pri-miR-196b promoter region

To assess whether CDX2 and 5' HOX proteins could indeed regulate transcription from the *pri-miR-196b* promoter, we generated a luciferase reporter construct, p196b(-249;+624), containing the *pri-miR-196b* promoter region tested in chick embryos. P19 cells were transiently co-transfected with p196b(-249;+624) or with the pLuc control reporter construct, together with increasing amounts of HOXA9, HOXD10, HOXD11, HOXD13, or CDX2 expression constructs (Fig. 7A). All tested 5' HOX proteins could efficiently activate the basal activity of the p196b(-249;+624) reporter, whereas no or only a marginal activation was observed on the control reporter (Fig. 7A). Similarly, CDX2 co-expression led to a significant increase of the p196b(-249;+624) reporter activity (Fig. 7A). We next tested whether two 3' HOX proteins, HOXD8 and HOXB1, could activate transcription from the *pri-miR-196b* promoter. Neither HOXD8 nor HOXB1 if co-

expressed with the p196b(-249;+624) reporter could significantly activate its transcription (Fig. 7A). As 3' located HOX proteins usually require the presence of a Pbx-family co-factor for their optimal transcriptional activation (29,55,56), we also tested the effect of HOXB1 on the p196b(-249;+624) reporter in the presence of PBX1. As a control PBX1 was also co-expressed with HOXA9 (Fig. 7B). While the co-expression of PBX1 with HOXA9 led only to a modest increase in the activation of p196b(-249;+624) (Fig. 7B), but not of the mutated versions of p196b(-249;+624) (Suppl. Fig. 8A), the co-expression of HOXB1 and PBX1 did not significantly alter the reporter basal activity, indicating that the lack of transcriptional activity of HOXB1 on the *pri-miR-196b* promoter did not depend on the absence of co-expressed HOX co-factors. Taken together these data show that of the HOX proteins tested only AbdB-related 5'Hox proved to be able to transactivate the *pri-miR-196b* promoter, whereas HOX proteins from loci located 3' to that of *miR-196* did not regulate its transcription in transient transcription assays.

We also generated luciferase reporter constructs containing mutated derivatives of the *pri-miR-196b* promoter. The p196b(-249;+624)M3, p196b(-249;+624)M4, and p196b(-249;+624)M3/4 reporter constructs contain mutations of site IIIab, site IV, and of both sites IIIab and IV, respectively (Suppl. Fig. 7), while p196b(-171;+508) represents a deletion mutant of both a 77 bp region 5' to the TSS, containing sites I and II, and a region of 116 bp 3' to the TSS (Suppl. Fig. 7). P19 cells were transiently transfected with the mutant reporter constructs together with increasing amounts of the CDX2 or HOXD13 expression constructs. The p196b(-171;+508) reporter showed only a moderate reduction in transcriptional activation by CDX2 with respect to its wild type counterpart (Fig. 8). Similarly, only a slight reduction in luciferase activity was observed with p196b(-249;+624)M3 upon CDX2 expression. Instead, a substantial decrease of transactivation by CDX2 was observed with p196b(-249;+624)M4 or p196b(-249;+624)M3/4 (Fig. 8), indicating that site IV is the principal CDX2 responsive element within the *pri-miR-196b* promoter. Transactivation of the *pri-miR-196b* promoter by 5' HOX proteins, conversely, revealed to require generally more than a single binding site. For instance, HOXA9, HOXD10, HOXD11, or HOXD13 upregulation of both the p196b(-249;+624)M3 and the p196b(-249;+624)M4 mutated reporter constructs was considerably lower than that of their wild type counterpart (p196b(-249;+624)). And the p196b(-171;+508) deletion mutant could not be fully transactivated by HOXA9, HOXD11, and HOXD13 (Fig. 8 and Suppl. Fig. 8).

Taken together these results establish that the evolutionarily conserved sites within the *pri-miR-196b* promoter are relevant for mediating transcriptional control by the CDX2 and 5' HOX proteins. In particular, CDX2 as opposed to 5'HOX proteins revealed to require selectively site IV for regulating *pri-miR-196b* transcription.

3.7. Cdx2 expression is required for endogenous miR-196b transcription

As Cdx2 proved to efficiently transactivate the *pri-miR-196b* promoter in transient transfection assays, we next tested whether its loss-of-function would affect the expression of endogenous *miR-196b*. To this end we took advantage of NTERA-2 human pluripotent embryonal carcinoma (EC) cells (57), which we found to express both CDX2 and *miR-196b* after their induction with retinoic acid (RA) to differentiate into a neuronal cell fate (Fig. 9A and B). NTERA-2 cells, which did not express *CDX2* in their undifferentiated state, showed an increase in *CDX2* expression after 14 days of 10^{-5}M RA treatment, which reached its peak after 21 days of RA induction (Fig. 9B). Similarly, *miR-196b*, which was also not expressed in undifferentiated NTERA-2, was maximally expressed after 14 days of RA treatment (Fig. 9B).

We knocked-down the expression of CDX2 by RNA interference using three different siRNAs (see Methods). CDX2 protein levels in 14d RA-treated NTERA-2 cells were markedly reduced by all of the transfected siRNAs (-60%, -54% and -77%, respectively) (Fig. 9C, top left panel). CDX2 protein decrease correlated with a reduction of mature *miR-196b* transcript levels by ~40% (Fig. 9C right panel). We then tested whether the gain-of-function of *Cdx2* would cause an increase in endogenous *miR-196b* expression. *Cdx2* was overexpressed by transfecting increasing amounts of a *Cdx2* expression construct in 14d RA-treated NTERA-2 cells (Fig. 9C, left bottom panel). The enforced exogenous expression of *Cdx2* was associated with an increase of endogenous *miR-196b* transcript levels proportional to the amount of Cdx2 produced (Fig. 9C, right and left panels). Taken together these results support the conclusion that Cdx2 is a major regulator of *miR-196b* expression in this cell context.

4. DISCUSSION

miRNAs embedded within the four *HOX* gene clusters play relevant roles in embryonic development and cancer pathogenesis (6,7). The transcriptional mechanisms controlling their expression, however, are still largely uncharacterised. We focused on the *miR-196b* locus, as it is located in an evolutionarily highly conserved region between *Hoxa9* and *Hoxa10* on mouse chromosome 6, and, according to available mouse EST clones, it appeared to originate from a long non-coding transcript starting from the *Hoxa10-Hoxa9* intergenic region. In this work, we establish by 5' and 3' RACE mapping that the mature *miR-196b* originates from a large, 2671 bp long, primary, non-coding transcript (*pri-miR-196b*), which starts within a TATA box promoter located at 260 bp 3' from the *Hoxa10* polyadenylation signal. In accordance with the assumption that the identified transcription start site (TSS) represents the bona fide core promoter of *pri-miR-196b*, we found in its vicinity by ChIP a significant enrichment for H3K4me3, a canonical epigenetic hallmark of active promoters (52). Furthermore, a genomic region of ~680 bp, comprising the identified *pri-miR-196b* TSS and two evolutionarily conserved 5'Hox/Cdx binding sites, proved by in ovo electroporation to be sufficient to generate an expression pattern within the chick embryo neural tube that closely recapitulates the developmental expression pattern of *miR-196b*. Thus, although it cannot be excluded that a larger read-through transcript(s) may exist, comprising the *miR-196b* locus and the adjacent *Hoxa10* and/or *Hoxa9* loci (see e. g. 58,59), our results are consistent with the conclusion that *pri-miR-196b* is transcribed from an autonomous promoter, independent of the characterised *Hoxa10* or *Hoxa9* promoters.

While *Hox*-embedded miRNAs are generally held to be expressed co-ordinately both in space and time with their adjacent *Hox* genes, this notion is not always supported by experimental evidence. Indeed, *miR-196a-1* a paralog of *miR-196b* has been shown to have a more restricted spatial expression pattern with respect to its neighbouring gene, *Hoxb9*. And in general, *miR-196* genes appear to have a more posteriorly restricted boundary of expression that the adjacent *Hox9*-paralogy group genes (9,60). This suggests that *miR-196* family members may not be simply regulated by promoters and regulatory elements belonging to adjacent *Hox* genes.

Our findings show a dynamic expression of *miR-196* during chick embryogenesis, starting at early stages in the most posterior portion of the embryonic anterior-posterior axis, and reaching its maximal anterior extension at HH stage 19-20. The *miR-196* posteriorly-restricted expression pattern observed in chick development closely matches the patterns of *miR-196* expression reported during mouse, *X. laevis*, and zebrafish embryogenesis as determined by WISH and/or by using miRNA-responsive sensor transgenes (9,23,27,61). This tempted to speculate that the transcriptional regulatory mechanisms driving *miR-196* expression, and hence the physical

elements they depend on, are evolutionarily conserved. We could indeed identify a number of conserved putative 5'Hox or Cdx transcription factor binding sites within the region spanning the *Hoxa10-Hoxa9* intergenic region and encompassing the *pri-miR-196b* TSS, some of which we found to be able to bind both 5'Hox proteins and Cdx2 in vitro. Of these, three binding sites located in proximity to the *pri-miR-196b* TSS (site IIIab and IV), proved to be critical for *pri-miR-196b* expression in vivo. Their combined mutation caused the complete loss of a reporter construct expression within the neural tube in electroporated chick embryos. Similarly, in cultured cells, mutation of site IIIab and IV, led to a substantial reduction of the transcriptional activation of the *miR-196b* promoter by 5'Hox and Cdx2 gene products. Our results thus implicate both 5'Hox and Cdx2 in the control of *pri-miR-196b* expression. Cdx2 was indeed found to be necessary for the optimal expression of endogenous *miR-196b* in human NTERA-2 EC cells, which, upon retinoic acid induction, express *CDX2* but only some 5'*Hox* (62), thus establishing Cdx2 as a regulator of *miR-196b* expression in this cell context.

Single mutations of either site IIIab or site IV showed that transcriptional activation by 5'Hox proteins (HOXA9, HOXD10, HOXD11, and HOXD13) was affected equally well by site IIIab or site IV mutation, while Cdx2-mediated transactivation was significantly impaired only by site IV mutation. These results correlate with our EMSA experiments showing that 5'Hox bind both site IIIab and IV, whereas Cdx2 binds strongly to site IV and only barely to site IIIab. Thus, among the identified conserved binding sites, site IV can be bona fide considered as both a CDRE (Cdx responsive element) as well as a 5'Hox responsive element. The fact that the *miR-196b* promoter can be bound both by 5'Hox and Cdx proteins would suggest a possible competition or sequentiality in vivo between these transcription factors for its regulation. Indeed, in vivo, selectively in the caudal portion of mouse embryos, the promoter region including site IV was bound at stage E10, both by the Cdx2 protein and by the Hoxd13 5'Hox protein. At E13.5, binding of Cdx2 was essentially undetectable, whereas binding of Hoxd13 persisted, prompting to speculate that at this stage the control of *miR-196b* expression is taken over by paralogy group (PG) 13 Hox proteins.

Regulation of *miR-196b* expression potentially by both 5'*Hox* and by *Cdx* gene products could point to different roles played by this miRNA, pertaining to different phases of anterior-posterior (A-P) axis formation and patterning during embryogenesis. *Cdx* genes have been shown to be crucial for posterior elongation of the main body axis during mouse development (20,21,63-65). Human and mouse genomes contain three paralogous members of the *Cdx* gene family of *ParaHox* transcription factors, *Cdx1*, *Cdx2*, and *Cdx4* (17). While all *Cdx* genes participate in the posterior extension of the A-P axis, *Cdx2* plays a primary role in this process, as *Cdx2* conditional, epiblast restricted mutant mice are posteriorly truncated, lack axial tissues posterior to the forelimbs, and die

in utero at E10.5 (63). Similarly, $Cdx2^{+/-}$;$Cdx4^{-/-}$ mice fail to develop caudal body regions posterior the hindlimbs, and axial extension prematurely terminates at ~E10.5 (21). *Cdx* genes have been shown to act downstream of Fgf signalling to regulate 5' *Hox* genes in posterior embryo regions during the specification of regional identities along the A-P axis (18,19). The gain of function of "central" Hox genes, such as *Hoxb8* and *Hoxa5*, expressed in developing trunk regions of the mouse embryo, has been shown to rescue the phenotype produced by *Cdx* gene mutation (21). Conversely, "posterior", 5' Hox genes belonging to PG 13, such as *Hoxa13*, *Hoxb13*, and *Hoxc13*, have been shown to prematurely arrest axial elongation if expressed precociously, recapitulating the defects observed in *Cdx* mutant embryos (21). These observations led to the conclusion that *Cdx* and "trunk" *Hox* genes promote posterior extension of axial structures by sustaining the growth of progenitor cells within axial tissues. This process would last in time until the onset of "posterior", 5' *Hox* gene expression. The gene products of 5' Hox, in particular of PG 13 *Hox*, would then exert their functional dominance over more anteriorly-expressed, more 3' located, "trunk" Hox proteins to arrest axial extension (see 21).

In this scenario, the binding and activation of the *pri-miR-196b* promoter by *Cdx* and by 5' *Hox* genes could reflect different roles in time played by *miR-196b* in dampening "trunk" Hox gene product expression. During the Cdx-controlled phase of axis formation and elongation regulation of *miR-196b* expression by Cdx would play a role in the fine-tuning of trunk "central" Hox gene product levels. In the subsequent, 5'Hox-controlled phase of axis extension termination, the regulation of *miR-196b* by 5'Hox would play a role in reinforcing the dominant competition, or "posterior prevalence" (see below), of these on trunk, "central" Hox gene activity by contributing to the clearance of their gene products. Intriguingly, transcription of Cdx2 was not significantly reduced in mice precociously expressing PG13 *Hox* genes, pointing to possible post-transcriptional mechanisms in the interference by PG13 5' Hox on Cdx and/or "central" Hox gene function (21).

5'Hox proteins are known to exert a functional dominant role over more "anterior", 3'Hox proteins in body regions in which these are co-expressed, an activity that is known under the name of "posterior prevalence" (66-68). Much has been speculated on the mechanism(s) underlying this feature of the Hox gene products, including the possible intervention of Hox cluster-embedded miRNAs. This latter assumption stems from the observation that Hox-embedded miRNAs preferentially target transcripts originating from Hox genes located 3' to the miRNA loci, while leaving 5' Hox unaffected (see 10). Thus, Hox-embedded miRNAs would act as mediators of posterior prevalence by selectively quenching the amounts of "anterior" Hox gene products, hence their function, in posterior body regions. In this context, our findings, showing binding in vivo during embryogenesis and transcriptional regulation of the *miR-196b* promoter by 5'Hox proteins,

prompt to speculate that these could employ *miR-196* family members, by activating their transcription, to control post-transcriptionally the protein levels of "anterior" 3'Hox, within posterior body regions in which they are co-expressed. Further work will be required to validate this hypothesis.

In conclusion, the results described in this paper establish an essential role for Cdx2 and 5'Hox responsive elements within the *pri-miR-196b* promoter region for the developmental expression of *miR-196*, with possible implications in understanding the misregulation of *miR-196* expression in tumorigenesis as Cdx and Hox genes are also involved in cancer pathogenesis.

5. ACKNOWLEDGEMENTS

Thanks are due to J.Deschamps for the kind gift of the pSG5-CDX2 expression construct. S.F. was the recipient of an EMBO short-term fellowship. This work was supported by grants from the Italian Association for Cancer Research (AIRC) (to V. Z.), and from the Fondazione Cassa di Risparmio di Modena (CRMO) (to V. Z.), as well as from the Swiss National Science Foundation and the Novartis Research Foundation (to F.M.R.).

6. REFERENCES

1. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116, 281-297.
2. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215-233.
3. Calin, G.A. and Croce, C.M. (2006) MicroRNA signatures in human cancers. *Nat Rev Cancer*, 6, 857-866.
4. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H. and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *Embo J*, 23, 4051-4060.
5. Schanen, B.C. and Li, X. (2011) Transcriptional regulation of mammalian miRNA genes. *Genomics*, 97, 1-6.
6. Mansfield, J.H. and McGlinn, E. (2012) Evolution, expression, and developmental function of Hox-embedded miRNAs. *Curr Top Dev Biol*, 99, 31-57.
7. Chen, C., Zhang, Y., Zhang, L., Weakley, S.M. and Yao, Q. (2011) MicroRNA-196: critical roles and clinical applications in development and cancer. *J Cell Mol Med*, 15, 14-23.
8. Tanzer, A., Amemiya, C.T., Kim, C.B. and Stadler, P.F. (2005) Evolution of microRNAs located within Hox gene clusters. *J Exp Zool B Mol Dev Evol*, 304, 75-85.
9. Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G. *et al.* (2004) MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet*, 36, 1079-1083.
10. Yekta, S., Tabin, C.J. and Bartel, D.P. (2008) MicroRNAs in the Hox network: an apparent link to posterior prevalence. *Nat Rev Genet*, 9, 789-796.
11. Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, 304, 594-596.
12. Kawasaki, H. and Taira, K. (2004) MicroRNA-196 inhibits HOXB8 expression in myeloid differentiation of HL60 cells. *Nucleic Acids Symp Ser (Oxf)*, 211-212.
13. Mallo, M., Wellik, D.M. and Deschamps, J. (2010) Hox genes and regional patterning of the vertebrate body plan. *Dev Biol*, 344, 7-15.
14. Alexander, T., Nolte, C. and Krumlauf, R. (2009) Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol*, 25, 431-456.
15. Shah, N. and Sukumar, S. (2010) The Hox genes and their roles in oncogenesis. *Nat Rev Cancer*, 10, 361-371.

16. Abramovich, C., Pineault, N., Ohta, H. and Humphries, R.K. (2005) Hox genes: from leukemia to hematopoietic stem cell expansion. *Ann N Y Acad Sci*, 1044, 109-116.
17. Ferrier, D.E., Dewar, K., Cook, A., Chang, J.L., Hill-Force, A. and Amemiya, C. (2005) The chordate ParaHox cluster. *Curr Biol*, 15, R820-822.
18. Isaacs, H.V., Pownall, M.E. and Slack, J.M. (1998) Regulation of Hox gene expression and posterior development by the Xenopus caudal homologue Xcad3. *Embo J*, 17, 3413-3427.
19. Pownall, M.E., Isaacs, H.V. and Slack, J.M. (1998) Two phases of Hox gene regulation during early Xenopus development. *Curr Biol*, 8, 673-676.
20. van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B.I. and Deschamps, J. (2002) Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development*, 129, 2181-2193.
21. Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.N. *et al.* (2009) Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell*, 17, 516-526.
22. Duboule, D. and Morata, G. (1994) Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet*, 10, 358-364.
23. Asli, N.S. and Kessel, M. (2010) Spatiotemporally restricted regulation of generic motor neuron programs by miR-196-mediated repression of Hoxb8. *Dev Biol*, 344, 857-868.
24. Hornstein, E., Mansfield, J.H., Yekta, S., Hu, J.K., Harfe, B.D., McManus, M.T., Baskerville, S., Bartel, D.P. and Tabin, C.J. (2005) The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature*, 438, 671-674.
25. McGlinn, E., Yekta, S., Mansfield, J.H., Soutschek, J., Bartel, D.P. and Tabin, C.J. (2009) In ovo application of antagomiRs indicates a role for miR-196 in patterning the chick axial skeleton through Hox gene regulation. *Proc Natl Acad Sci U S A*, 106, 18610-18615.
26. He, X., Yan, Y.L., DeLaurier, A. and Postlethwait, J.H. (2011) Observation of miRNA gene expression in zebrafish embryos by in situ hybridization to microRNA primary transcripts. *Zebrafish*, 8, 1-8.
27. He, X., Yan, Y.L., Eberhart, J.K., Herpin, A., Wagner, T.U., Schartl, M. and Postlethwait, J.H. (2011) miR-196 regulates axial patterning and pectoral appendage initiation. *Dev Biol*, 357, 463-477.
28. Nordeen, S.K. (1988) Luciferase reporter gene vectors for analysis of promoters and enhancers. *Biotechniques*, 6, 454-457.
29. Di Rocco, G., Mavilio, F. and Zappavigna, V. (1997) Functional dissection of a transcriptionally active, target-specific Hox-Pbx complex. *EMBO J.*, 16, 3644-3654.

30. Fantini, S., Vaccari, G., Brison, N., Debeer, P., Tylzanowski, P. and Zappavigna, V. (2008) A G220V substitution within the N-terminal transcription regulating domain of HOXD13 causes a variant synpolydactyly phenotype. *Hum Mol Genet*.

31. Zappavigna, V., Renucci, A., Izpisua-Belmonte, J.C., Urier, G., Peschle, C. and Duboule, D. (1991) HOX4 genes encode transcription factors with potential auto- and cross-regulatory capacities. *EMBO J.*, 10, 4177-4187.

32. Darnell, D.K., Kaur, S., Stanislaw, S., Konieczka, J.H., Yatskievych, T.A. and Antin, P.B. (2006) MicroRNA expression during chick embryo development. *Dev Dyn*, 235, 3156-3165.

33. Hamburger, V. and Hamilton, H.L. (1992) A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn*, 195, 231-272.

34. Salsi, V. and Zappavigna, V. (2006) Hoxd13 and Hoxa13 Directly Control the Expression of the EphA7 Ephrin Tyrosine Kinase Receptor in Developing Limbs. *J Biol Chem*, 281, 1992-1999.

35. Salsi, V., Ferrari, S., Ferraresi, R., Cossarizza, A., Grande, A. and Zappavigna, V. (2009) HOXD13 binds DNA replication origins to promote origin licensing and is inhibited by geminin. *Mol Cell Biol*, 29, 5775-5788.

36. Hu, Y.L., Passegue, E., Fong, S., Largman, C. and Lawrence, H.J. (2007) Evidence that the Pim1 kinase gene is a direct target of HOXA9. *Blood*, 109, 4732-4738.

37. Capellini, T.D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D., Zappavigna, V. and Selleri, L. (2006) Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development*, 133, 2263-2273.

38. Caronia, G., Goodman, F.R., McKeown, C.M., Scambler, P.J. and Zappavigna, V. (2003) An I47L substitution in the HOXD13 homeodomain causes a novel human limb malformation by producing a selective loss of function. *Development (Cambridge, England)*, 130, 1701-1712.

39. Dignam, J.R., Lebovitz, R.M. and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, 11, 1475-1489.

40. Itasaki, N., Bel-Vialar, S. and Krumlauf, R. (1999) 'Shocking' developments in chick embryology: electroporation and in ovo gene expression. *Nat Cell Biol*, 1, E203-207.

41. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25, 402-408.

42. Pearson, J.C., Lemons, D. and McGinnis, W. (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet*, 6, 893-904.
43. Beck, F., Erler, T., Russell, A. and James, R. (1995) Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn*, 204, 219-227.
44. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol*, 14, 116-127.
45. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420, 563-573.
46. Margalit, Y., Yarus, S., Shapira, E., Gruenbaum, Y. and Fainsod, A. (1993) Isolation and characterization of target sequences of the chicken CdxA homeobox gene. *Nucleic Acids Res*, 21, 4915-4922.
47. Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. and Werner, T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21, 2933-2942.
48. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, 12(6), 996-1006.
49. Garcia-Fernandez, J. (2005) The genesis and evolution of homeobox gene clusters. *Nat Rev Genet*, 6, 881-892.
50. Schneider, B.F. and Norton, S. (1979) Equivalent ages in rat, mouse and chick embryos. *Teratology*, 19, 273-278.
51. Kloosterman, W.P., Wienholds, E., de Bruijn, E., Kauppinen, S. and Plasterk, R.H. (2006) In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat Methods*, 3, 27-29.
52. Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J. and Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature*, 419, 407-411.
53. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298, 1039-1043.
54. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac

separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107, 21931-21936.

55. Di Rocco, G., Gavalas, A., Popperl, H., Krumlauf, R., Mavilio, F. and Zappavigna, V. (2001) The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. *J Biol Chem*, 276, 20506-20515.
56. Moens, C.B. and Selleri, L. (2006) Hox cofactors in vertebrate development. *Dev Biol*, 291, 193-206.
57. Andrews, P.W. (1988) Human teratocarcinomas. *Biochim. Biophys. Acta*, 948, 17-36.
58. Mainguy, G., Koster, J., Woltering, J., Jansen, H. and Durston, A. (2007) Extensive polycistronism and antisense transcription in the mammalian Hox clusters. *PLoS One*, 2, e356.
59. Tang, G.Q. and Maxwell, E.S. (2008) Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome Res*, 18, 104-112.
60. Burke, A.C., Nelson, C.E., Morgan, B.A. and Tabin, C. (1995) Hox genes and the evolution of vertebrate axial morphology. *Development*, 121, 333-346.
61. Qiu, R., Liu, Y., Wu, J.Y., Liu, K., Mo, W. and He, R. (2009) Misexpression of miR-196a induces eye anomaly in Xenopus laevis. *Brain Res Bull*, 79, 26-31.
62. Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F. and Boncinelli, E. (1991) Differential regulation by retinoic acid of the homeobox genes of the four Hox loci in human embryonal carcinoma cells. *Mech. Dev.*, 33, 215-228.
63. Savory, J.G., Bouchard, N., Pierre, V., Rijli, F.M., De Repentigny, Y., Kothary, R. and Lohnes, D. (2009) Cdx2 regulation of posterior development through non-Hox targets. *Development*, 136, 4099-4110.
64. van de Ven, C., Bialecka, M., Neijts, R., Young, T., Rowland, J.E., Stringer, E.J., Van Rooijen, C., Meijlink, F., Novoa, A., Freund, J.N. *et al.* (2011) Concerted involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone. *Development*, 138, 3451-3462.
65. Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. and Beck, F. (2004) Cdx2 is essential for axial elongation in mouse development. *Proc Natl Acad Sci U S A*, 101, 7641-7645.

66. Duboule, D. (1991) Patterning in the vertebrate limb. *Current Opinion in Genetics and Development*, 1, 211-216.
67. Duboule, D. and Morata, G. (1994) Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.*, 10, 358-364.
68. Tschopp, P. and Duboule, D. (2011) A regulatory 'landscape effect' over the HoxD cluster. *Dev Biol*, 351, 288-296.

7. LEGENDS TO FIGURES

Fig. 1. Temporal and spatial expression pattern of miR-196 in chick embryos. Expression of *miR-196* was assessed by whole-mount *in situ* hybridization (WISH). Mature microRNA expression was detected with a digoxigenin-labelled, LNA-modified probe (Exiqon) revealing all three miR-196 family members in chicken (*miR-196-1*, *miR-196-2*, *miR-196-3*, corresponding to *miR-196a-1*, *miR-196b*, and *miR-196a-2*, respectively, in human and mouse). Their expression was analysed during *in ovo* development at the Hamburger-Hamilton stages (HH) indicated in the panels (**A-L**). For each stage the left panel shows the dorsal view, the right panel the lateral one. For each stage 10-12 embryos were analysed, the presented embryos exemplify the obtained expression patterns. Black arrowheads represent the anterior boundary of *miR-196* expression.

Fig. 2. Identification of primary *miR-196b* 5' and 3' ends. Schematic representation of the mouse *Hoxa9-Hoxa10* genomic region. The *Hoxa10* and *Hoxa9* genes from their TSSs (+1) to their poly-A sequences are represented as grey block arrows. A black block arrowhead indicates the miR-196b pre-miR sequence, located between Hoxa9 and Hoxa10. The indicated *miR-196b* TSS and poly-A sequence were identified using 5'- and 3'-RACE, respectively. The resulting primary transcript (*pri-miR-196b)* and the mRNAs of neighbouring *Hox* genes are indicated on top. Primers used in the PCRs are indicated at the bottom. GSP, Gene Specific Primer.

Fig. 3. The region upstream to the *miR-196b* gene contains putative 5'HOX/CDX2 binding sites. **(A)** Alignment of *pri-miR-196b* and of its 5' promoter region to Chr.6 52231000-52227500 mouse genomic sequence was performed using the UCSC Genome Browser. UCSC genes, mouse mRNA, CpG islands and conservation tracks are shown in dense mode, with each item displayed separately, as indicated on the right. mRNAs are transcribed from left to right. The highest degree of conservation is found within a region of ~300bp located upstream from the transcription start site and in close proximity to the *pre-miR-196b* sequence. **(B)** Sequence comparison of the *miR-196b* promoter region in reference species was performed using the Clustal-Omega software. Seven different putative 5'HOX/CDX2 binding sites (boxed) were identified using the MatInspector software. Grey boxes highlight a canonical TATA box and an Initiator element. A black arrow indicates the mouse TSS, small boxes surmounted by an asterisk indicate the human TSSs. A previously mapped mouse CpG island is indicated by a grey rectangle.

Fig. 4. 5' HOX and CDX2 bind *in vitro* to the *miR-196b* promoter region. **(A, B)** EMSAs using oligonucleotides corresponding to sites I-V of the *miR-196b* promoter region and the

indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). **(C)** EMSA using a site IV oligonucleotide. Nuclear extracts (10 μg) prepared from HEK293, cells expressing CDX2 and all the analyzed 5'HOX proteins, led to the formation of four retarded bands. Black arrowheads indicate specific bands; asterisks point out non-specific bands. A specific band, comigrating with in vitro synthesized CDX2 (lane 1) was supershifted by an anti-Cdx2 antibody (white arrowhead). C2= Cdx2 antibody; A9, D10, D11 indicate antibodies against HOXA9, HOXD10 and HOXD11, respectively. WT= 500-fold excess of wild-type unlabelled competitor; MUT= 500-fold excess of unlabelled competitor carrying a mutation of the TTTATG sequence (site IVMUT see Materials and Methods). **(D)** miR-196b levels detected by qRT-PCR in total RNA from NTERA2, P19, and HEK293 cells. Expression levels were calculated according to the ΔΔCt method using RNU6B as a reference gene.

Fig. 5. 5'HOX and CDX2 bind *in vivo* to the miR-196b promoter. **(A)** Representative mouse embryo with boxes indicating the dissected portions used to prepare chromatin for ChIP assays. Schematic representation of the murine chromosome 6 region encompassing the *miR-196b* locus, showing the primer set amplifying the *miR-196b* promoter region and two alternative negative control regions, one located within the intron of *Hoxa9* (CTRL 1), and the other within a gene desert located ~70Mb 3' to the *miR-196b* locus (CTRL 2). **(B, C, D)** ChIPs on mouse embryo chromatin were performed with anti-me3K4, anti-me3K27, anti-AcK27 antibodies (B), or with anti-HOXA9, anti-HOXD10, anti-HOXD11, anti-HOXD13, and anti-CDX2 antibodies (C). (CA) Control antibody. IP enrichments were analysed by qPCR using oligonucleotides amplifying the three regions indicated. Dark grey bars represent the binding on the miR-196b promoter region, whereas light grey bars represent the enrichment on the two respective control regions. Bars represent the IP/INPUT ratio ±SEM of at least three independent experiments. An asterisk represents $p < 0.05$.

Fig. 6. The miR-196b promoter region recapitulates *in vivo* the endogenous miR-196b expression pattern. Chicken embryos were electroporated at HH10 stage and collected 24h later. A CMV-driven GFP construct was always co-electroporated as a positive control and only embryos showing labelling in the whole neural tube (as shown in **A**) were considered in the analysis and subjected to β-galactosidase staining. Lateral (**B**), and dorsal (**C**) views of embryos electroporated with the *miR-196b* promoter region (p196b(-249;+624)WT). (**D**), Embryos electroporated with a construct, p196b(-171;+508), carrying only the central portion of the promoter. Embryos injected

with mutant constructs, p196b(-249;+624)M3/4 (**E**), and p196b(-249;+624)INV (**F**), displayed no staining. The electroporated constructs are indicated under the respective panels. (**G**), Electroporation results are summarized in a stacked column chart. Expressing embryos are represented as percentage of the total number of analysed embryos (n= 30 for the wild-type construct, n=26 for p196b(-171;+508), and n=20 for the each of the promoter mutants). Expressing embryos were classified in three categories according to staining levels: strong, medium and weak/no expression.

Fig. 7. CDX2 and 5' HOX proteins activate transcription from the pri-miR-196b promoter region. Luciferase activities assayed in cell extracts of P19 murine cells, transiently transfected with the empty, pLuc (light grey) or the p196b(-249;+624) reporter constructs (dark grey), together with increasing amounts (0,12-0,25 μg) of expression constructs for the indicated HOX and CDX2 proteins. **(A)** HOXA9, HOXD10, HOXD11 and HOXD13 were tested as representative members of the 5' or "posterior" subclass of HOX proteins; HOXB1 and HOXD8 as representative members of "anterior" and "central" HOX proteins, respectively. **(B)** Fixed amounts of HOXA9 and HOXB1 (0.12 μg) were co-transfected together with a increasing amounts (0,25-0.5 μg) of a construct expressing their cofactor PBX1a. Bars represent the mean luciferase activity±SEM of at least three independent experiments. * = $p < 0,05$, ** = $p < 0,005$, *** = $p < 0,001$, NS = not statistically significant.

Fig. 8. Central portion of *miR-196b* promoter is crucial for HOX/CDX transactivation. Luciferase activities assayed in cell extracts of P19 murine cells, transiently transfected with increasing amounts (0,12-0,25 μg) of the expression constructs for CDX2 **(A)**, HOXA9 **(B)** or HOXD10 **(C),** together with a fixed amount of empty (light grey) and wild-type or mutant reporter constructs (dark grey) carrying the *pri-miR-196b* promoter sequence. A schematic representation of the transfected constructs is shown in Suppl. Fig. 7. Bars represent the mean luciferase activity±SEM of at least three independent experiments. * = $p < 0,05$, ** = $p < 0,005$.

Fig. 9 Cdx2 expression affects *miR-196b* transcription. NTERA-2 human EC cells were induced to differentiate towards a neural fate by 10^{-5}M all-trans retinoic acid (RA) treatment. (**A**) Immunoblotting of Cdx2 expression using anti-Cdx2 antibodies (Biogenex) of whole-cell extracts from untreated (-) or RA-treated (7, 14, 21 days) NTERA-2 cells. (**B**) *miR-196b* and *CDX2* mRNA levels detected by qRT-PCR in total RNA from NTERA-2 cells. Expression levels were calculated according to the ΔΔCt method using RNU6B and actin as reference genes, respectively, and using untreated NTERA-2 cells as calibrator. (**C**) Top left, immunoblotting of Cdx2 expression after

transfection of NTERA-2 cells treated with RA for 14 days with three different siRNA, or with the control scrambled siRNA (SCR). (NT) Non-transfected cells. Bottom left, immunoblotting to detect Cdx2 protein levels on whole-cell extracts from 14d RA-treated NTERA-2 cells transfected with increasing amounts of a Cdx2 expression construct. Right panel, *miR-196b* expression assayed by qRT-PCR in total RNA from siRNA-transfected or Cdx2-overexpressing NTERA-2 cells. Expression levels were calculated according to the ΔΔCt method using RNU6B as a reference gene and using nontransfected induced NTERA-2(RA-14d) cells as calibrator. Anti-Tubulin antibodies were used in all immunoblottings as a loading control.

LEGENDS TO SUPPLEMENTAL FIGURES

Suppl. Fig. 1. (**A**) Mature miR-196b expression detected by qRT-PCR in total RNA extracted from whole embryo (W) or from tissues dissected from head (H), anterior (AT) or posterior trunk (PT) of mouse embryos at the indicated stages of development. miR-196b expression was normalized using the small RNA RNU6B as a reference endogenous control. (**B**) Expression profile of the internal reference control RNU6B in the tissues and stages analysed showing its constancy.

Suppl. Fig. 2. Sequencing products of 5' and 3' RACE are shown on left and right, respectively. Black arrows indicate TSS and 3' end of transcripts. Two alternative TSSs for *pre-miR-196b* were discovered in humans (termed I and II).

Suppl. Fig. 3. Nucleotide sequence comparison of the *pri-miR-196b* and *pri-miR-196a2* promoter regions in the mouse genome was performed using the Clustal-Omega software. The putative 5'HOX/CDX2 binding sites termed IIIa and IV display a high degree of conservation in mouse. Concomitantly, a canonical TATA box is located at the same relative position in the genome. No initiator element was detected in *pri-miR-196a2* promoter region. The conservation of these elements is highlighted by black rectangles.

Suppl. Fig. 4. EMSAs using oligonucleotides corresponding to site IIab (SITE II WT) and to its derivatives carrying single or compound mutations of sites IIa and IIb (MUT IIa, MUT IIb, and MUT IIab), together with the indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). An arrowhead indicates the specific retarded complex, asterisks indicate non-specific bands.

Suppl. Fig. 5. EMSAs using oligonucleotides corresponding to site IIIab (SITE III WT) and to its derivatives, carrying single or compound mutations of sites IIIa and IIIb (MUT IIIa, MUT IIIb, and MUT IIIab), together with the indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). An arrowhead indicates the specific retarded complex, asterisks indicate non-specific bands.

Suppl. Fig. 6. EMSAs using oligonucleotides corresponding to site IV (SITE IV WT) and to its derivative, carrying a mutation within sites IV (MUT IV), together with the indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). An arrowhead indicates the specific retarded complex.

Suppl. Fig. 7. Schematic representation of the genomic fragments containing the *pri-miR-196b* promoter region used to generate luciferase reporter constructs. The annotated 5'HOX/CDX2 binding sites are represented with green vertical rectangles. Red boxes indicate mutations of the corresponding binding sites. A black arrow indicates the TSS.

Suppl. Fig. 8. Luciferase activities assayed in cell extracts of P19 murine cells. **(A)** Cells were transiently transfected with fixed amounts of the HOXA9 or PBX1a expression constructs, or with both HOXA9 and PBX1a expressors, together with a fixed amount of empty (light grey), wild-type, or mutant indicated reporter constructs (dark grey). **(B and C)** P19 cells were transfected with increasing amount of HOXD11 and HOXD13 expression constructs respectively, together with a fixed amount of the indicated reporter constructs. Bars represent the mean luciferase activity±SEM of at least three independent experiments. * = $p < 0,05$, ** = $p < 0,005$.

MicroRNA-196b is transcribed from an autonomous promoter and is directly regulated by Cdx2 and by posterior Hox proteins during embryogenesis

Sebastian Fantini°, Valentina Salsi°, Antonio Vitobello[#,*], Filippo M. Rijli[#] and Vincenzo Zappavigna°

°Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 213/d, Modena 41125, Italy; [#] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland; [*] Present address: Preclinical Safety, Novartis Institutes for Biomedical Research, CH-4057 Basel, Switzerland

Address correspondence to: Vincenzo Zappavigna, Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 213/d, Modena 41125, Italy, Tel. 0039 059 2055537, Fax 0039 059 2055548, E-Mail: zappavigna.vincenzo@unimore.it

ABSTRACT

The *miR-196* miRNA gene family located within the Hox gene clusters has been shown to function during embryogenesis and to be aberrantly expressed in various malignancies, including leukemia, melanoma, and colorectal cancer. Despite its involvement in numerous biological processes, the control of *miR-196* expression is still poorly defined. We identified the *miR-196b* promoter and found that the mature *miR-196b* originates from a large, non-coding primary transcript, which starts within an autonomous TATA box promoter and is not in physical continuity with either the *Hoxa10* or *Hoxa9* main primary transcripts. A ~680 bp genomic fragment, spanning the *pri-miR-196b* transcription start site, is sufficient to recapitulate the neural tube expression pattern of *miR-196* during embryogenesis. This region contains potential binding sites for Cdx and 5'Hox transcription factors. Two of these sites revealed to be necessary for neural tube expression and were bound in vivo by Cdx2 and Hoxd13. We show that Cdx2 is required for *miR-196* expression and that both Cdx2 and 5'Hox, but not 3'Hox, are able to activate the *miR-196b* promoter. The possible role of Cdx2- and 5'Hox-mediated regulation of *miR-196* expression in vertebrate anterior-posterior (AP) axis formation during embryogenesis is discussed.

1. INTRODUCTION

MicroRNAs (miRNAs) are evolutionarily conserved, small non-coding RNAs of ~20-24 nucleotides (nt) in length, which negatively regulate gene expression at the post-transcriptional level. They typically bind to the 3' UTR of their target mRNAs, and either inhibit translation or induce mRNA degradation (1, 2). miRNAs derive from larger primary miRNAs (pri-miRNAs), in many cases non-coding, which are processed by endonucleases, to precursors (pre-miRNAs) in the nucleus and eventually to their mature form in the cytoplasm. miRNAs participate in the regulation of a vast array of biological processes, ranging from cell cycle regulation to differentiation and development. Consequently, the deregulation of miRNA expression is implicated in several human diseases including cancer (3). The mechanisms controlling miRNA gene expression, especially those regarding the transcription of their pri-miRNA precursors, have been only partially elucidated. These studies reveal a primary role for RNA polymerase II (RNA polII) in miRNA transcription (4), and that many miRNA genes use their own promoter and transcriptional regulatory elements, even if located within gene clusters (intergenic) or within introns of coding genes (intragenic) (5). Still less known are the transcription factors directing miRNA gene transcription. Their knowledge, however, would greatly increase our understanding of miRNA gene deregulation in disease, and would allow to define novel gene regulatory networks, comprising miRNAs, their transcriptional regulators, and their target mRNAs, thus connecting transcriptional and post-transcriptional controls in single pathways.

The *miR-196* gene family is evolutionarily conserved, having arisen in a common ancestor of chordates and urochordates, and includes three genes, all located within the *HOX* gene clusters (6). In the human genome, the *miR-196a-1* gene is located in the intergenic region between *HOXB9* and *HOXB10* on chromosome 17, *miR-196a-2* between *HOXC9* and *HOXC10* on chromosome 12 (7). *miR-196b* is located between *HOXA9* and *HOXA10* on chromosome 7 in human, and on chromosome 6 in mice, in a region evolutionarily conserved among vertebrates (7). *miR-196a-1* and *miR-196a-2* share the same mature, functional miRNA sequence, while *miR-196b* differs from the *miR-196a* sequence by one nucleotide (8). *miR-196* family members have been reported to participate in relevant biological processes such as embryonic development (reviewed in 9) and tumorigenesis where they have been found to be aberrantly expressed in various malignancies, including glioblastoma, melanoma, leukemia, and colorectal cancer (reviewed in 7).

Among all putative miR-196 targets, found by in silico prediction, *Hox* mRNAs appear to be the class of transcripts that are preferentially targeted by the miR-196 gene family (10). Indeed, if the percentage of conserved miRNA target sites within the 3' UTRs of *Hox* transcripts is taken into account, the *miR-196* gene family ranks first among all conserved miRNA families (10). The

"central" *Hoxa7*, *Hoxb8*, *Hoxc8*, and *Hoxd8* genes, expressed in the developing trunk during embryogenesis, have been demonstrated experimentally to be conserved targets of miR-196 (9,11,12).

Hox homeodomain-containing transcription factors have been shown to play crucial roles in embryonic development by controlling determination of cell fate, motility, adhesion, proliferation, and apoptosis, eventually orchestrating the generation of different morphological identities along the primary body and limb axes (reviewed in 13,14). Alterations of HOX gene expression and/or structure have been moreover found to be associated with tumorigenesis (reviewed in 15,16). In addition to *Hox* genes, the *ParaHox* family *Cdx* genes (17) are relevant to anterior-posterior axis patterning and elongation during embryogenesis, in part also as upstream regulators of *Hox* genes (18-21).

The remarkably close proximity of miR-196 and its most relevant target genes has prompted speculations about its biological significance. It has been proposed that HOX-embedded miRNAs (*miR-10* and *miR-196*) may play a role in the phenomenon termed "posterior prevalence". Posterior prevalence, represents one of the characteristic features of the *Hox* gene system (reviewed in 13). It accounts for the fact that more posteriorly-expressed *Hox* genes, such as the *AbdB*-related, 5' located Hox (5'Hox), are functionally dominant over anteriorly-expressed genes, as single loss-of-function *Hox* mutants display phenotypes only in the most anterior regions of their expression domains (22). The putative role of Hox cluster-linked miRNAs in posterior prevalence rests on the observation that their *Hox* target mRNAs are asymmetrically distributed with respect to their location within the HOX clusters. For instance, the majority of *Hox* target mRNAs of *miR-196* are located 3' to the *miR-196* loci, thus the activity of *miR-196* would repress expression of *Hox* genes located 3' to its locus, while sparing 5' located *Hox* (reviewed in 10).

Both loss- and gain-of-function experiments have shown that the miR-196 gene family participate in developmental processes such as limb A-P, nervous system, and axial skeletal patterning (reviewed in 6). The majority of these studies have focused on the capability of *miR-196* to negatively regulate the expression of *Hoxb8* (23-26), a key representative of "central" Hox genes, expressed in developing trunk regions of the embryo. *miR-196* has been shown to reinforce the repression and to contribute to the clearance of the *Hoxb8* gene product in caudal regions of the embryo, to define the posterior boundary of *Hoxb8* function (23,24). In more anterior regions where it is co-expressed with *Hoxb8*, *miR-196* has been shown to increase the robustness of the *Hoxb8*-controlled genetic program by eliminating fluctuations in its expression (25). Finally, overexpression of *miR-196* in zebrafish causes transformations of anterior vertebrae identities and the failure of pectoral fin bud initiation (27).

Despite its roles in development and cancer (reviewed in 7) the regulation of *miR-196* gene expression has never been thoroughly characterised (see 6). Still unknown are its promoters, their upstream regulators as well as the regulatory elements to which these bind. We thus set out to characterize the miR-196b promoter, and to explore the possibility that it might be controlled by upstream regulators of the Hox gene clusters and/or by Hox gene products themselves possibly as part of the mechanism(s) underlying posterior prevalence. We found that the mature *miR-196b* originates from a large, 2671 bp, primary, non-coding transcript (*pri-miR-196b*), which is, unlike previously assumed, not in physical continuity with either the *Hoxa10* or *Hoxa9* main primary transcripts, and starts within an autonomous TATA box promoter. The evolutionarily conserved genomic region upstream to the *pri-miR-196b* transcriptional start site (TSS) was found to contain several potential binding sites for Cdx and 5'Hox transcription factors. We establish that a ~680 bp genomic fragment spanning the *pri-miR-196b* TSS is sufficient to recapitulate the neural tube expression pattern of *miR-196* and that two of the analysed sites are necessary in vivo for *pri-miR-196b* promoter function. We show that these sites are bound in vivo by the Cdx2 and Hoxd13 proteins in posterior regions of E10 mouse embryos. Cdx2 and 5'Hox, but not 3'Hox, were found to transactivate the *pri-miR-196b* promoter and Cdx2 to be necessary for the expression of *miR-196* in human embryonal carcinoma (EC) cells. We discuss the possible significance of the regulation of *miR-196* expression by Cdx2 and/or 5'Hox in the context of vertebrate anterior-posterior (AP) axis formation.

2. MATERIALS AND METHODS

2.1. Plasmid constructs

The *pri-miR-196b* murine promoter sequence was obtained by PCR amplification from genomic DNA of P19 cells using the following primers: 196b(-249)for: 5'-AAAAGCTGTGGAATGAAGC-3'; 196b(+624)rev: 5'-CTCCGGAAAGGTACGCCTAG-3'. The amplified fragment was sequence-verified and cloned into the HindIII site of the pXP2-luciferase vector (28) to obtain the p196b(-249;+624) reporter construct. A 680-bp fragment, encompassing sites III and IV, and V, was PCR-amplified using the following primers: 196b(-171)for: 5'-GAGGGCTAGCCCGAGCCC-3'; 196b(+508)rev: 5'-CGGTGTGAGGCGGGTTCC-3'. The fragment was verified by sequencing and cloned into the HindIII site of pXP2 to generate p196b(-171;+508). The p196b(-249;+624)M3 mutant was generated via splicing by overlapping extension (SOE), using the wild type promoter as a template and mutating site IIIab using the following primers: 196b(M3)for 5'-TGCCTTTAGGCATCTATTTAAATATGCCCTAGACGGTCGGCATT-TGTCTGGGCCCTATAG-3'; 196b(M3)rev: 5'-ATGCCGACCGTCTAGGGCATATTTAAATA-GATGCCTAAAGGCATTGCAGGGGGCTCGGG-3'. Starting from the wild type promoter and from the p196b(-249;+624)M3 constructs, respectively, p196b(-249;+624)M4 and p196b(-249;+624)M3/4 were generated using the following primers: 196b(M4)for: 5'-TTTGTTCAGGGGG-CGTCATTGCTAGTATATCTGGATTG-3'; 196b(M4)rev: 5'-TACTAGCAATGACGCCCCCTG-AACAAAAGCCGACAAACC-3'. All mutant fragments were cloned into the HindIII site of pXP2.

The promoter sequences described above were also cloned into the pGL3basic luciferase reporter vector (Promega) using the same strategies. The reporter constructs used in electroporation experiments were generated using the pNASSβ reporter vector (Clontech). They were generated by cloning the PCR fragments described above into the EcoRI-blunted site of pNASSβ. Following the same strategy, the wild-type promoter was cloned in addition into pNASSβ in antisense orientation.

Expression vectors for PBX1a, HOXB1, HOXD8, HOXD10, and HOXD13 were described previously (29-31). The pSG-FHOXA9 expression construct was obtained by PCR amplification of the human HOXA9 coding sequence, which was cloned into the BamHI site of the pSG5-FLAG vector.

2.2. Whole mount in situ hybridization (WISH)

WISH was performed using the method described in (32) with some minor modifications. Embryos at 17-25 Hamburger-Hamilton (HH) (33) stages were dissected in PBS and fixed

overnight in 4% paraformaldehyde. Embryos were then permeabilized with proteinase K and hybridized with 5nM miR-196 digoxygenin-labeled LNA probe (Exiqon) in hybridization buffer (50% formamide, 5X SSC pH 7, 0,1% Tween-20, 50 µg/mL total yeast RNA, 50 µg/mL heparin, 0,1% CHAPS, 5mM EDTA and 2% blocking powder) overnight at 44°C.

2.3. Chromatin immunoprecipitation (ChIP)

ChIP assays were performed as described in (34). Briefly, formaldehyde-crosslinked chromatin was extracted from caudal trunk and forebrain dissected from ~200 E10 mouse embryos, and subjected to immunoprecipitation using 1 µg anti-me3K4 (Abcam, cat#ab8580), 5 µg anti-me3K27 (Abcam, cat#ab6002), 5 µg anti-AcK27 (Abcam, cat#ab4729), 5 µg anti-HOXD11 (Abcam, cat#ab60715) (35), 5 µg anti-HOXD13 (Abcam, cat#ab19866), 5 µl anti-CDX2 (Biogenex, cat#MU392A-UC clone Cdx2-88), 5 µg anti-HOXA9 (Upstate, cat#07-178) (36) and 5 µl anti-HOXD10 (37), 5 µg anti-FLAG (F3165, Sigma, clone M2) was used as control antibody. The enrichment of amplified DNA sequences in immunoprecipitates was calculated as a percentage of total input chromatin. Mean enrichments for each ChIP were assessed using at least three independent ChIP quantitative real-time PCR (qPCR) results. Primers used in qPCR analysis were: miR196b F_: 5'-GAGCCCCCTGCAATTACTTT-3'; miR196b R_: 5'-CAAGACAGCCACAATCCAGA-3'; CTRL1 F_: 5'-CTCCGCCGCTCTCATTCTC-3'; CTRL1 R_: 5'- ACGCTTGACACT-CACACTTTG-3'; CTRL2 F_: 5'-GGAGGGATTCAAAGTCAGTGTC-3'; CTRL2 R_: 5'-CTGCCTT-ATCTTGGAGCCATC-3'.

2.4. RRL, nuclear extracts, and EMSAs

CDX2 and human 5'HOX proteins were synthesized *in vitro* using the TNT-coupled transcription/translation system (Promega). In EMSAs the total amount of reticulocyte lysate utilized was adjusted to normalize for translated protein content, based on the amount of ^{35}S-Met incorporated and on the number of Met residues for each protein. The amount of the control, "empty" RRL used corresponded to the highest amount of total RRL used in the other binding reactions for each experiment. The binding reactions were performed as described previously by (38). Briefly, proteins were pre-incubated with 100 ng poly-(dI-dC) in 1x binding buffer (0.1 M KCl, 2 mM MgCl2, 4 mM spermidine, 0.1 mg/ml BSA) for 15 minutes on ice. 20,000-50,000 cpm of ^{32}P-labelled probe were then added and samples were incubated for 30 minutes on ice. For supershift experiments, 2 µg of the indicated antibodies, or 2 µL αCDX2 ascite (Biogenex), were added to the binding reactions. In competition assays, a 500-fold excess of cold probe was used.

Reactions were separated on 6% polyacrylamide gels in 0.5x TBE at 4°C. The oligonucleotides probes used are:

SITE I: 5'-GAAGCTTTTTAATTTCACCTCTTTTTTGACTTCTTATCTTCACCTTC-3';

SITE IIab: 5'-CACCTTCCATTTTATCCTGTCCACCACTTTTACAACAGGAGGGC-3';

SITE IIIab: 5'-CTGCAATTACTTTAGGCATCTATTTAAATATTACCTAGACGGTCGT-AATTTGTCTGG-3';

SITE IV: 5'-CGGCTTTTGTTCAGTTTTATGACTTGCTAGTATATCTGG-3';

SITE V: 5'-CGCAGGAGCCAGTTATTTTGCGGGTATCCAGGTCCC-3'.

SITE IIa MUT: 5'-CACCTTCCAT**GGGCGA**CTGTCCACCACTTTTACAACAGGAGGGC-3'

SITE IIb MUT: 5'-CACCTTCCATTTTATCCTGTCCACCAC**GGGGCA**AACAGGAGGGC-3'

SITE IIab MUT: 5'-CACCTTCCAT**GGGCGA**CTGTCCACCAC**GGGGCA**AACAGGAGGGC-3'

SITE IIIa MUT:

5'CTGCAAT**GCA**TTTAGGCATCTATTTAAATATTACCTAGACGGTCGTAATTTGTCTGG-3'

SITE IIIb MUT: 5'-

CTGCAATTACTTTAGGCATCTATTTAAATAT**GCA**CTAGACGGTCGTAATTTGTCTGG-3'

SITE IIIab MUT: 5'-

CTGCAAT**GCA**TTTAGGCATCTATTTAAATAT**GCA**CTAGACGGTCGTAATTTGTCTGG-3'

SITE IV MUT: 5'-CGGCTTTTGTTCAG**GGGGC**GTCATTGCTAGTATATCTGG-3'

Nuclear extracts from HEK293 cells were prepared by lysing them in cytoplasmic buffer (10mM HEPES pH 7,9, 1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT) and subsequent extraction in nuclear buffer (20mM HEPES pH 7.9, 420mM NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA, 0.5mM PMSF and 0.5mM DTT) (39).

2.5. 5' and 3' rapid amplification of cDNA ends (RACE)

For 5' RACE experiments, total RNA was isolated from posterior trunk of E10.5 CD1 mouse embryos by using TRIzol (Invitrogen) and treated with RQ1 DNAse (Promega) before the reverse-transcription step with 196b-5'-GSP1. A tail of dCTP was then added to the cDNA by using terminal deoxynucleotidyl transferase (TdT, Promega). The tailed DNA was initially amplified with a dG-adapter primer and 196b-5'-GSP2 and finally nested-amplified with an abridged primer (AUAP) anchoring to the adapter and 196b-5'-GSP3 primer. The human *miR-196b* promoter was investigated with the same strategy using three human gene-specific primers and RNA from HEK293 cells.

In 3'-RACE experiments, cDNA was retrotranscribed with dT-anchor primer annealing on the poly-A tail of the pri-miR-196b, initially amplified with AUAP and 196b-3'-GSP1 and finally nested-amplified with the AUAP and 196b-3'-GSP2 primers. Both 5'- and 3'-RACE products were visualized by gel electrophoresis , cloned into the pBluescipt SK+ vector and sequence verified.
The primers used in RACE experiments are listed below:
196b-5'-GSP1: 5'-TGATGCTGCGGGCGTAGGG-3';
dG-adapter: 5'-GGCCACGCGTCGACTAGTACGGGGGGGGGGGGGGG-3';
196b-5'-GSP2: 5'-GGAAACGCGTGGCCTAGCGG-3';
AUAP: 5'-GGCCACGCGTCGACTAGTAC-3';
196b-5'-GSP3: 5'-CCCTCCTCCGCGGATTTCGGG-3'.
hsa 196b-5'-GSP1: 5'-CAGTAAGGTATTCCTGGGAGG-3';
hsa 196b-5'-GSP2: 5'-ACTCTGGCAAGCAGGAAACG-3';
hsa 196b-5'-GSP3: 5'-ATGCTGCAGAGAGATGCGCTGTGGGCCAG-3';
dT-anchor primer: 5'-GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTT-3';
196b-3'-GSP1: 5'-GTCGCCCGACTACTGAGAAC-3';
196b-3'-GSP2: 5'-ATGCTGCAGTGGATGCAACCACACAACAC-3'.

2.6. Cell culture, transfection and luciferase assays

P19 cells (ATCC# CRL-1825) mouse embryonal carcinoma cells were cultured in Mininum Essential Medium Alpha (αMEM, Gibco) with nuleosides. HEK293 (ATCC# CRL-1573) and NTERA-2 (ATCC# CRL-1973) human cells were cultured in Dulbecco modified Eagle medium. All the media were supplemented with 10% fetal bovine serum, 2mM L-glutamine, 100U/mL penicillin and 100µg/mL streptomycin. NTERA-2 neuronal differentiation was induced by treatment with 10^{-5} all-trans retinoic acid. Transfections were carried out by $CaPO_4$ precipitation. In a typical experiment, P19 cells were transfected with 2,5-6 µg of reporter plasmid, 0,12-0,25 µg of expression construct and 0,1 µg of pCMV-β-Gal (Clontech) per 6cm dish. Cells were washed and 48 hours after transfection lysed and assayed for luciferase and β-galactosidase as described in (30).

2.7. In ovo electroporation

Chicken eggs were incubated in a humidified incubator at 37°C. Embryos at stage HH10-12 were injected and electroporated into the neural tube as described by (40) Construct concentrations were as follows: 1.0 µg/µL for pNASSβ::lacZ reporter construct and 0.5 µg/µL

for pCMV::eGFP, coinjected as positive control of electroporated tissues. Embryos were collected 24 hrs after electroporation and subjected to β-galactosidase staining.

2.8. siRNA depletion of CDX2

Three different small interfering RNA (siRNA) duplexes were designed targeting the human CDX2 mRNA (Stealth siRNA; Life Technologies). siRNA and CDX2 transfection were performed in induced NTERA-2(RA14) with 100nM RNAi duplexes using Lipofectamine 2000 (Invitrogen) according to the manufacturer instructions. CDX2 depletion and overexpression were verified by immunoblotting 48 h after siRNA transfection. Whole cell extracts were prepared through repeated freeze and thaw cycles and extraction in 10mM Hepes PH7.9, 100mM NaCl, 0,1mM EGTA, 5% glycerol and proteases and phosphatases inhibitors. Immunoblotting analyses were performed with anti-CDX2 antibody (Biogenex, cat#MU392A-UC clone Cdx2-88) and anti-α-Tubulin (Sigma, T6074, clone B-5-1-2). Protein expression in immunoblots was quantified using the ImageJ software.

2.9. RNA extraction and miRNA RT-qPCR detection

Total RNA from dissected embryonic tissues or cultured cells was isolated by TRIzol (Invitrogen) according to the manufacturer instructions. Mature miR-196b and the endogenous control RNU6B (Suppl. Fig. 1B) were reverse transcribed and detected in qPCR with TaqMan® microRNA assays (Applied Biosystems). Real-time (qPCR) relative quantification was performed according to the ΔΔCt method (41). Synthesis of cDNA was done starting from 3 µg of total RNA using M-MLV reverse transcriptase (Promega). Quantitative PCR was performed with the following oligonucleotides: hCDX2 F: 5'-GTGCTAAACCCCACCGTCAC-3'; hCDX2 R: 5'-CTGAGGAGTCTAGCAGAGTC-3'; hACTB F: 5'-AGGCACCAGGGCGTGAT-3'; hACTB R: 5'-GCCCACATAGGAATCCTTCTGAC-3'.

Statistical analysis

Results are expressed as mean ± SEM from at least triplicate experiments. Statistical analyses were performed with GraphPad Prism version 5.0. In transfection experiments the non-parametric Kruskal-Wallis with by Dunn's post test were used. ChIP statistical analyses were performed with the Student's t test.

3. RESULTS

3.1. miR-196 displays a dynamic expression pattern along the anterior-posterior axis during chick embryogenesis

miR-196 expression during chick embryonic development was analysed by whole-mount in situ hybridisation (WISH) using a locked nucleic acid (LNA)-modified oligonucleotide probe (Exiqon). At Hamburger-Hamilton (HH) (33) stage HH17 *miR-196* expression was relatively weak and confined to the caudal-most regions of the embryo within the neural tube and flanking tissue (Fig. 1 panels A and B). At HH stage 18, the expression of *miR-196* was consistently increased and extended more anteriorly, with an expression boundary located posterior to the hindlimb buds (Fig. 1 panels C and D). The maximal *miR-196* expression levels were reached at HH stage 19, where *miR-196* was uniformly expressed in the posterior chick embryo including the hind limbs (Fig. 1E and F). At this stage the anterior-most boundary of expression was found to lie within the posterior trunk, approximately at the level of somites 27-28. At later stages (HH 20-24) *miR-196* expression remained high with an anterior boundary within the lumbar region (somites 27-28, at HH 24). At HH24, within the neural tube, the anterior expression boundary was consistently more rostral (by approx. 3-4 somites) than in the flanking tissues (Fig. 1 I and J). At this stage, a possible additional, weaker expression domain could be detected at the posterior margin of the forelimbs (Fig. 1 I and J). At HH stage 25 *miR-196* expression was still elevated in the whole caudal portion of the embryo (Fig. 1 K and L). In conclusion, *miR-196* expression revealed to be dynamic, both in time and spatial restriction, gradually increasing and extending towards more anterior territories from HH stage 17 onwards, to reach a peak of expression at stages HH 19-20. The posteriorly restricted expression of *miR-196* overlaps the previously reported embryonic expression of 5'Hox genes in the neural tube and flanking tissues (reviewed in 10,13,42), and of Cdx2 (43).

miR-196b expression during mouse embryogenesis was analyzed using real-time reverse transcriptase PCR (qRT-PCR) on dissected embryos at different developmental stages (Suppl. Fig. 1). At E8.5, the earliest stage analysed, expression was relatively low and confined to the posterior portion of the trunk. At later stages, *miR-196b* expression gradually increased, reaching its maximal expression at E10.5 in the posterior trunk (Suppl. Fig. 1). A low level of expression was observed also in the anterior trunk at E10.5 and at later stages (Suppl. Fig. 1). From E11.5 onwards the expression in the posterior trunk remained approximately constant up to stage E14.5 (Suppl. Fig. 1). These results parallel those obtained in chick embryos, confirming a dynamic, posteriorly-restricted expression of *miR-196b*.

3.2. The pri-miRNA giving rise to miR-196b is transcribed from an independent promoter located in the Hoxa10-Hoxa9 intergenic region

To characterise the murine primary miRNA (pri-miRNA) originating the mature *miR-196b*, we mapped its boundaries within the intergenic region between the *Hoxa10* and *Hoxa9* genes, located on chromosome 6, by 5' and 3' RACE (Fig. 2). Three different primers (5'GSP1, 5'GSP2, 5'GSP3, Fig. 2A) allowed us to identify the *pri-miR-196b* transcription start site (TSS) at a distance of 260 bp from the polyadenylation signal of the *Hoxa10* gene, and of ~780 bp from the mature *miR-196b* sequence (Fig. 2 and Suppl. Fig. 2). A putative TATA box (5'-TATATC-3', see Fig. 3B) was found to be located at 35 bp 5' from the mapped TSS, which itself is located within an Initiator sequence (5'-TCAGTT-3', Fig. 3B) (44). Finally, a CpG island was identified at 118 bp 3' to the TSS.

Similarly, the 3' end of the *pri-miR-196b* transcript was mapped, using two different primers (3'GSP1, 3'GSP2, Fig. 2 and Suppl. Fig. 2), and found to be located at a distance of ~920 bp from the *Hoxa9* gene transcription start site (TSS). A canonical polyadenylation signal (5'-AATAAA-3') was identified in close proximity (19 bp) of the mapped transcript 3' end (Fig. 2 and Suppl. Fig. 2), hinting at a possible polyadenylation of the *pri-miR-196b*. Interestingly, a RIKEN cDNA clone (45), deposited in GeneBank (AK054058) (Fig. 3A), covers most of the length of the mapped transcript further supporting the existence of a long non-coding transcript representing the *pri-miR-196b.*

These results suggest that *miR-196b* originates from a large 2671 bp transcript, starting within the Hoxa10-Hoxa9 intergenic region, which is not in physical continuity with either the *Hoxa10* or *Hoxa9* main primary transcripts (Fig. 3A). Our data also indicate that the *pri-miR-196b* is transcribed from an autonomous promoter, independent of the known *Hoxa10* or *Hoxa9* promoters.

3.3. The miR-196b promoter region contains potential responsive elements for Hox and/or Cdx proteins

To characterise the *miR-196b* promoter, and to identify possible relevant regulatory proteins controlling its activity, we performed an in silico analysis of 873 bp of the *Hoxa10-Hoxa9* intergenic region, spanning from the polyadenylation signal of *Hoxa10* to the TSS of *Hoxa9*. Binding sites for 5' (posterior) Hox and Cdx proteins, as defined in (38) (5'-TTTTATTGG-3' and 5'-TTT(T/A)ACGAG-3') and (46) (5'-A/CTTTATA/G-3'), were searched using the Matinspector software (47). We focused on 5' (posterior) Hox and Cdx proteins, as we hypothesized that these would negatively control the function of 3' located Hox gene products, via an indirect, post-

transcriptional mechanism based on the activation of *miR-196b* expression. In parallel, we performed an interspecies comparison using the UCSC genome browser (48), as we reasoned that evolutionarily sequence conservation might highlight relevant regulatory elements. We could thus identify seven evolutionarily conserved putative binding sites for 5' (posterior) Hox proteins and/or for Cdx proteins (sites I to V, Fig. 3B) within a region of ~330 bp encompassing the *pri-miR-196b* TSS (Fig. 3B). Of these sites IIb, IIIb, and IV showed full sequence conservation in all species analysed (Fig. 3B) whereas sites I, IIa, IIIa, and V showed a more limited conservation (Fig. 3B).

To test whether these sites could indeed bind 5' HOX proteins, we initially performed electrophoretic mobility shift assays (EMSAs) using the HOXD10, HOXD11, and HOXD13 5'HOX proteins produced in rabbit reticulocyte lysates (RRLs) (Fig. 4A). Of the five potential binding sites tested, sites IIIab and IV were bound strongly by HOXD10 and HOXD11, and to a lesser extent by HOXD13, whereas site IIab was bound weakly by all the three HOX proteins tested. Sites I and V showed no or only barely detectable binding (Fig. 4A). We next tested for binding of the CDX2 protein, a member of the CDX ParaHox family of transcription factors. CDX2 also showed a distinct preference for site IV (Fig. 4B) and a weak binding to site IIab (Fig. 4B). The binding of CDX2 to site IV was also verified in EMSAs using nuclear extracts from HEK293 cells (Fig. 4), a cell line which was found to endogenously express CDX2 (data not shown) and *miR-196b* (Fig. 4D). A retarded band, comigrating with the CDX2 shifted complex observed using RRL, was detected in HEK293 cells nuclear extract (Fig. 4C, lane 3). This complex was outcompeted by a molar excess of unlabelled site IV oligonucleotide probe (Fig. 4C, lane 5), but was not affected by a molar excess of unlabelled mutated site IV oligonucleotide (Fig. 4C, lane 6). A specific anti-CDX2 antibody (Biogenex) was able to supershift the CDX2 retarded band (Fig. 4C, lane 4), whereas the addition of the available anti-HOXA9, anti-HOXD10, or anti-HOXD11 antibodies did not alter CDX2 binding (Fig. 4C, lanes 7-9). Anti-HOXA9 addition led to the disruption of a faster migrating complex (Fig. 4C, lane 7), indicating a weak binding of HOXA9 and/or of a paralogy group 9 HOX protein to site IV.

To verify whether the sites identified via in silico analysis are indeed required for the observed binding by the HOXD10, HOXD11, HOXD13, and CDX2 proteins, we tested in EMSAs mutant derivatives (see Materials and Methods) of the oligonucleotide probes representing sites II to IV (Suppl. Figs. 4, 5, and 6). While the mutation of site IIa did not affect the binding by HOXD10, HOXD11, HOXD13, and CDX2 (Suppl. Fig. 4A, B, C, and D, lanes 3), the mutation of site IIb abolished the binding by all proteins tested (Suppl. Fig. 4A, B, C, and D, lanes 6). The mutation of site IIIa led to a substantial reduction of the binding by HOXD10, HOXD11, and HOXD13 (Suppl. Fig. 5A, B, C, and D, lanes 3). The mutation of site IIIb, conversely, did not

significantly affect the binding by HOXD10 and HOXD13, while the binding by HOXD11 was reduced (Suppl. Fig. 5A, B, C, and D, lanes 6). Finally, the mutation of site IV completely abolished the binding by HOXD10, HOXD11, HOXD13, and CDX2 (Suppl. Fig. 6A, B, C, and D, lanes 3).

These results show that the 330 bp region encompassing the *pri-miR-196b* TSS contains several potential 5'HOX and CDX binding sites of which only one, site IV, is strongly bound in vitro by both 5' HOX proteins and CDX2.

A sequence alignment between the mouse *pri-miR-196b* promoter region and the corresponding region of the paralogous *miR-196a-2* locus within the HOXC cluster on mouse chromosome 15 revealed a conservation of binding sites IIIa, IIIb, and IV and of the putative TATA box (Suppl. Fig. 3) together with their relative spacing, suggesting the evolutionary conservation of these relevant regulatory elements following the major genome duplication events that led to the formation of the HOXA and HOXC clusters (49). In addition, this finding suggests that the expression of *miR-196b* and its paralog *miR-196a-2* might be regulated in concert by the same upstream transcription factors. Interestingly, no obvious sequence conservation was detected analyzing the genomic region upstream to the paralogous *miR-196a-1* locus located within the HOXB cluster on chromosome 11, leaving its transcriptional layout and regulation still an open question.

3.4. Cdx2 and Hoxd13 bind the pri-miR-196b promoter region in vivo, during embryogenesis

We next tested, using in vivo chromatin crosslinking and immunoprecipitation (ChIP), whether we could identify, during embryogenesis when *miR-196b* is maximally expressed, the canonical epigenetic hallmarks of promoter activity and the binding of 5' HOX and/or CDX within the genomic region encompassing the TSS. To this end, we prepared crosslinked chromatin from E10 mouse embryos, a stage comparable to HH stage 21 in chick (50) when *miR-196b* is highly expressed (see also 51). E10 mouse embryos were dissected to obtain chromatin from anterior regions (including hind-, mid-, and fore-brain) and from posterior ones (including posterior trunk, hindlimbs and tail) (Fig. 5A). Immunoprecipitations were performed using specific antibodies against trimethylated lysine 4 of histone H3 (me3K4), acetylated lysine 27 of histone H3 (AcK27), and trimethylated lysine 27 of histone H3 (me3K27) (Abcam). The first two modifications represent hallmarks of transcriptionally active promoters, with me3K4 usually found at transcriptional start sites, whereas the latter modification is characteristic of transcriptionally silent regions (52-54). The immunoprecipitated chromatin associated DNA was assayed for enrichment of specific sequences by real-time PCR (qPCR) analysis. Three primer sets were used, one amplifying

a fragment located close to the TSS, another amplifying a control region located within the first intron of the Hoxa9 gene, and a third one amplifying a fragment located within a chromosome 6 gene desert ~75 Mb 3' from the TSS of *miR-196b* (Fig. 5A). Substantial enrichments of the *miR-196b* promoter region were observed with the anti-me3K4 and anti-AcK27 in chromatin extracted from the tail region (Fig. 5B, left panel), whereas considerably lower enrichments were observed in chromatin derived from the anterior portion of the embryo (Fig. 5B, right panel). Conversely, a significant enrichment was observed using anti-me3K27 only in chromatin from the anterior E10 embryo (Fig. 5B, right panel). No, or only comparably low enrichments of the gene desert fragment were observed with all three antibodies (Fig. 5B). These data confirm that the identified *pri-miR-196b* promoter region carries the hallmarks of a transcriptionally active promoter region in the posterior part of E10 mouse embryos, i. e. where and when *miR-196b* is maximally expressed.

We then analysed by ChIP the binding in vivo, in mouse embryos, of select 5' Hox proteins and of the Cdx2 transcription factor. Out of four different 5' Hox proteins (Hoxa9, Hoxd10, Hoxd11, Hoxd13), for which specific, ChIP-grade antibodies were available, only Hoxd13 showed an appreciable binding to the *pri-miR-196b* promoter region in E10 embryo trunk chromatin (Fig. 5C, left panel). Similarly, Cdx2 was found to bind the *pri-miR-196b* promoter region selectively in E10 embryonic posterior trunk chromatin (Fig. 5C, left panel). No significant binding by 5' Hox or Cdx2 was observed at the *pri-miR-196b* promoter region in E10 head chromatin (Fig. 5C, right panel), and both in head and trunk E10 chromatin at the control genomic region (Fig. 5C right and left panels, respectively). At a later stage, in E13.5 embryonic trunk chromatin, Cdx2 binding to the *pri-miR-196b* promoter region was considerably reduced, whereas Hoxd13 binding was unchanged (Fig. 5D).

Thus, among the homeodomain proteins that were found in EMSAs to be potentially capable of binding individual sites within the *pri-miR-196b* putative promoter, Cdx2 and Hoxd13 were found to actually bind it in vivo, during mouse development. These data hence establish Cdx2 and Hoxd13 as possible regulators of *miR-196b* expression during embryogenesis.

3.5. The miR-196b promoter region is capable of driving expression in the chick embryo neural tube

To verify whether the identified *pri-miR-196b* promoter regulatory elements were sufficient and necessary to recapitulate the embryonic expression pattern of *miR-196b*, we carried out in ovo electroporation of chick embryos (40) as this represents a profitable system to rapidly characterise conserved transcriptional regulatory regions. Four different reporter constructs were generated driving the lacZ reporter gene (Fig. 6). The p196b(-249;+624)WT reporter contains a ~870 bp

genomic region spanning the TSS of *pri-miR-196b* (see also Fig. 3B) and comprises all five analysed binding sites (I-V), the p196b(-171;+508) reporter contains a deletion of 77 bp 5' to the TSS and of 116 bp 3' to it, the p196b(-249;+624)MUT3/4 reporter represents the entire 870 bp promoter region carrying DNA-binding disrupting mutations within sites IIIab and IV, and finally, the p196b(-249;+624)INV reporter represents a control construct containing the whole promoter region in inverted orientation (Fig. 6). Stage HH 10 chick embryos were co-electroporated in ovo with the indicated reporter constructs and with a GFP control construct (pCMV::eGFP Fig. 6A). The expression of the reporter gene was monitored 24 hours after electroporation (at stage HH 17-19).

Expression of the p196b(-249;+624)WT lacZ reporter was consistently detected in GFP positive stage HH19 chick embryos in neuroepithelial cells of the neural tube, and showed reproducibly an anterior expression front reaching approximately somites 23-24 (Fig. 6, panels B and C). The p196b(-171;+508) reporter also displayed with high frequency an expression pattern within the neural tube that was superimposable to that of p196b(-249;+624)WT. Conversely, both the p196b(-249;+624)MUT3/4 and the p196b(-249;+624)INV reporters showed little or no expression in the neural tube, respectively (Fig. 6, panels E and F). The results are summarized as a graph in Fig. 6 (panel G).

Overall, these data indicate that a ~680 bp genomic region spanning the TSS of *pri-miR-196b* is sufficient to recapitulate the neural tube expression pattern of *miR-196* and that either 5'HOX/CDX binding site IIIab or site IV, or both, are necessary for driving *miR-196b* transcription in the neural tube.

3.6. CDX2 and 5' HOX proteins but not 3' HOX proteins activate transcription from the pri-miR-196b promoter region

To assess whether CDX2 and 5' HOX proteins could indeed regulate transcription from the *pri-miR-196b* promoter, we generated a luciferase reporter construct, p196b(-249;+624), containing the *pri-miR-196b* promoter region tested in chick embryos. P19 cells were transiently co-transfected with p196b(-249;+624) or with the pLuc control reporter construct, together with increasing amounts of HOXA9, HOXD10, HOXD11, HOXD13, or CDX2 expression constructs (Fig. 7A). All tested 5' HOX proteins could efficiently activate the basal activity of the p196b(-249;+624) reporter, whereas no or only a marginal activation was observed on the control reporter (Fig. 7A). Similarly, CDX2 co-expression led to a significant increase of the p196b(-249;+624) reporter activity (Fig. 7A). We next tested whether two 3' HOX proteins, HOXD8 and HOXB1, could activate transcription from the *pri-miR-196b* promoter. Neither HOXD8 nor HOXB1 if co-

expressed with the p196b(-249;+624) reporter could significantly activate its transcription (Fig. 7A). As 3' located HOX proteins usually require the presence of a Pbx-family co-factor for their optimal transcriptional activation (29,55,56), we also tested the effect of HOXB1 on the p196b(-249;+624) reporter in the presence of PBX1. As a control PBX1 was also co-expressed with HOXA9 (Fig. 7B). While the co-expression of PBX1 with HOXA9 led only to a modest increase in the activation of p196b(-249;+624) (Fig. 7B), but not of the mutated versions of p196b(-249;+624) (Suppl. Fig. 8A), the co-expression of HOXB1 and PBX1 did not significantly alter the reporter basal activity, indicating that the lack of transcriptional activity of HOXB1 on the *pri-miR-196b* promoter did not depend on the absence of co-expressed HOX co-factors. Taken together these data show that of the HOX proteins tested only AbdB-related 5'Hox proved to be able to transactivate the *pri-miR-196b* promoter, whereas HOX proteins from loci located 3' to that of *miR-196* did not regulate its transcription in transient transcription assays.

We also generated luciferase reporter constructs containing mutated derivatives of the *pri-miR-196b* promoter. The p196b(-249;+624)M3, p196b(-249;+624)M4, and p196b(-249;+624)M3/4 reporter constructs contain mutations of site IIIab, site IV, and of both sites IIIab and IV, respectively (Suppl. Fig. 7), while p196b(-171;+508) represents a deletion mutant of both a 77 bp region 5' to the TSS, containing sites I and II, and a region of 116 bp 3' to the TSS (Suppl. Fig. 7). P19 cells were transiently transfected with the mutant reporter constructs together with increasing amounts of the CDX2 or HOXD13 expression constructs. The p196b(-171;+508) reporter showed only a moderate reduction in transcriptional activation by CDX2 with respect to its wild type counterpart (Fig. 8). Similarly, only a slight reduction in luciferase activity was observed with p196b(-249;+624)M3 upon CDX2 expression. Instead, a substantial decrease of transactivation by CDX2 was observed with p196b(-249;+624)M4 or p196b(-249;+624)M3/4 (Fig. 8), indicating that site IV is the principal CDX2 responsive element within the *pri-miR-196b* promoter. Transactivation of the *pri-miR-196b* promoter by 5' HOX proteins, conversely, revealed to require generally more than a single binding site. For instance, HOXA9, HOXD10, HOXD11, or HOXD13 upregulation of both the p196b(-249;+624)M3 and the p196b(-249;+624)M4 mutated reporter constructs was considerably lower than that of their wild type counterpart (p196b(-249;+624)). And the p196b(-171;+508) deletion mutant could not be fully transactivated by HOXA9, HOXD11, and HOXD13 (Fig. 8 and Suppl. Fig. 8).

Taken together these results establish that the evolutionarily conserved sites within the *pri-miR-196b* promoter are relevant for mediating transcriptional control by the CDX2 and 5' HOX proteins. In particular, CDX2 as opposed to 5'HOX proteins revealed to require selectively site IV for regulating *pri-miR-196b* transcription.

3.7. Cdx2 expression is required for endogenous miR-196b transcription

As Cdx2 proved to efficiently transactivate the *pri-miR-196b* promoter in transient transfection assays, we next tested whether its loss-of-function would affect the expression of endogenous *miR-196b*. To this end we took advantage of NTERA-2 human pluripotent embryonal carcinoma (EC) cells (57), which we found to express both CDX2 and *miR-196b* after their induction with retinoic acid (RA) to differentiate into a neuronal cell fate (Fig. 9A and B). NTERA-2 cells, which did not express *CDX2* in their undifferentiated state, showed an increase in *CDX2* expression after 14 days of 10^{-5}M RA treatment, which reached its peak after 21 days of RA induction (Fig. 9B). Similarly, *miR-196b*, which was also not expressed in undifferentiated NTERA-2, was maximally expressed after 14 days of RA treatment (Fig. 9B).

We knocked-down the expression of CDX2 by RNA interference using three different siRNAs (see Methods). CDX2 protein levels in 14d RA-treated NTERA-2 cells were markedly reduced by all of the transfected siRNAs (-60%, -54% and -77%, respectively) (Fig. 9C, top left panel). CDX2 protein decrease correlated with a reduction of mature *miR-196b* transcript levels by ~40% (Fig. 9C right panel). We then tested whether the gain-of-function of *Cdx2* would cause an increase in endogenous *miR-196b* expression. *Cdx2* was overexpressed by transfecting increasing amounts of a *Cdx2* expression construct in 14d RA-treated NTERA-2 cells (Fig. 9C, left bottom panel). The enforced exogenous expression of *Cdx2* was associated with an increase of endogenous *miR-196b* transcript levels proportional to the amount of Cdx2 produced (Fig. 9C, right and left panels). Taken together these results support the conclusion that Cdx2 is a major regulator of *miR-196b* expression in this cell context.

4. DISCUSSION

miRNAs embedded within the four *HOX* gene clusters play relevant roles in embryonic development and cancer pathogenesis (6,7). The transcriptional mechanisms controlling their expression, however, are still largely uncharacterised. We focused on the *miR-196b* locus, as it is located in an evolutionarily highly conserved region between *Hoxa9* and *Hoxa10* on mouse chromosome 6, and, according to available mouse EST clones, it appeared to originate from a long non-coding transcript starting from the *Hoxa10-Hoxa9* intergenic region. In this work, we establish by 5' and 3' RACE mapping that the mature *miR-196b* originates from a large, 2671 bp long, primary, non-coding transcript (*pri-miR-196b*), which starts within a TATA box promoter located at 260 bp 3' from the *Hoxa10* polyadenylation signal. In accordance with the assumption that the identified transcription start site (TSS) represents the bona fide core promoter of *pri-miR-196b*, we found in its vicinity by ChIP a significant enrichment for H3K4me3, a canonical epigenetic hallmark of active promoters (52). Furthermore, a genomic region of ~680 bp, comprising the identified *pri-miR-196b* TSS and two evolutionarily conserved 5'Hox/Cdx binding sites, proved by in ovo electroporation to be sufficient to generate an expression pattern within the chick embryo neural tube that closely recapitulates the developmental expression pattern of *miR-196b*. Thus, although it cannot be excluded that a larger read-through transcript(s) may exist, comprising the *miR-196b* locus and the adjacent *Hoxa10* and/or *Hoxa9* loci (see e. g. 58,59), our results are consistent with the conclusion that *pri-miR-196b* is transcribed from an autonomous promoter, independent of the characterised *Hoxa10* or *Hoxa9* promoters.

While *Hox*-embedded miRNAs are generally held to be expressed co-ordinately both in space and time with their adjacent *Hox* genes, this notion is not always supported by experimental evidence. Indeed, *miR-196a-1* a paralog of *miR-196b* has been shown to have a more restricted spatial expression pattern with respect to its neighbouring gene, *Hoxb9*. And in general, *miR-196* genes appear to have a more posteriorly restricted boundary of expression that the adjacent *Hox9*-paralogy group genes (9,60). This suggests that *miR-196* family members may not be simply regulated by promoters and regulatory elements belonging to adjacent *Hox* genes.

Our findings show a dynamic expression of *miR-196* during chick embryogenesis, starting at early stages in the most posterior portion of the embryonic anterior-posterior axis, and reaching its maximal anterior extension at HH stage 19-20. The *miR-196* posteriorly-restricted expression pattern observed in chick development closely matches the patterns of *miR-196* expression reported during mouse, *X. laevis*, and zebrafish embryogenesis as determined by WISH and/or by using miRNA-responsive sensor transgenes (9,23,27,61). This tempted to speculate that the transcriptional regulatory mechanisms driving *miR-196* expression, and hence the physical

elements they depend on, are evolutionarily conserved. We could indeed identify a number of conserved putative 5'Hox or Cdx transcription factor binding sites within the region spanning the *Hoxa10-Hoxa9* intergenic region and encompassing the *pri-miR-196b* TSS, some of which we found to be able to bind both 5'Hox proteins and Cdx2 in vitro. Of these, three binding sites located in proximity to the *pri-miR-196b* TSS (site IIIab and IV), proved to be critical for *pri-miR-196b* expression in vivo. Their combined mutation caused the complete loss of a reporter construct expression within the neural tube in electroporated chick embryos. Similarly, in cultured cells, mutation of site IIIab and IV, led to a substantial reduction of the transcriptional activation of the *miR-196b* promoter by 5'Hox and Cdx2 gene products. Our results thus implicate both 5'Hox and Cdx2 in the control of *pri-miR-196b* expression. Cdx2 was indeed found to be necessary for the optimal expression of endogenous *miR-196b* in human NTERA-2 EC cells, which, upon retinoic acid induction, express *CDX2* but only some 5'*Hox* (62), thus establishing Cdx2 as a regulator of *miR-196b* expression in this cell context.

Single mutations of either site IIIab or site IV showed that transcriptional activation by 5'Hox proteins (HOXA9, HOXD10, HOXD11, and HOXD13) was affected equally well by site IIIab or site IV mutation, while Cdx2-mediated transactivation was significantly impaired only by site IV mutation. These results correlate with our EMSA experiments showing that 5'Hox bind both site IIIab and IV, whereas Cdx2 binds strongly to site IV and only barely to site IIIab. Thus, among the identified conserved binding sites, site IV can be bona fide considered as both a CDRE (Cdx responsive element) as well as a 5'Hox responsive element. The fact that the *miR-196b* promoter can be bound both by 5'Hox and Cdx proteins would suggest a possible competition or sequentiality in vivo between these transcription factors for its regulation. Indeed, in vivo, selectively in the caudal portion of mouse embryos, the promoter region including site IV was bound at stage E10, both by the Cdx2 protein and by the Hoxd13 5'Hox protein. At E13.5, binding of Cdx2 was essentially undetectable, whereas binding of Hoxd13 persisted, prompting to speculate that at this stage the control of *miR-196b* expression is taken over by paralogy group (PG) 13 Hox proteins.

Regulation of *miR-196b* expression potentially by both 5'*Hox* and by *Cdx* gene products could point to different roles played by this miRNA, pertaining to different phases of anterior-posterior (A-P) axis formation and patterning during embryogenesis. *Cdx* genes have been shown to be crucial for posterior elongation of the main body axis during mouse development (20,21,63-65). Human and mouse genomes contain three paralogous members of the *Cdx* gene family of *ParaHox* transcription factors, *Cdx1*, *Cdx2*, and *Cdx4* (17). While all *Cdx* genes participate in the posterior extension of the A-P axis, *Cdx2* plays a primary role in this process, as *Cdx2* conditional, epiblast restricted mutant mice are posteriorly truncated, lack axial tissues posterior to the forelimbs, and die

in utero at E10.5 (63). Similarly, $Cdx2^{+/-}$;$Cdx4^{-/-}$ mice fail to develop caudal body regions posterior the hindlimbs, and axial extension prematurely terminates at ~E10.5 (21). *Cdx* genes have been shown to act downstream of Fgf signalling to regulate 5' *Hox* genes in posterior embryo regions during the specification of regional identities along the A-P axis (18,19). The gain of function of "central" Hox genes, such as *Hoxb8* and *Hoxa5*, expressed in developing trunk regions of the mouse embryo, has been shown to rescue the phenotype produced by *Cdx* gene mutation (21). Conversely, "posterior", 5' Hox genes belonging to PG 13, such as *Hoxa13*, *Hoxb13*, and *Hoxc13*, have been shown to prematurely arrest axial elongation if expressed precociously, recapitulating the defects observed in *Cdx* mutant embryos (21). These observations led to the conclusion that *Cdx* and "trunk" *Hox* genes promote posterior extension of axial structures by sustaining the growth of progenitor cells within axial tissues. This process would last in time until the onset of "posterior", 5' *Hox* gene expression. The gene products of 5' Hox, in particular of PG 13 *Hox*, would then exert their functional dominance over more anteriorly-expressed, more 3' located, "trunk" Hox proteins to arrest axial extension (see 21).

In this scenario, the binding and activation of the *pri-miR-196b* promoter by *Cdx* and by 5' *Hox* genes could reflect different roles in time played by *miR-196b* in dampening "trunk" Hox gene product expression. During the Cdx-controlled phase of axis formation and elongation regulation of *miR-196b* expression by Cdx would play a role in the fine-tuning of trunk "central" Hox gene product levels. In the subsequent, 5'Hox-controlled phase of axis extension termination, the regulation of *miR-196b* by 5'Hox would play a role in reinforcing the dominant competition, or "posterior prevalence" (see below), of these on trunk, "central" Hox gene activity by contributing to the clearance of their gene products. Intriguingly, transcription of Cdx2 was not significantly reduced in mice precociously expressing PG13 *Hox* genes, pointing to possible post-transcriptional mechanisms in the interference by PG13 5' Hox on Cdx and/or "central" Hox gene function (21).

5'Hox proteins are known to exert a functional dominant role over more "anterior", 3'Hox proteins in body regions in which these are co-expressed, an activity that is known under the name of "posterior prevalence" (66-68). Much has been speculated on the mechanism(s) underlying this feature of the Hox gene products, including the possible intervention of Hox cluster-embedded miRNAs. This latter assumption stems from the observation that Hox-embedded miRNAs preferentially target transcripts originating from Hox genes located 3' to the miRNA loci, while leaving 5' Hox unaffected (see 10). Thus, Hox-embedded miRNAs would act as mediators of posterior prevalence by selectively quenching the amounts of "anterior" Hox gene products, hence their function, in posterior body regions. In this context, our findings, showing binding in vivo during embryogenesis and transcriptional regulation of the *miR-196b* promoter by 5'Hox proteins,

prompt to speculate that these could employ *miR-196* family members, by activating their transcription, to control post-transcriptionally the protein levels of "anterior" 3'Hox, within posterior body regions in which they are co-expressed. Further work will be required to validate this hypothesis.

In conclusion, the results described in this paper establish an essential role for Cdx2 and 5'Hox responsive elements within the *pri-miR-196b* promoter region for the developmental expression of *miR-196*, with possible implications in understanding the misregulation of *miR-196* expression in tumorigenesis as Cdx and Hox genes are also involved in cancer pathogenesis.

5. ACKNOWLEDGEMENTS

Thanks are due to J.Deschamps for the kind gift of the pSG5-CDX2 expression construct. S.F. was the recipient of an EMBO short-term fellowship. This work was supported by grants from the Italian Association for Cancer Research (AIRC) (to V. Z.), and from the Fondazione Cassa di Risparmio di Modena (CRMO) (to V. Z.), as well as from the Swiss National Science Foundation and the Novartis Research Foundation (to F.M.R.).

6. REFERENCES

1. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116, 281-297.
2. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215-233.
3. Calin, G.A. and Croce, C.M. (2006) MicroRNA signatures in human cancers. *Nat Rev Cancer*, 6, 857-866.
4. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H. and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *Embo J*, 23, 4051-4060.
5. Schanen, B.C. and Li, X. (2011) Transcriptional regulation of mammalian miRNA genes. *Genomics*, 97, 1-6.
6. Mansfield, J.H. and McGlinn, E. (2012) Evolution, expression, and developmental function of Hox-embedded miRNAs. *Curr Top Dev Biol*, 99, 31-57.
7. Chen, C., Zhang, Y., Zhang, L., Weakley, S.M. and Yao, Q. (2011) MicroRNA-196: critical roles and clinical applications in development and cancer. *J Cell Mol Med*, 15, 14-23.
8. Tanzer, A., Amemiya, C.T., Kim, C.B. and Stadler, P.F. (2005) Evolution of microRNAs located within Hox gene clusters. *J Exp Zool B Mol Dev Evol*, 304, 75-85.
9. Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G. *et al.* (2004) MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet*, 36, 1079-1083.
10. Yekta, S., Tabin, C.J. and Bartel, D.P. (2008) MicroRNAs in the Hox network: an apparent link to posterior prevalence. *Nat Rev Genet*, 9, 789-796.
11. Yekta, S., Shih, I.H. and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science*, 304, 594-596.
12. Kawasaki, H. and Taira, K. (2004) MicroRNA-196 inhibits HOXB8 expression in myeloid differentiation of HL60 cells. *Nucleic Acids Symp Ser (Oxf)*, 211-212.
13. Mallo, M., Wellik, D.M. and Deschamps, J. (2010) Hox genes and regional patterning of the vertebrate body plan. *Dev Biol*, 344, 7-15.
14. Alexander, T., Nolte, C. and Krumlauf, R. (2009) Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol*, 25, 431-456.
15. Shah, N. and Sukumar, S. (2010) The Hox genes and their roles in oncogenesis. *Nat Rev Cancer*, 10, 361-371.

16. Abramovich, C., Pineault, N., Ohta, H. and Humphries, R.K. (2005) Hox genes: from leukemia to hematopoietic stem cell expansion. *Ann N Y Acad Sci*, 1044, 109-116.
17. Ferrier, D.E., Dewar, K., Cook, A., Chang, J.L., Hill-Force, A. and Amemiya, C. (2005) The chordate ParaHox cluster. *Curr Biol*, 15, R820-822.
18. Isaacs, H.V., Pownall, M.E. and Slack, J.M. (1998) Regulation of Hox gene expression and posterior development by the Xenopus caudal homologue Xcad3. *Embo J*, 17, 3413-3427.
19. Pownall, M.E., Isaacs, H.V. and Slack, J.M. (1998) Two phases of Hox gene regulation during early Xenopus development. *Curr Biol*, 8, 673-676.
20. van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B.I. and Deschamps, J. (2002) Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development*, 129, 2181-2193.
21. Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.N. *et al.* (2009) Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell*, 17, 516-526.
22. Duboule, D. and Morata, G. (1994) Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet*, 10, 358-364.
23. Asli, N.S. and Kessel, M. (2010) Spatiotemporally restricted regulation of generic motor neuron programs by miR-196-mediated repression of Hoxb8. *Dev Biol*, 344, 857-868.
24. Hornstein, E., Mansfield, J.H., Yekta, S., Hu, J.K., Harfe, B.D., McManus, M.T., Baskerville, S., Bartel, D.P. and Tabin, C.J. (2005) The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature*, 438, 671-674.
25. McGlinn, E., Yekta, S., Mansfield, J.H., Soutschek, J., Bartel, D.P. and Tabin, C.J. (2009) In ovo application of antagomiRs indicates a role for miR-196 in patterning the chick axial skeleton through Hox gene regulation. *Proc Natl Acad Sci U S A*, 106, 18610-18615.
26. He, X., Yan, Y.L., DeLaurier, A. and Postlethwait, J.H. (2011) Observation of miRNA gene expression in zebrafish embryos by in situ hybridization to microRNA primary transcripts. *Zebrafish*, 8, 1-8.
27. He, X., Yan, Y.L., Eberhart, J.K., Herpin, A., Wagner, T.U., Schartl, M. and Postlethwait, J.H. (2011) miR-196 regulates axial patterning and pectoral appendage initiation. *Dev Biol*, 357, 463-477.
28. Nordeen, S.K. (1988) Luciferase reporter gene vectors for analysis of promoters and enhancers. *Biotechniques*, 6, 454-457.
29. Di Rocco, G., Mavilio, F. and Zappavigna, V. (1997) Functional dissection of a transcriptionally active, target-specific Hox-Pbx complex. *EMBO J.*, 16, 3644-3654.

30. Fantini, S., Vaccari, G., Brison, N., Debeer, P., Tylzanowski, P. and Zappavigna, V. (2008) A G220V substitution within the N-terminal transcription regulating domain of HOXD13 causes a variant synpolydactyly phenotype. *Hum Mol Genet*.

31. Zappavigna, V., Renucci, A., Izpisua-Belmonte, J.C., Urier, G., Peschle, C. and Duboule, D. (1991) HOX4 genes encode transcription factors with potential auto- and cross-regulatory capacities. *EMBO J.*, 10, 4177-4187.

32. Darnell, D.K., Kaur, S., Stanislaw, S., Konieczka, J.H., Yatskievych, T.A. and Antin, P.B. (2006) MicroRNA expression during chick embryo development. *Dev Dyn*, 235, 3156-3165.

33. Hamburger, V. and Hamilton, H.L. (1992) A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn*, 195, 231-272.

34. Salsi, V. and Zappavigna, V. (2006) Hoxd13 and Hoxa13 Directly Control the Expression of the EphA7 Ephrin Tyrosine Kinase Receptor in Developing Limbs. *J Biol Chem*, 281, 1992-1999.

35. Salsi, V., Ferrari, S., Ferraresi, R., Cossarizza, A., Grande, A. and Zappavigna, V. (2009) HOXD13 binds DNA replication origins to promote origin licensing and is inhibited by geminin. *Mol Cell Biol*, 29, 5775-5788.

36. Hu, Y.L., Passegue, E., Fong, S., Largman, C. and Lawrence, H.J. (2007) Evidence that the Pim1 kinase gene is a direct target of HOXA9. *Blood*, 109, 4732-4738.

37. Capellini, T.D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D., Zappavigna, V. and Selleri, L. (2006) Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development*, 133, 2263-2273.

38. Caronia, G., Goodman, F.R., McKeown, C.M., Scambler, P.J. and Zappavigna, V. (2003) An I47L substitution in the HOXD13 homeodomain causes a novel human limb malformation by producing a selective loss of function. *Development (Cambridge, England)*, 130, 1701-1712.

39. Dignam, J.R., Lebovitz, R.M. and Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, 11, 1475-1489.

40. Itasaki, N., Bel-Vialar, S. and Krumlauf, R. (1999) 'Shocking' developments in chick embryology: electroporation and in ovo gene expression. *Nat Cell Biol*, 1, E203-207.

41. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25, 402-408.

42. Pearson, J.C., Lemons, D. and McGinnis, W. (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet*, 6, 893-904.
43. Beck, F., Erler, T., Russell, A. and James, R. (1995) Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev Dyn*, 204, 219-227.
44. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol Cell Biol*, 14, 116-127.
45. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420, 563-573.
46. Margalit, Y., Yarus, S., Shapira, E., Gruenbaum, Y. and Fainsod, A. (1993) Isolation and characterization of target sequences of the chicken CdxA homeobox gene. *Nucleic Acids Res*, 21, 4915-4922.
47. Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. and Werner, T. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21, 2933-2942.
48. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, 12(6), 996-1006.
49. Garcia-Fernandez, J. (2005) The genesis and evolution of homeobox gene clusters. *Nat Rev Genet*, 6, 881-892.
50. Schneider, B.F. and Norton, S. (1979) Equivalent ages in rat, mouse and chick embryos. *Teratology*, 19, 273-278.
51. Kloosterman, W.P., Wienholds, E., de Bruijn, E., Kauppinen, S. and Plasterk, R.H. (2006) In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat Methods*, 3, 27-29.
52. Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J. and Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature*, 419, 407-411.
53. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298, 1039-1043.
54. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A. *et al.* (2010) Histone H3K27ac

separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107, 21931-21936.

55. Di Rocco, G., Gavalas, A., Popperl, H., Krumlauf, R., Mavilio, F. and Zappavigna, V. (2001) The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. *J Biol Chem*, 276, 20506-20515.
56. Moens, C.B. and Selleri, L. (2006) Hox cofactors in vertebrate development. *Dev Biol*, 291, 193-206.
57. Andrews, P.W. (1988) Human teratocarcinomas. *Biochim. Biophys. Acta*, 948, 17-36.
58. Mainguy, G., Koster, J., Woltering, J., Jansen, H. and Durston, A. (2007) Extensive polycistronism and antisense transcription in the mammalian Hox clusters. *PLoS One*, 2, e356.
59. Tang, G.Q. and Maxwell, E.S. (2008) Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome Res*, 18, 104-112.
60. Burke, A.C., Nelson, C.E., Morgan, B.A. and Tabin, C. (1995) Hox genes and the evolution of vertebrate axial morphology. *Development*, 121, 333-346.
61. Qiu, R., Liu, Y., Wu, J.Y., Liu, K., Mo, W. and He, R. (2009) Misexpression of miR-196a induces eye anomaly in Xenopus laevis. *Brain Res Bull*, 79, 26-31.
62. Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F. and Boncinelli, E. (1991) Differential regulation by retinoic acid of the homeobox genes of the four Hox loci in human embryonal carcinoma cells. *Mech. Dev.*, 33, 215-228.
63. Savory, J.G., Bouchard, N., Pierre, V., Rijli, F.M., De Repentigny, Y., Kothary, R. and Lohnes, D. (2009) Cdx2 regulation of posterior development through non-Hox targets. *Development*, 136, 4099-4110.
64. van de Ven, C., Bialecka, M., Neijts, R., Young, T., Rowland, J.E., Stringer, E.J., Van Rooijen, C., Meijlink, F., Novoa, A., Freund, J.N. *et al.* (2011) Concerted involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone. *Development*, 138, 3451-3462.
65. Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. and Beck, F. (2004) Cdx2 is essential for axial elongation in mouse development. *Proc Natl Acad Sci U S A*, 101, 7641-7645.

66. Duboule, D. (1991) Patterning in the vertebrate limb. *Current Opinion in Genetics and Development*, 1, 211-216.
67. Duboule, D. and Morata, G. (1994) Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.*, 10, 358-364.
68. Tschopp, P. and Duboule, D. (2011) A regulatory 'landscape effect' over the HoxD cluster. *Dev Biol*, 351, 288-296.

7. LEGENDS TO FIGURES

Fig. 1. Temporal and spatial expression pattern of miR-196 in chick embryos. Expression of *miR-196* was assessed by whole-mount *in situ* hybridization (WISH). Mature microRNA expression was detected with a digoxigenin-labelled, LNA-modified probe (Exiqon) revealing all three miR-196 family members in chicken (*miR-196-1*, *miR-196-2*, *miR-196-3*, corresponding to *miR-196a-1*, *miR-196b*, and *miR-196a-2*, respectively, in human and mouse). Their expression was analysed during *in ovo* development at the Hamburger-Hamilton stages (HH) indicated in the panels (**A-L**). For each stage the left panel shows the dorsal view, the right panel the lateral one. For each stage 10-12 embryos were analysed, the presented embryos exemplify the obtained expression patterns. Black arrowheads represent the anterior boundary of *miR-196* expression.

Fig. 2. Identification of primary *miR-196b* 5' and 3' ends. Schematic representation of the mouse *Hoxa9-Hoxa10* genomic region. The *Hoxa10* and *Hoxa9* genes from their TSSs (+1) to their poly-A sequences are represented as grey block arrows. A black block arrowhead indicates the miR-196b pre-miR sequence, located between Hoxa9 and Hoxa10. The indicated *miR-196b* TSS and poly-A sequence were identified using 5'- and 3'-RACE, respectively. The resulting primary transcript (*pri-miR-196b)* and the mRNAs of neighbouring *Hox* genes are indicated on top. Primers used in the PCRs are indicated at the bottom. GSP, Gene Specific Primer.

Fig. 3. The region upstream to the *miR-196b* gene contains putative 5'HOX/CDX2 binding sites. **(A)** Alignment of *pri-miR-196b* and of its 5' promoter region to Chr.6 52231000-52227500 mouse genomic sequence was performed using the UCSC Genome Browser. UCSC genes, mouse mRNA, CpG islands and conservation tracks are shown in dense mode, with each item displayed separately, as indicated on the right. mRNAs are transcribed from left to right. The highest degree of conservation is found within a region of ~300bp located upstream from the transcription start site and in close proximity to the *pre-miR-196b* sequence. **(B)** Sequence comparison of the *miR-196b* promoter region in reference species was performed using the Clustal-Omega software. Seven different putative 5'HOX/CDX2 binding sites (boxed) were identified using the MatInspector software. Grey boxes highlight a canonical TATA box and an Initiator element. A black arrow indicates the mouse TSS, small boxes surmounted by an asterisk indicate the human TSSs. A previously mapped mouse CpG island is indicated by a grey rectangle.

Fig. 4. 5' HOX and CDX2 bind *in vitro* to the *miR-196b* promoter region. **(A, B)** EMSAs using oligonucleotides corresponding to sites I-V of the *miR-196b* promoter region and the

indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). **(C)** EMSA using a site IV oligonucleotide. Nuclear extracts (10 μg) prepared from HEK293, cells expressing CDX2 and all the analyzed 5'HOX proteins, led to the formation of four retarded bands. Black arrowheads indicate specific bands; asterisks point out non-specific bands. A specific band, comigrating with in vitro synthesized CDX2 (lane 1) was supershifted by an anti-Cdx2 antibody (white arrowhead). C2= Cdx2 antibody; A9, D10, D11 indicate antibodies against HOXA9, HOXD10 and HOXD11, respectively. WT= 500-fold excess of wild-type unlabelled competitor; MUT= 500-fold excess of unlabelled competitor carrying a mutation of the TTTATG sequence (site IVMUT see Materials and Methods). **(D)** miR-196b levels detected by qRT-PCR in total RNA from NTERA2, P19, and HEK293 cells. Expression levels were calculated according to the ΔΔCt method using RNU6B as a reference gene.

Fig. 5. 5'HOX and CDX2 bind *in vivo* to the miR-196b promoter. **(A)** Representative mouse embryo with boxes indicating the dissected portions used to prepare chromatin for ChIP assays. Schematic representation of the murine chromosome 6 region encompassing the *miR-196b* locus, showing the primer set amplifying the *miR-196b* promoter region and two alternative negative control regions, one located within the intron of *Hoxa9* (CTRL 1), and the other within a gene desert located ~70Mb 3' to the *miR-196b* locus (CTRL 2). **(B, C, D)** ChIPs on mouse embryo chromatin were performed with anti-me3K4, anti-me3K27, anti-AcK27 antibodies (B), or with anti-HOXA9, anti-HOXD10, anti-HOXD11, anti-HOXD13, and anti-CDX2 antibodies (C). (CA) Control antibody. IP enrichments were analysed by qPCR using oligonucleotides amplifying the three regions indicated. Dark grey bars represent the binding on the miR-196b promoter region, whereas light grey bars represent the enrichment on the two respective control regions. Bars represent the IP/INPUT ratio ±SEM of at least three independent experiments. An asterisk represents $p < 0.05$.

Fig. 6. The miR-196b promoter region recapitulates *in vivo* the endogenous miR-196b expression pattern. Chicken embryos were electroporated at HH10 stage and collected 24h later. A CMV-driven GFP construct was always co-electroporated as a positive control and only embryos showing labelling in the whole neural tube (as shown in **A**) were considered in the analysis and subjected to β-galactosidase staining. Lateral (**B**), and dorsal (**C**) views of embryos electroporated with the *miR-196b* promoter region (p196b(-249;+624)WT). (**D**), Embryos electroporated with a construct, p196b(-171;+508), carrying only the central portion of the promoter. Embryos injected

with mutant constructs, p196b(-249;+624)M3/4 (**E**), and p196b(-249;+624)INV (**F**), displayed no staining. The electroporated constructs are indicated under the respective panels. (**G**), Electroporation results are summarized in a stacked column chart. Expressing embryos are represented as percentage of the total number of analysed embryos (n= 30 for the wild-type construct, n=26 for p196b(-171;+508), and n=20 for the each of the promoter mutants). Expressing embryos were classified in three categories according to staining levels: strong, medium and weak/no expression.

Fig. 7. CDX2 and 5' HOX proteins activate transcription from the pri-miR-196b promoter region. Luciferase activities assayed in cell extracts of P19 murine cells, transiently transfected with the empty, pLuc (light grey) or the p196b(-249;+624) reporter constructs (dark grey), together with increasing amounts (0,12-0,25 μg) of expression constructs for the indicated HOX and CDX2 proteins. **(A)** HOXA9, HOXD10, HOXD11 and HOXD13 were tested as representative members of the 5' or "posterior" subclass of HOX proteins; HOXB1 and HOXD8 as representative members of "anterior" and "central" HOX proteins, respectively. **(B)** Fixed amounts of HOXA9 and HOXB1 (0.12 μg) were co-transfected together with a increasing amounts (0,25-0.5 μg) of a construct expressing their cofactor PBX1a. Bars represent the mean luciferase activity±SEM of at least three independent experiments. * = $p < 0,05$, ** = $p < 0,005$, *** = $p < 0,001$, NS = not statistically significant.

Fig. 8. Central portion of *miR-196b* promoter is crucial for HOX/CDX transactivation. Luciferase activities assayed in cell extracts of P19 murine cells, transiently transfected with increasing amounts (0,12-0,25 μg) of the expression constructs for CDX2 **(A)**, HOXA9 **(B)** or HOXD10 **(C),** together with a fixed amount of empty (light grey) and wild-type or mutant reporter constructs (dark grey) carrying the *pri-miR-196b* promoter sequence. A schematic representation of the transfected constructs is shown in Suppl. Fig. 7. Bars represent the mean luciferase activity±SEM of at least three independent experiments. * = $p < 0,05$, ** = $p < 0,005$.

Fig. 9 Cdx2 expression affects *miR-196b* transcription. NTERA-2 human EC cells were induced to differentiate towards a neural fate by 10^{-5}M all-trans retinoic acid (RA) treatment. (**A**) Immunoblotting of Cdx2 expression using anti-Cdx2 antibodies (Biogenex) of whole-cell extracts from untreated (-) or RA-treated (7, 14, 21 days) NTERA-2 cells. (**B**) *miR-196b* and *CDX2* mRNA levels detected by qRT-PCR in total RNA from NTERA-2 cells. Expression levels were calculated according to the ΔΔCt method using RNU6B and actin as reference genes, respectively, and using untreated NTERA-2 cells as calibrator. (**C**) Top left, immunoblotting of Cdx2 expression after

transfection of NTERA-2 cells treated with RA for 14 days with three different siRNA, or with the control scrambled siRNA (SCR). (NT) Non-transfected cells. Bottom left, immunoblotting to detect Cdx2 protein levels on whole-cell extracts from 14d RA-treated NTERA-2 cells transfected with increasing amounts of a Cdx2 expression construct. Right panel, *miR-196b* expression assayed by qRT-PCR in total RNA from siRNA-transfected or Cdx2-overexpressing NTERA-2 cells. Expression levels were calculated according to the ΔΔCt method using RNU6B as a reference gene and using nontransfected induced NTERA-2(RA-14d) cells as calibrator. Anti-Tubulin antibodies were used in all immunoblottings as a loading control.

LEGENDS TO SUPPLEMENTAL FIGURES

Suppl. Fig. 1. (**A**) Mature miR-196b expression detected by qRT-PCR in total RNA extracted from whole embryo (W) or from tissues dissected from head (H), anterior (AT) or posterior trunk (PT) of mouse embryos at the indicated stages of development. miR-196b expression was normalized using the small RNA RNU6B as a reference endogenous control. (**B**) Expression profile of the internal reference control RNU6B in the tissues and stages analysed showing its constancy.

Suppl. Fig. 2. Sequencing products of 5' and 3' RACE are shown on left and right, respectively. Black arrows indicate TSS and 3' end of transcripts. Two alternative TSSs for *pre-miR-196b* were discovered in humans (termed I and II).

Suppl. Fig. 3. Nucleotide sequence comparison of the *pri-miR-196b* and *pri-miR-196a2* promoter regions in the mouse genome was performed using the Clustal-Omega software. The putative 5'HOX/CDX2 binding sites termed IIIa and IV display a high degree of conservation in mouse. Concomitantly, a canonical TATA box is located at the same relative position in the genome. No initiator element was detected in *pri-miR-196a2* promoter region. The conservation of these elements is highlighted by black rectangles.

Suppl. Fig. 4. EMSAs using oligonucleotides corresponding to site IIab (SITE II WT) and to its derivatives carrying single or compound mutations of sites IIa and IIb (MUT IIa, MUT IIb, and MUT IIab), together with the indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). An arrowhead indicates the specific retarded complex, asterisks indicate non-specific bands.

Suppl. Fig. 5. EMSAs using oligonucleotides corresponding to site IIIab (SITE III WT) and to its derivatives, carrying single or compound mutations of sites IIIa and IIIb (MUT IIIa, MUT IIIb, and MUT IIIab), together with the indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). An arrowhead indicates the specific retarded complex, asterisks indicate non-specific bands.

Suppl. Fig. 6. EMSAs using oligonucleotides corresponding to site IV (SITE IV WT) and to its derivative, carrying a mutation within sites IV (MUT IV), together with the indicated *in vitro* transcribed/translated proteins. The amount of reticulocyte lysate in each binding reaction was adjusted to normalize for translated protein content. (RRL=empty rabbit reticulocyte lysate). An arrowhead indicates the specific retarded complex.

Suppl. Fig. 7. Schematic representation of the genomic fragments containing the *pri-miR-196b* promoter region used to generate luciferase reporter constructs. The annotated 5'HOX/CDX2 binding sites are represented with green vertical rectangles. Red boxes indicate mutations of the corresponding binding sites. A black arrow indicates the TSS.

Suppl. Fig. 8. Luciferase activities assayed in cell extracts of P19 murine cells. **(A)** Cells were transiently transfected with fixed amounts of the HOXA9 or PBX1a expression constructs, or with both HOXA9 and PBX1a expressors, together with a fixed amount of empty (light grey), wild-type, or mutant indicated reporter constructs (dark grey). **(B and C)** P19 cells were transfected with increasing amount of HOXD11 and HOXD13 expression constructs respectively, together with a fixed amount of the indicated reporter constructs. Bars represent the mean luciferase activity±SEM of at least three independent experiments. * = $p < 0,05$, ** = $p < 0,005$.

Figure 1

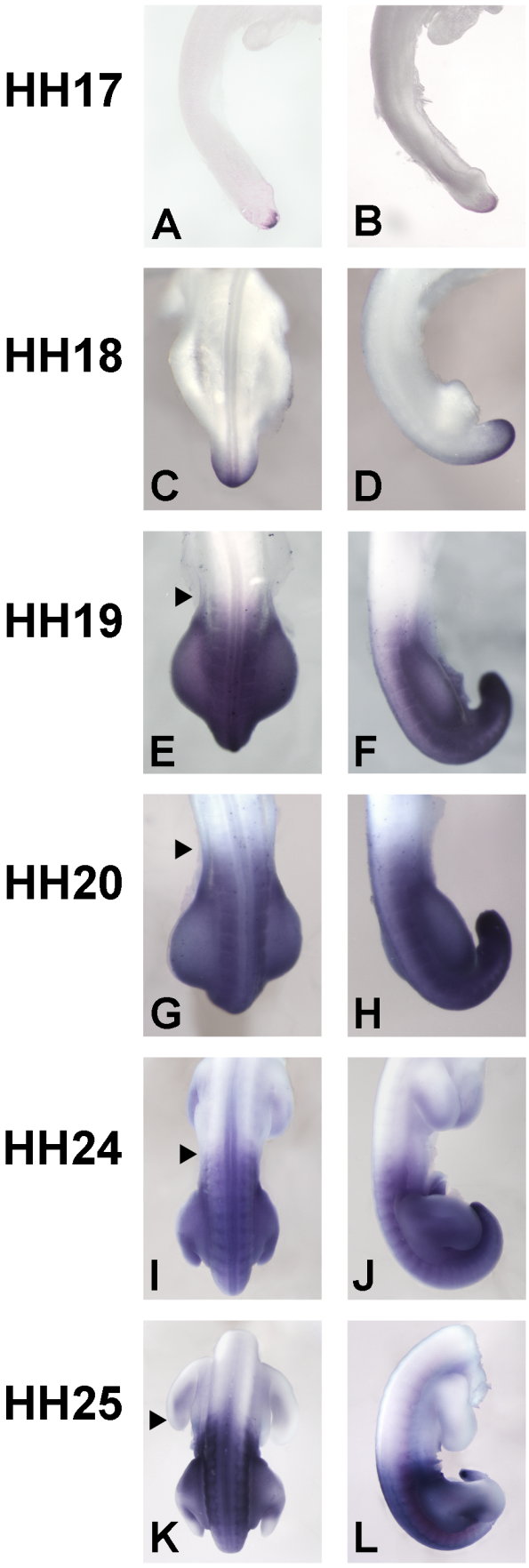


Fig.1

Figure 2

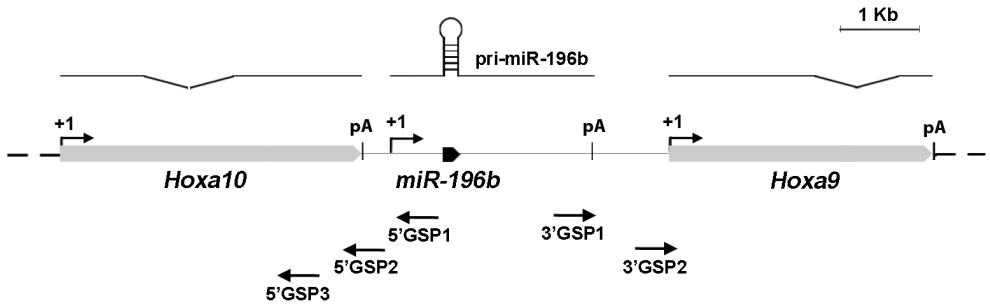
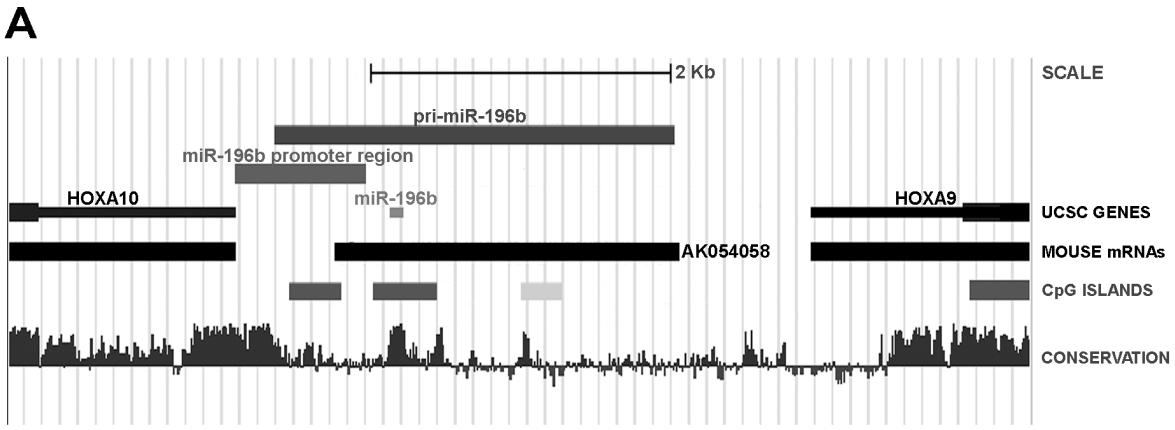


FIG.2

Figure 3



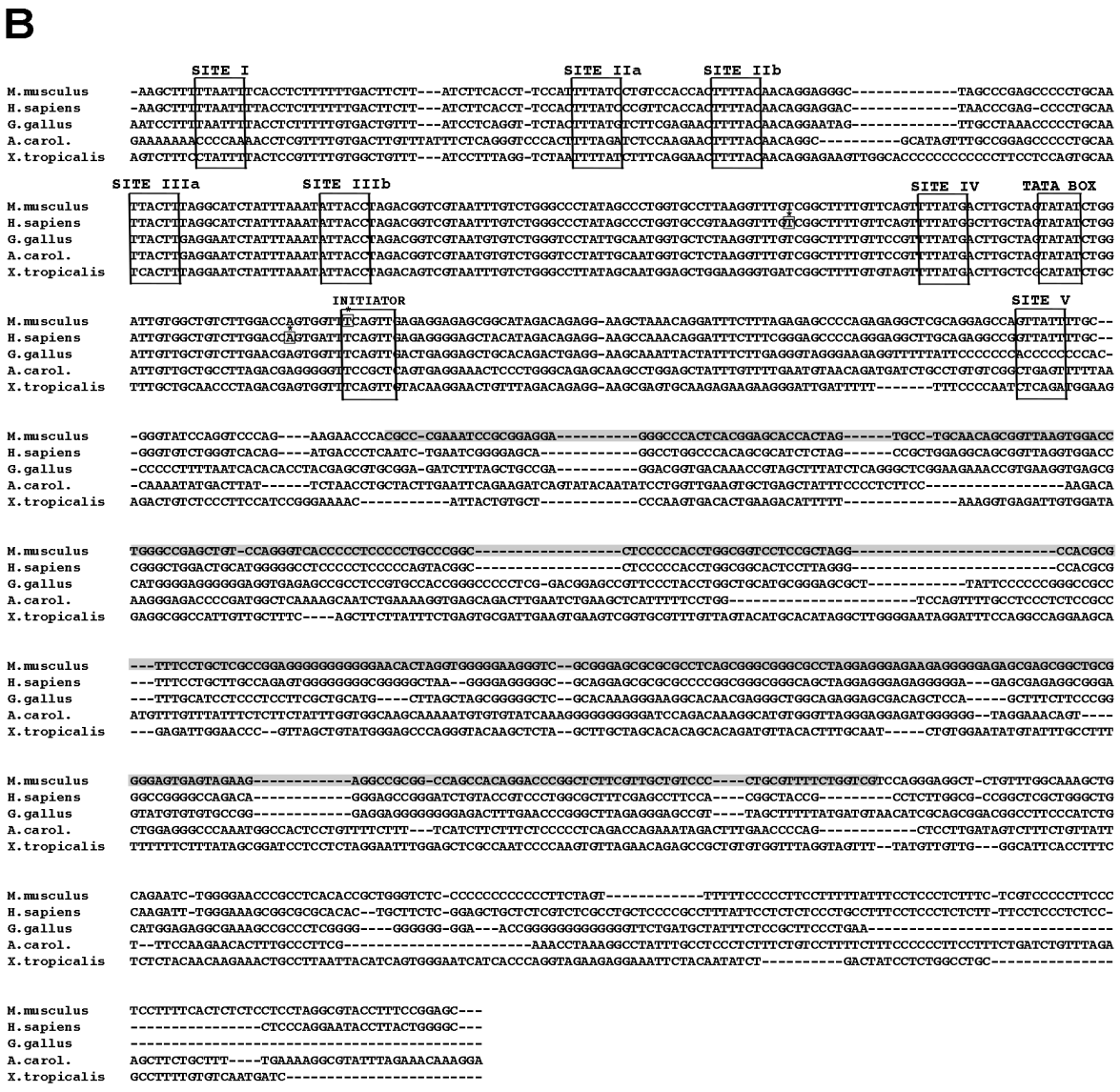


Fig3

Figure 4

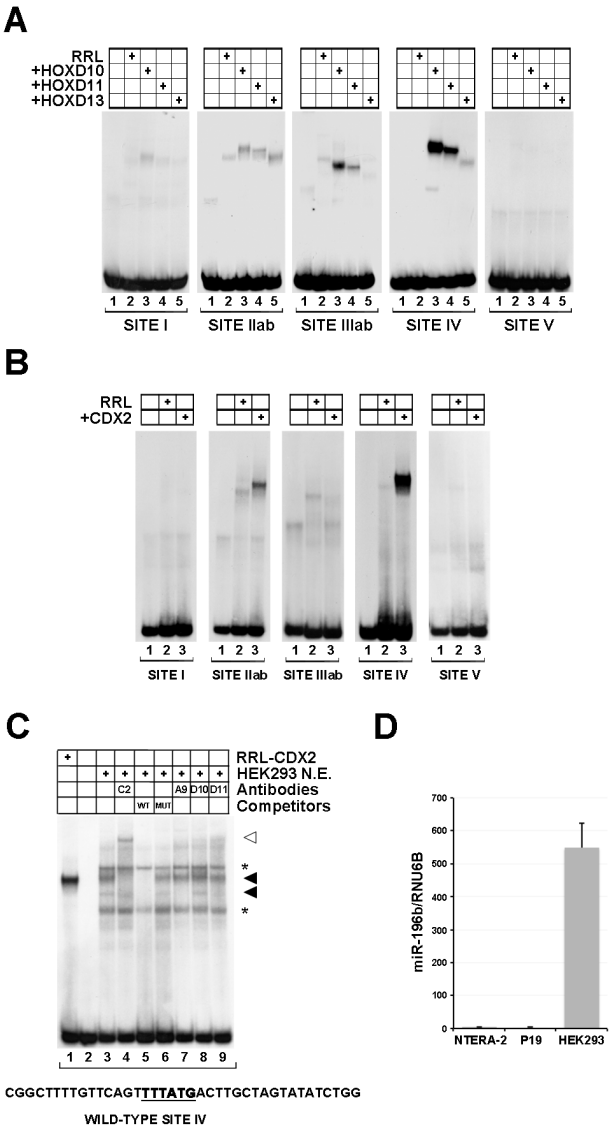


Fig.4

Figure 5

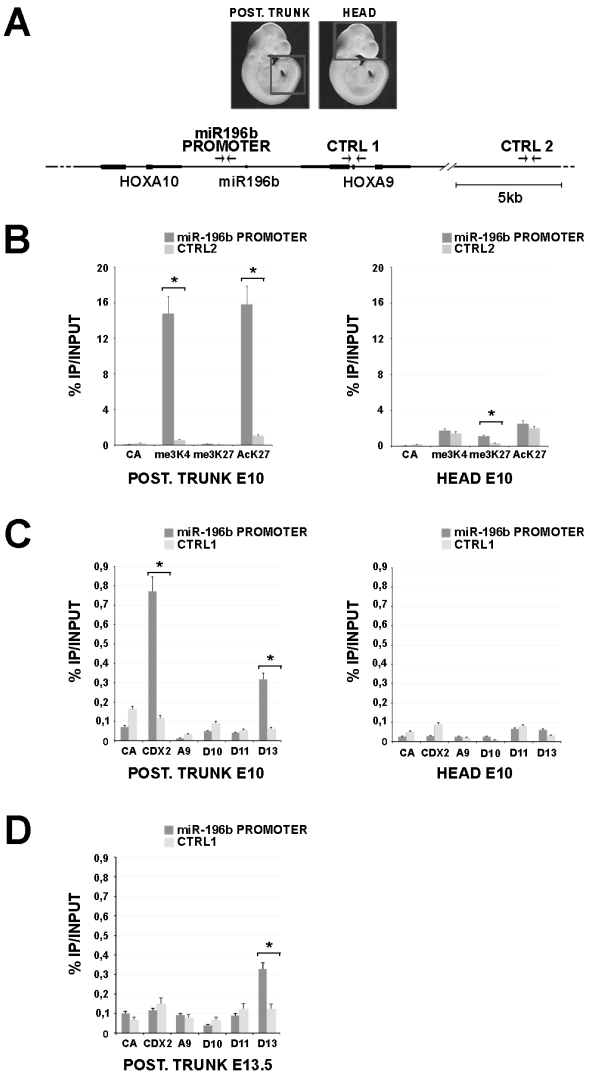


Fig.5

Figure 6

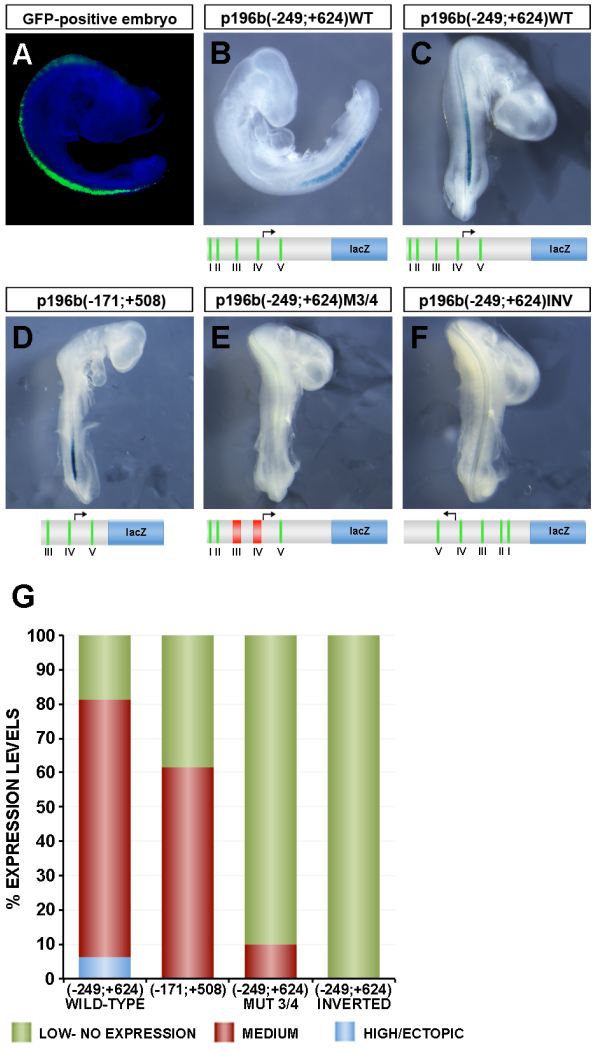


Fig.6

Figure 7A

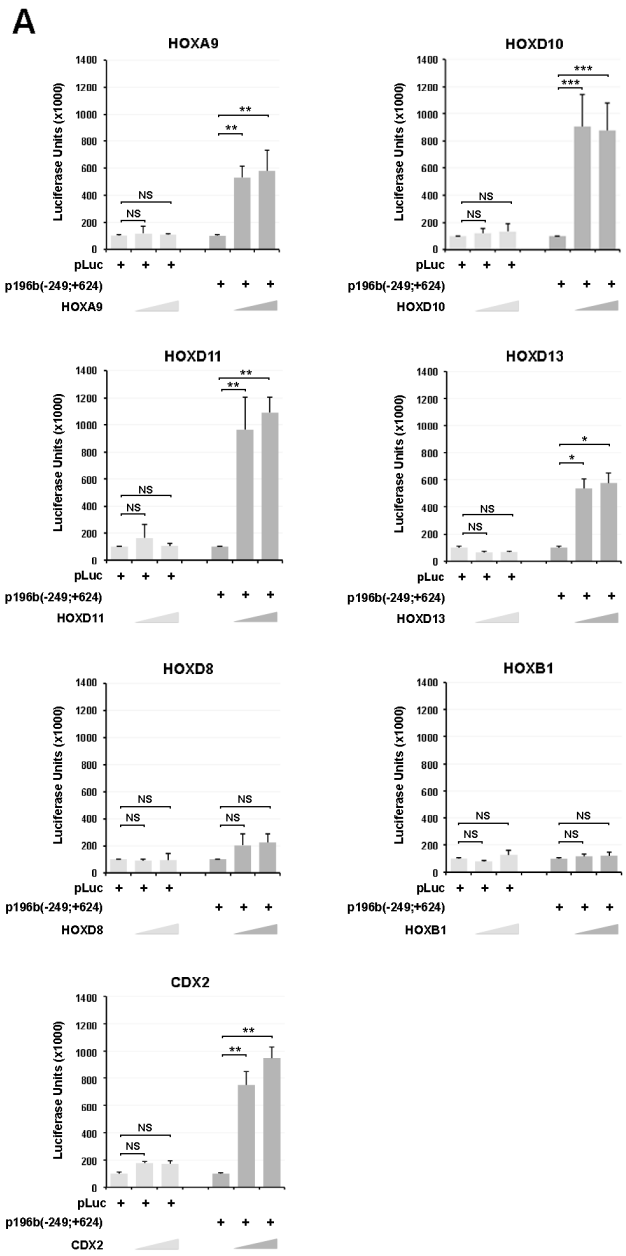


Fig.7A

Figure 7B

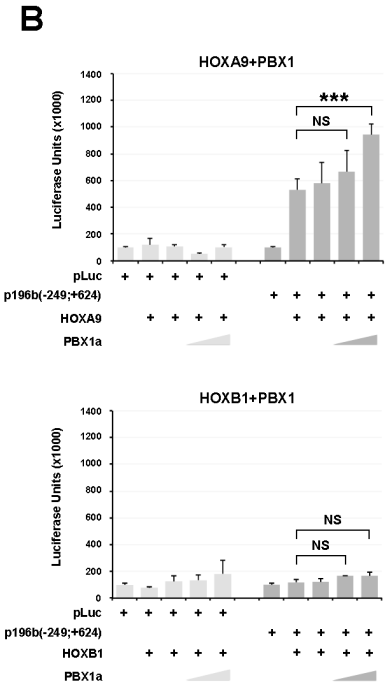


Fig.7B

Figure 8

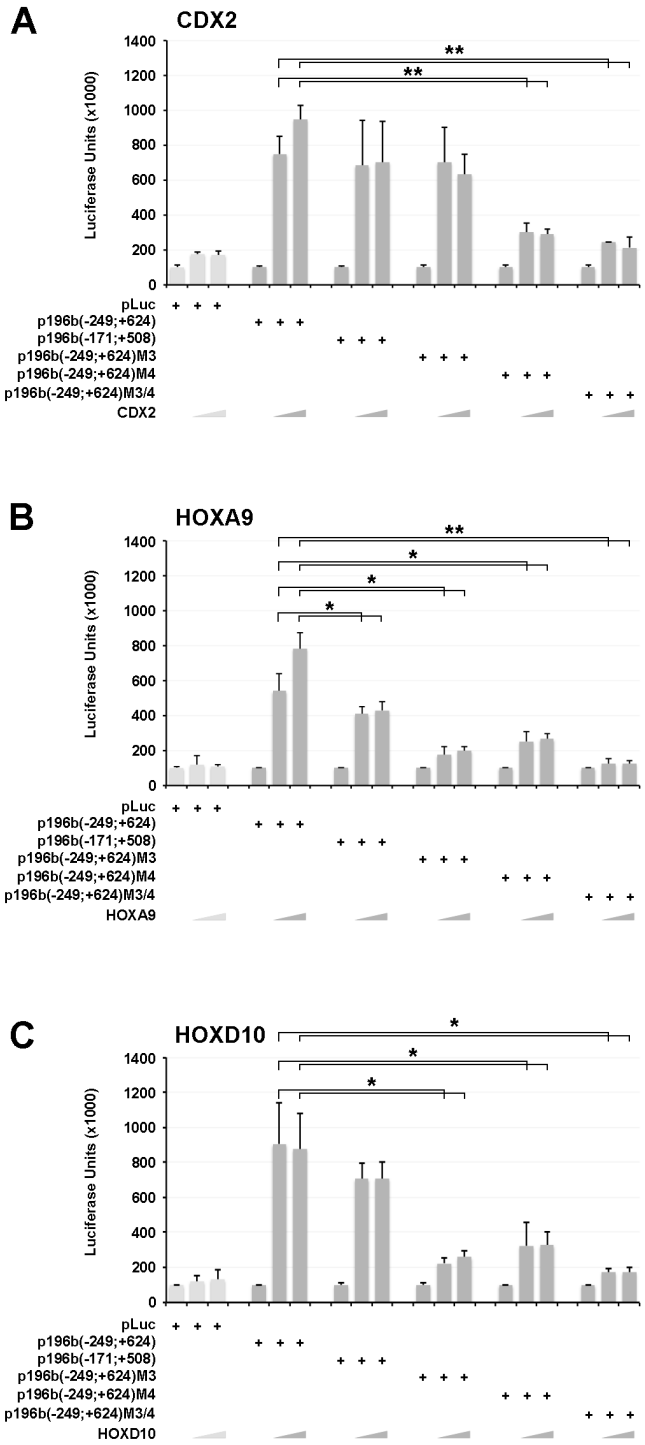


Fig.8

Figure 9

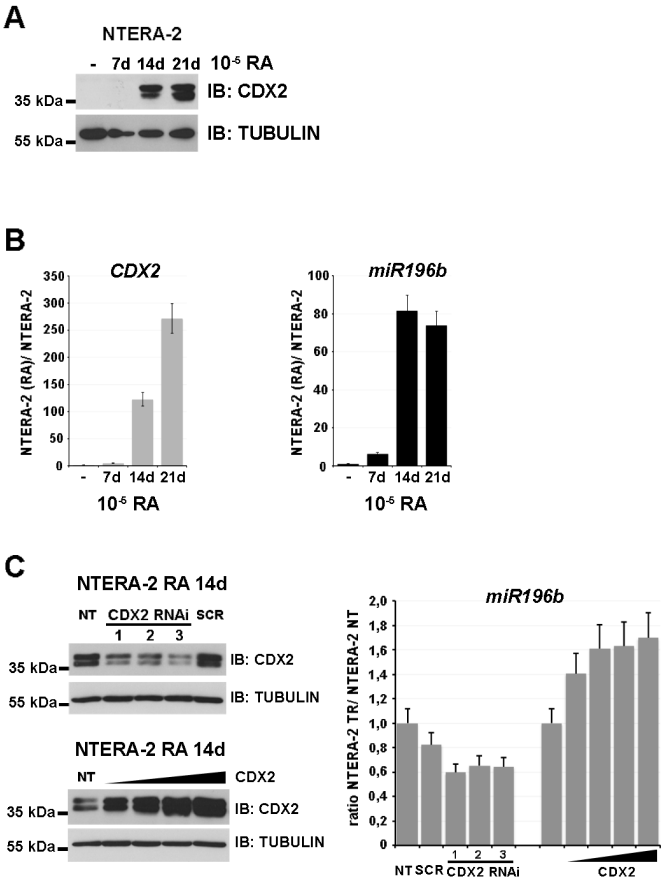


FIG.9

Supplementary Figure 1

Click here to download Supplementary Material (for online publication): Suppl Fig 1.pdf

Supplementary Figure 2

Click here to download Supplementary Material (for online publication): Suppl Fig 2.pdf

Supplementary Figure 3

Click here to download Supplementary Material (for online publication): Suppl Fig 3.pdf

Supplementary Figure 4

Click here to download Supplementary Material (for online publication): Suppl Fig 4.pdf

Supplementary Figure 5

Click here to download Supplementary Material (for online publication): Suppl Fig 5.pdf

Supplementary Figure 6

Click here to download Supplementary Material (for online publication): Suppl Fig 6.pdf

Supplementary Figure 7

Click here to download Supplementary Material (for online publication): Suppl Fig 7.pdf

Supplementary Figure 8

Click here to download Supplementary Material (for online publication): Suppl Fig 8.pdf

***Conflict of Interest**

Click here to download Conflict of Interest: Conflict of Interest declaration.pdf