

Article

Microsyntenic Clusters Reveal Conservation of lncRNAs in Chordates Despite Absence of Sequence Conservation

Carlos Herrera-Úbeda ¹, Marta Marín-Barba ², Enrique Navas-Pérez ¹, Jan Gravemeyer ³, Beatriz Albuixech-Crespo ¹, Grant N. Wheeler ² and Jordi Garcia-Fernández ^{1,*}

¹ Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain

² School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TU, UK

³ German Cancer Research Center, 69120 Heidelberg, Germany

* Correspondence: jordigarcia@ub.edu

Received: 23 July 2019; Accepted: 21 August 2019; Published: 24 August 2019



Abstract: Homologous long non-coding RNAs (lncRNAs) are elusive to identify by sequence similarity due to their fast-evolutionary rate. Here we develop LincOFinder, a pipeline that finds conserved intergenic lncRNAs (lincRNAs) between distant related species by means of microsynteny analyses. Using this tool, we have identified 16 bona fide homologous lincRNAs between the amphioxus and human genomes. We characterized and compared in amphioxus and *Xenopus* the expression domain of one of them, *Hotairm1*, located in the anterior part of the Hox cluster. In addition, we analyzed the function of this lincRNA in *Xenopus*, showing that its disruption produces a severe headless phenotype, most probably by interfering with the regulation of the Hox cluster. Our results strongly suggest that this lincRNA has probably been regulating the Hox cluster since the early origin of chordates. Our work pioneers the use of synteny searches to identify non-coding genes over long evolutionary distances and helps to further understand lncRNA evolution.

Keywords: lncRNAs; genome_evolution; synteny; amphioxus

1. Introduction

Identifying and understanding the factors that underlie the evolution of morphological complexity is one of the central issues in the field of evolutionary developmental biology (or evo-devo). From the initial claims that gene duplication and neofunctionalization were at the core of phenotypic change [1], the current view also takes into account the fine-tuning of gene regulation [2] and increasing the proteome and interactome complexity through additional processes. In this regard, molecular mechanisms such as alternative splicing or RNA-editing, and the RNA world, with molecules like small miRNA or long non coding RNAs (lncRNAs), allow deeper and multifaceted levels of gene regulation [3]. LncRNA-mediated regulation stands out as a quick and efficient mechanism to modulate gene expression, as these molecules are function-ready almost immediately after or even during transcription and can be rapidly degraded by the cellular machinery [3–5]. These characteristics make lncRNAs sharp regulators of the myriad biological processes in which they are involved, such as chromatin remodeling, protein scaffolding or gene expression regulation through direct binding to genomic enhancers [6,7].

The study of lncRNAs from an evolutionary perspective has been hindered by their lack of strong primary sequence conservation [8,9], their apparent lack of secondary structure conservation [10], and their massive genomic generation and decay rate [11]. The cephalochordate *Branchiostoma lanceolatum*

represents the earliest branching chordate lineage and holds a genome that seems to have retained many of the features of the ancestral pre-duplicative vertebrate [12,13]. Searches of lncRNAs conserved between amphioxus and vertebrates based on sequence similarity have been unsuccessful [14,15], probably due to the long evolutionary distance that separates these lineages [16]. Recently, however, a strategy to identify conserved intergenic lncRNAs (lincRNAs) by means of syntenic analyses has been successfully attempted, but limited to closely related species [17,18].

Here, we have developed a pipeline called LincOFinder that finds conserved clusters of microsynteny between two distant organisms surrounding an intergenic lncRNA. Furthermore, we use this tool to study the conservation and evolution of the lincRNA repertoire in the chordate lineage, finding up to 16 lincRNAs putatively conserved between amphioxus and human. Finally, we further study the case of *Hotairm1*, assessing its developmental expression in amphioxus and *Xenopus* and showing that its inhibition during *X. tropicalis* development produces a severe headless phenotype, probably by disrupting the chromatin dynamics of the anterior Hox cluster. Overall, our work pioneers the use of syntenic searches to identify non-coding genes over long evolutionary distances and helps to further understand lincRNA evolution in the frame of the invertebrate-vertebrate transition.

2. Materials and Methods

2.1. Amphioxus and Human Coding and lincRNA Datasets

We used the intergenic and bidirectional fractions from the lncRNAs dataset provided by Marlétaz et al. [15] to obtain an amphioxus lncRNA fraction (1318 genes), and their protein-coding genes supported by orthology as the coding fraction (10,832 genes). The human coding genes were obtained from the Ensembl annotation of the Ch38.96 genome assembly [19]. Finally, the orthologous gene families described in Marlétaz et al. [15] were used to assess the amphioxus/human gene orthologies.

2.2. LincOFinder (lincRNA Orthology Finder)

LincOFinder (<https://github.com/cherreira1990/LincOFinder>) is a program designed to identify shared microsyntenic clusters surrounding lncRNAs between two species. These are named the “reference” (*Ref*) and the “interrogated” species (*Int*). In the first, nonautomated step, the genes of both species (only the coding genes for *Int*) need to be arranged according to their position within the corresponding chromosome or scaffold, then a virtual coordinate according to their position is established (e.g., the first gene in chromosome A will be chrA-1, the second chrA-2, etc.). Furthermore, the orthologies between the genes of both species are established, using sets of known orthologous families or with the help of programs like Orthofinder [20]. Once the data is properly formatted as indicated in the ReadMe.md of LincOFinder, each annotated lncRNA from *Ref* is used as a reference point. The three upstream and three downstream genes neighboring the lncRNA are selected, and the orthology coordinates from *Int* are parsed into a distance matrix (Figure 1). The reasoning behind selecting only the three upstream and three downstream genes is to try to be astringent enough to comply with the definition of microsynteny [21] but at the same time allowing insertions and deletions up to a certain degree and the discovering, in case they exist, of larger clusters. Only genes present in the same chromosome are taken into account for distance assessment, and comparisons between paralogs of the same *Ref* gene are avoided. Then, a UPGMA hierarchical clustering algorithm [22] is used to create viable distance clusters (the ones that comply with the previously stated restrictions), and the cluster with the minimum distance between two neighboring genes is selected, thus identifying possible microsyntenic clusters. If several possible clusters are formed, then they are displayed separately. These microsyntenic clusters should be further filtered by selecting only those that harbor adjacent genes. Finally, candidates are manually curated by looking for the presence of lncRNAs in the microsyntenic region of *Int* (the algorithm is blind to *Int* ncRNAs due to the possibility of missing unannotated syntenic lncRNAs that could be, for example, present in the form of ESTs). This step can be done using a genome browser such us UCSC [23] (Figure 1).

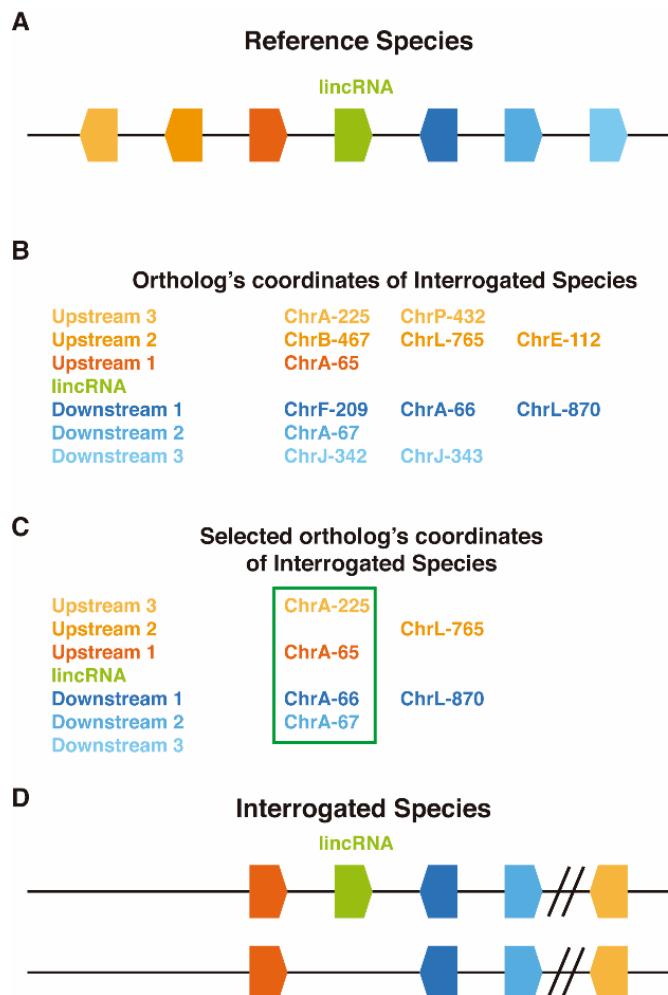


Figure 1. Diagram of LincOFinder mechanism. (A) Representation of the *Ref* species region where a lincRNA is present. (B) Formatted table of orthologs and virtual coordinates from the three upstream and downstream coding genes fed to the algorithm. (C) Selection of the best cluster according to the minimum distance between genes. (D) Representation of a conserved mycosyntenic cluster in the *Int* species, where the presence of a lincRNA is manually confirmed (above) or discarded (below).

2.3. Xenopus Embryos and MO Injections

All experiments were performed in compliance with the relevant laws and institutional guidelines of the University of East Anglia. The research has been approved by the local ethical review committee according to UK Home Office under Project License PPL 70/8876. *X. tropicalis* females were primed 24 h before the eggs were required and induced 5 h prior the experiment both with Chorulon (human Chorionic Gonadotropin). *X. tropicalis* males were also primed with Chorulon. Eggs were naturally obtained and fertilized with a sperm solution in Leibovitz's L-15 Medium supplemented with 10% calf serum and left at room temperature for 5 min. After that, embryos were immersed in 0.05× MMR (Marc's Modified Ringer's) (50 mM NaCl, 1 mM KCl, 0.5 mM MgCl₂, 1 mM CaCl₂, 2.5 mM HEPES pH 7.5) for 20 min at room temperature, then washed in 2% L-cysteine pH = 8 for 7 min and rinsed several times with 0.05× MMR. Embryos were incubated in a BSA (bovine serum albumin) coated Petri dish in 0.05× MMR at 26 °C.

Morpholino oligos (MO) were designed and provided by Gene Tools. Morpholino sequences are: Standard control CCTCTTACCTCAGTTACAATTATA, Hotairm1 AATCACTATTGCTCCTT ACCGGGT. Microinjections were carried out in 3% Ficoll PM400 (Sigma St. Louis, MO, USA) at 2-cell stage in both cells, and then embryos were incubated at 26 °C. Once *Xenopus* embryos reached the

appropriate stage, they were snap frozen in liquid nitrogen for RNA extraction or fixed for whole mount *in situ* hybridization using MEMFA (3.7% formaldehyde, 1× MEM salts), then washed in PBST (PBS, 0.1% Tween), dehydrated in a serial dilution of ethanol and kept at 4 °C or –20 °C.

2.4. Xenopus RNA Extraction, cDNA Synthesis, PCR and Quantitative PCR

RNA was extracted using High Pure RNA isolation kit (Roche Basel, Switzerland) and 1 µg of RNA was taken to synthesize cDNA using Maxima First Strand cDNA synthesis kit (Thermofisher Waltham, MA, USA). RT-PCR was performed using SYBR Green detection method. Primers were designed using Primer3plus; *gapdh* was used as a control housekeeper gene. Primer sequences are indicated in 5'-3' direction.

hoxa1—F: AGAAGTTGCCGGTCTCCTT, R: AAGCCATATTCCCCAGCTT
hoxa4—F: CAGTATCCACCCGAAAAGA, R: GGGTTCCCCTCCACTGTAAAT
hoxa5—F: GTCAGTGCAACCCCAAATCT, R: TTTCCTCTGGCCCTCCTAT
hoxa6—F: GGAAGTACAGCAGCCCTGTC, R: GTAGGTCTGCCTCCCTCTCC
hoxa7—F: GACTCCCATTCCGCATCTA, R: GGTAACGGGTGTAGGTCTGG
gapdh—F: ACTACCGTCCATGCCTTCAC, R: TCAGGGATGACTTCCAAAC

For RT-PCR, ~100 ng of cDNA was used for amplification and the total PCR product was loaded to a 10 mg/mL agarose gel.

P300—F: GATTGCTACACCACCTCTC, R: CCATGGGAGTCTTGACAATC
hotairm1—F1: CACAGTGCAGATGTCAGTGC, F2: CTACGGAGAGATACTTGCAC, R1: ATGCA CGGTGTGATCAGTCG, R2: AAGCAATAACCGAGGCCTCT

2.5. Xenopus Whole-Mount *In Situ* Hybridization

In situ hybridization was carried out as previously described [24]. Probes were synthetized for *X. tropicalis* *hoxa1* and *hotairm1* using the following primers (sequences 5'-3') *hoxa1*F: GATCGTTTGTGGTCGGACG, *hoxa1*R: GCAGCAATTCTACCCCTGCG, *hotairm1*F: CTACGGAGA GATACTTGCAC, *hotairm1*R: AAGCAATAACCGAGGCCTCT.

Otx2 and engrailed vectors for probe synthesis were kindly provided by Professor N. Papalopulu (Manchester).

2.6. Amphioxus Embryo Collection and Whole-Mount *In Situ* Hybridization

Ripe adult amphioxus specimens were collected in Banyuls-sur-mer, France. Spawning was induced as previously described [25] in a dry laboratory. After in vitro fertilization, embryos were cultured at 18 °C until they reach the desired stage and fixed with 4% PFA in MOPS buffer overnight at 4 °C.

The hybridization chain reaction (HCR) [26] *in situ* v3.0 kit by Molecular Instruments (Los Angeles, CA, USA) was used following the protocol provided by the manufacturer for zebrafish embryos, with some adjustments to the probe and hairpins concentration (2pMol and 18pMol respectively) and using nests with a 0.4 µm mesh. The sequence provided for the probe synthesis for *Hotairm1* was the following (5'-3') AAGGAGAGACGAAAGTCACCGGGACAAACCGGAGGATGTCTCGG AGGACCTTACCAACCGCTCCGCCTGTGCTCTACAGGTACCAAGGTGGGATAGCACACATG TCCTCTCTAGACATCTACTACACGCAGCTTGCTACCTGAAAGTTATCATATCTAGAATGTATAT CTGCTTCAGTGTAAAGCAACG.

3. Results and Discussion

3.1. Conserved Microsynteny Clusters

In order to identify conserved lincRNAs across chordates, we developed a pipeline called LincOFinder and used previously described *Branchiostoma lanceolatum-Homo sapiens* orthology families to detect conserved microsynteny clusters around specific amphioxus lincRNAs [15,23] (see Methods).

Here we present the most reliable set of homologous lincRNAs that we were able to produce. Although in our pipeline three upstream and three downstream genes are considered, the output must be trimmed to extract the bona fide orthologous lincRNAs. In this case, we decided to restrict the distance between coding genes to one, and to consider only the clusters formed by one upstream gene, the lincRNA and one downstream gene. From the 32 clusters, only the 16 presented in Table 1 were considered to have a bona fide orthologous lincRNA. We also analyzed under these restrictions the clusters formed by two upstream genes and the lincRNA and by the lincRNA and two downstream genes (Table S1, Table S2). The rate of orthologous lincRNA finding was around 45% in the aforementioned analyses and the whole raw output is available in the supplementary info (File S1). Using this approach, we were able to obtain a list of 16 lincRNAs putatively conserved between human and amphioxus (Table 1). To our knowledge, this list represents the best set of highly curated lincRNAs with the deepest evolutionary conservation reported to date [27]. The main advantage of LincOFinder over other methods based on lincRNA sequence conservation is that it relies on microsyntenic conservation and a proper establishment of interspecific orthology relationships, which are more evolutionary constrained than the highly mutable nucleotide sequences of lincRNAs. In conclusion, LincOFinder can help to uncover conserved lincRNAs over deep evolutionary distances, in any species for which proper gene annotation data is available.

Table 1. Putatively conserved lincRNAs between *Homo sapiens* and *Branchiostoma lanceolatum*. Analysis of the genes surrounding the lincRNA focusing on the three core genes (upstream1, lincRNA and downstream1). Some lincRNAs can be ascribed to more than one hypothetically conserved microsyntenic cluster.

Orthologous lncRNAs ¹	State of the Cluster in Human ²	Human Orthologous lncRNA ³
BL20528 Sc0000000 28 +	* Conserved microsynteny	ENST00000623777.1_1
BL38782 Sc0000000 30 +	* Conserved microsynteny	HOTAIRM1
BL90848 Sc0000001 150 −	Correct order but strands inverted. In addition, there are two lncRNAs surrounding the cluster	AL354977.2
BL79733 Sc0000007 52 +	One gene with the strand inverted	BANCR
BL84418 Sc0000009 86 +	One gene with the strand inverted * Conserved microsynteny	TCONS_12_00008513
BL91140 Sc0000010 144 −	Couple of lncRNAs in amphioxus, synteny conserved in the coding genes, but not in the lincRNA	TCONS_00027655
BL91143 Sc0000010 145 −	* Conserved microsynteny	TCONS_00006308
BL53024 Sc0000015 55 +	One gene with the strand inverted	RP11-181G12.4
BL82992 Sc0000016 15 −	One gene with the strand inverted	TCONS_00027115
BL78145 Sc0000039 54 +	* Conserved microsynteny	TCONS_00000550
BL55463 Sc0000050 45 +	* Conserved microsynteny	TCONS_00024711
BL68900 Sc0000072 2 +	Synteny conserved in the coding genes Problematic region with several lncRNAs and massive distances	TCONS_00011710
BL54861 Sc0000089 4 −	Problematic region with several lncRNAs and massive distances	AI219887
BL41904 Sc0000219 3 −	* Conserved microsynteny	AC109136.1
BL72725 Sc0000229 14 +	* Conserved microsynteny	AC124852.1
BL59605 Sc0000234 6 −	* Conserved microsynteny	TCONS_00007813
BL38170 Sc0000240 5 +	* Conserved microsynteny	BC043517
		LINC00114
		TCONS_00011870
		LOC100132215

¹ GeneID, Scaffold, virtual coordinates and strand separated by "|". ² Description of the synteny of the cluster status in human. ³ ID of the putative human orthologous lncRNA. * indicates a perfect match in strand and order of the three core genes.

3.2. Conservation of HOTAIRM1 across Chordates

HOTAIRM1 was selected for further study because its conservation across several vertebrate lineages has been previously underscored [28,29], and its mechanism of action has been thoroughly studied [30]. HOTAIRM1 was identified for the first time in myelopoietic human cells [31], during a screen for transcriptionally active intergenic elements within the HoxA cluster. In amphioxus it is situated in the Hox cluster between *Hox1* and *Hox2*, and between *Hoxa1* and *Hoxa2* in vertebrates (Figure 2). According to our microsynteny-based analysis, *Hotairm1* is conserved in most of the chordate species analyzed, with the notable exception of zebrafish (Figure 2). Nonetheless, *Hotairm1* appears to be present in other actinopterigians like spotted gar (data not shown), as well as in teleosts like medaka (Figure 2). Given that the Hox cluster is disintegrated in tunicates [32] and due to the absence of an antisense transcript 5' of *Ciona intestinalis Hoxa1*, we were unable to confirm the presence of *Hotairm1* in this chordate subphylum. Finally, we could not find any trace of this lincRNA in the genomes of the cyclostomes *Eptapretus burgeri* and *Petromyzon marinus* possibly because the microsynteny is lost in this region, due to a lineage-specific loss or alternatively because the lncRNA annotation was deficient [33].

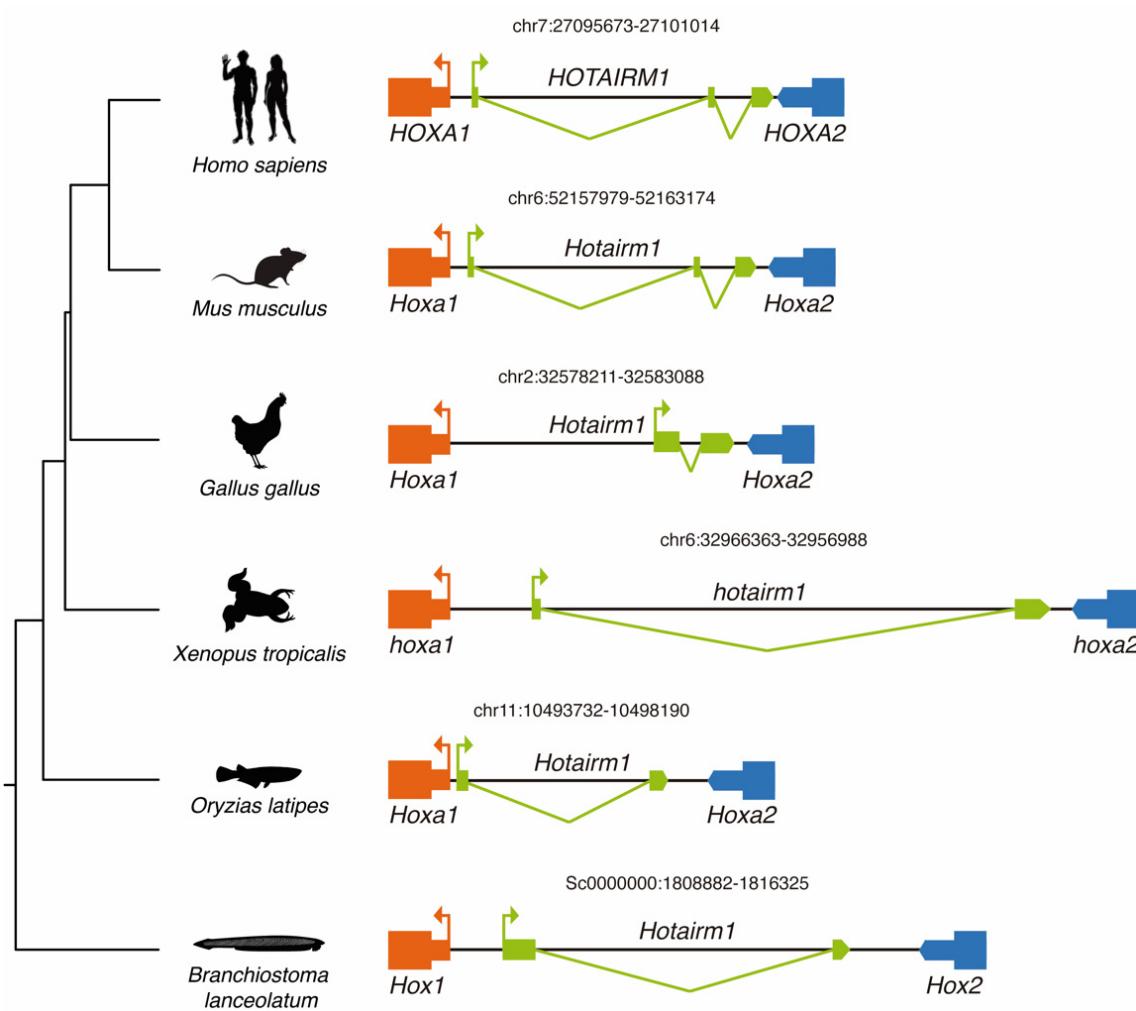


Figure 2. Schematic representation of the genomic locus of *Hotairm1* across several chordate species. Genome or scaffold position is indicated above each *Hotairm1* locus.

Our results add up to previous studies that have described the presence of *Hotairm1* in mammals [28], birds and reptiles [29], strongly suggesting that *Hotairm1* was retained within the HoxA cluster after the vertebrate-specific rounds of genome duplication [34], and that its origin predates, at least, the appearance of extant chordate lineages more than 500 million years ago.

3.3. HOTAIRM1 Expression Patterns in Amphioxus and Xenopus

The expression domain of *Hotairm1* is mostly unknown, although it has been observed to be significantly increased or decreased in several types of cancer [35–37]. Furthermore, its expression is dynamically regulated during neuronal differentiation, showing a sharp increase in early differentiating neurons [38]. According to available RNA-seq data [15], the expression of *Hotairm1* during *B. lanceolatum* development peaks at 27 h post fertilization (hpf) (File S1). To investigate *Hotairm1* expression during amphioxus development, we performed fluorescent in situ hybridization in embryos from 18 hpf to 48 hpf. At 18 hpf we couldn't detect any signal, while at 21 hpf the expression of *Hotairm1* appeared in scattered cells in the presomitic mesoderm and in the neural plate partly overlapping *Hox1* expression domain (data not shown) [39]. At 30 hpf, 36 hpf and 48 hpf *Hotairm1* expression is restricted to the neural tube from the 5th somite towards the anterior developing neural tube, probably reaching the Di-Mesencephalic primordium (DiMes) [39,40] (Figure 3A'). Relevantly, the expression domain of *Hotairm1* in this developmental stage overlaps with *Hox1* which is also expressed in the developing neural tube and localized in the hindbrain (Figure 3C,C') [41,42].

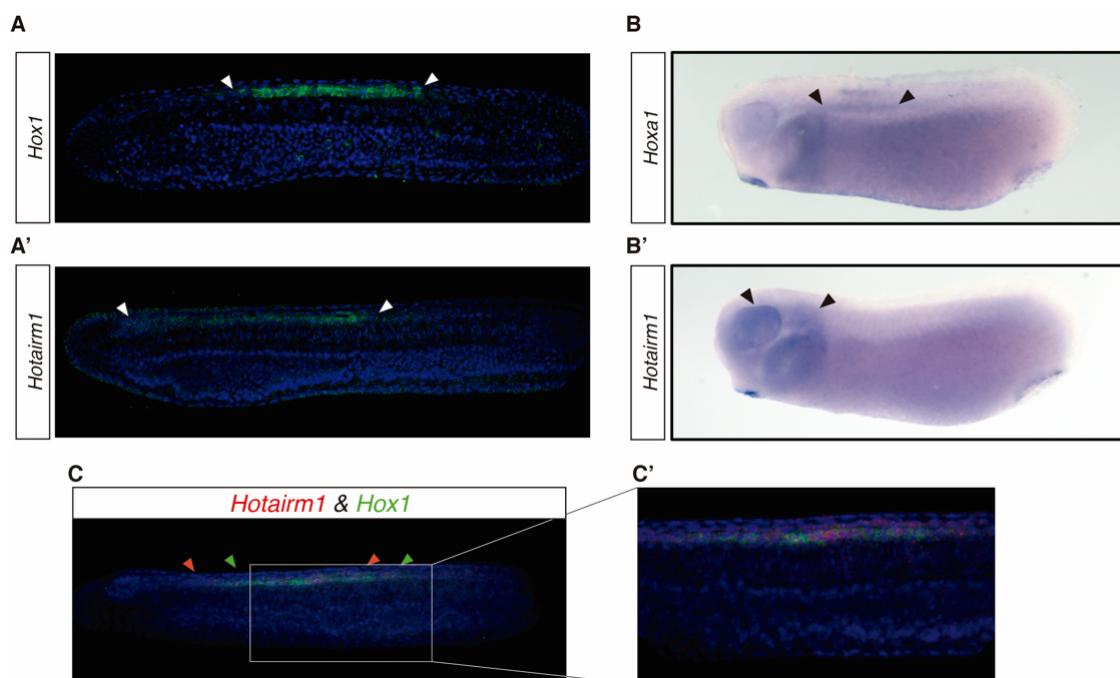


Figure 3. In situ hybridization (*ish*) in *B. lanceolatum* and *X. tropicalis*. Anterior to the left, dorsal is up. (A,A') Fluorescent HCR *ish* in *B. lanceolatum* in whole mount for (A) *Hox1* and (A') *Hotairm1* in 30 hpf embryos. White arrows mark the anterior and posterior limits of the expression domain. (B,B') Colorimetric whole mount *ish* in *X. tropicalis* tadpoles for (B) *hoxa1* and (B') *hotairm1*. Black arrows mark the anterior and posterior limits of the expression domain. (C,C') Fluorescent double HCR *ish* in *B. lanceolatum* in whole mount *ish* of (C) *Hox1* and *Hotairm1* in a 36 hpf embryo and (C') the detailed zone where *Hotairm1* peaks its expression. Green arrows mark the anterior and posterior limits of *Hox1* expression and red arrows mark the ones of *Hotairm1*.

In the vertebrate *X. tropicalis*, the expression of *hotairm1* was detected in the midbrain, hindbrain and the pharyngeal arch (Figure 3B'). At the same time, *hoxa1* in this species is expressed in the pharyngeal arch and the anterior developing neural tube (Figure 3B) [43].

These results suggest that the expression domain of *hotairm1* is conserved between *B. lanceolatum* and *X. tropicalis*, being expressed in the anterior half of the developing neural tube and partly overlapping with *Hox1* and *hoxa1* expression, respectively (Figure 3).

3.4. HOTAIRM1 Function and Expression Conservation

Remarkably, *HOTAIRM1* has been described to act as a regulator of the chromatin state within the Hox cluster in human cells. Wang & Dostle [30] found that two *HOTAIRM1* isoforms, one spliced and one unspliced, play diverging roles in the regulation of the HoxA cluster chromatin state in presence of retinoic acid. Their findings indicate that the spliced isoform binds to the polycomb repressive complex 2 (PRC2) and changes the chromatin state repressing the medial HoxA genes (*HOXA4*, *HOXA5*, *HOXA6*). The unspliced isoform, on the other hand, binds to the UTX/MLL complex and promotes the expression of the proximal HoxA genes (*HOXA1* and *HOXA2*).

To gain insight into the function of *hotairm1* we tried to alter the expression balance of its isoforms during *X. tropicalis* development. In order to achieve this, we used a morpholino oligonucleotide targeting the 3' splice junction, thus forcing an isoform switch towards the unspliced state (Figure 4).

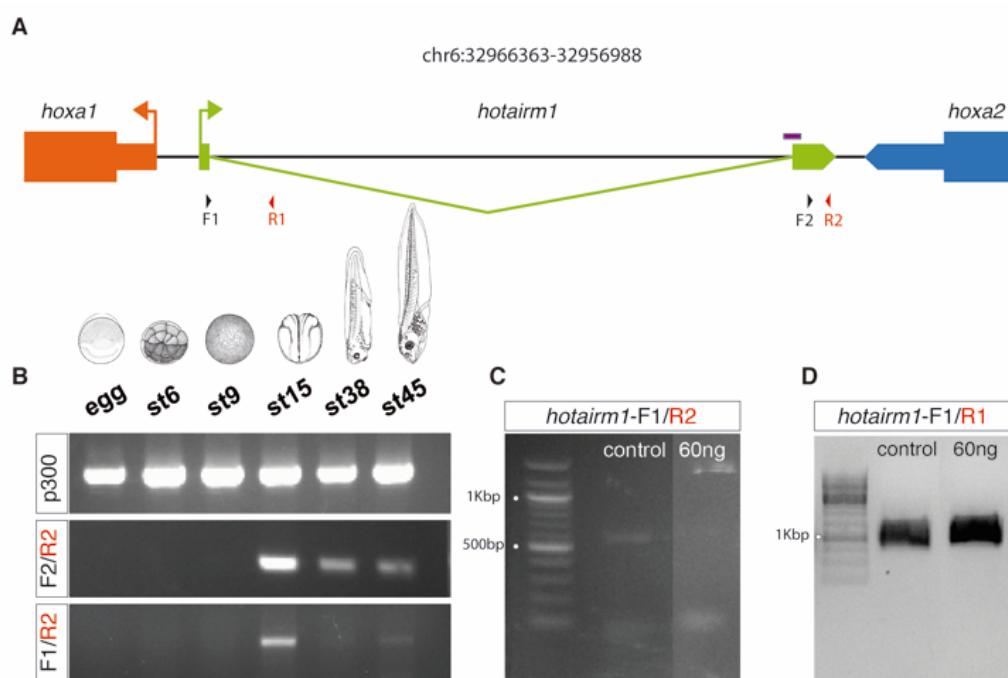


Figure 4. Isoform switch of *hotairm1* expression towards the unspliced state using a morpholino. (A) Detail of the primers (black and red arrowheads) and morpholino (purple box) used for the amplification of the spliced and unspliced isoforms of *hotairm1* in *X. tropicalis* and for the impairment of the splicing in the MO-treated embryos. (B) Expression of *hotairm1* across *Xenopus* developmental stages. (C) Inhibition of the spliced isoform in MO treated embryos of *Xenopus* at st18. (D) Assessment of the presence of the unspliced isoform of *hotairm1* in MO treated embryos as well as in the control embryos at st18.

Strikingly, the morpholino treatment resulted in a headless tadpole-stage embryo (Figure 5A). This phenotype is characterized by the decrease of expression of brain markers such as *otx2* (forebrain-midbrain boundary marker) and *engrailed* (midbrain-hindbrain boundary marker) (Figure 5B,C). These results suggest that alterations in the balance of *Hotairm1* isoforms produce a severe disruption in the development of the anterior part of the central neural system (Figure 5).

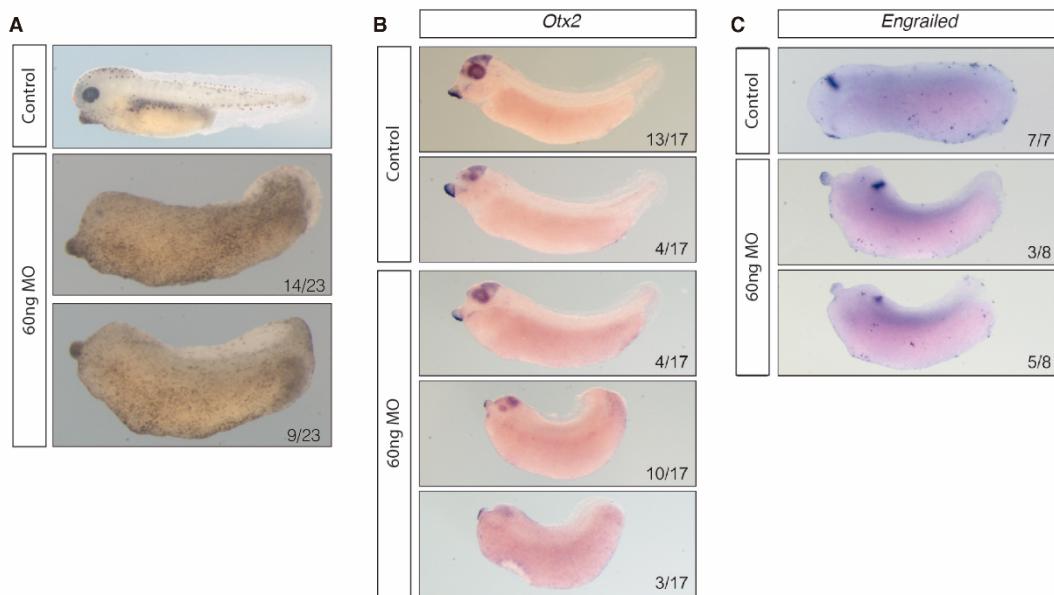


Figure 5. MO treated embryos and in situ hybridization in MO treated embryos. Anterior to the left, dorsal is up. (A) Control *X. tropicalis* MO treated embryos with normal development. 60ng hotairm1-MO treated embryos with a posteriorization of the anterior part of the embryo. (B) Whole mount colorimetric *ish* of *otx2* in *X. tropicalis* stage 26 control embryos and MO treated embryos showing the reduced expression domain of *otx2* in MO treated embryos. (C) Whole mount colorimetric *ish* of *engrailed* in *X. tropicalis* stage 26 control embryos and MO treated embryos showing a clear reduction in the expression in the MO treated embryos.

Finally, in order to check whether the expression of HoxA genes was altered in MO-treated *Xenopus* embryos, we performed Real Time quantitative PCR at stage 18, when neurulation is taking place. Our results show a significant upregulation of medial Hox genes *hoxa5* and *hoxa6*, and a downregulation of *hoxa4*, compared with control embryos. Remarkably, no significant change in the expression of the proximal HoxA gene, *hoxa1*, was observed. These results suggest that *HOTAIRM1* function is partially conserved between *Xenopus* and human (Figure 6).

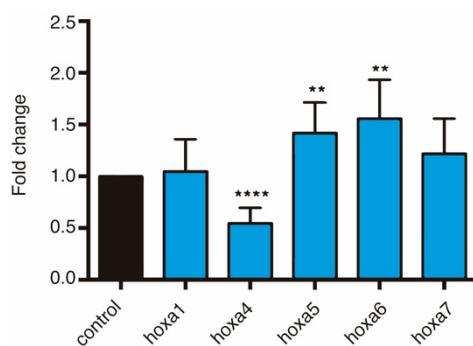


Figure 6. qPCRs between 60 ng MO treated embryos and control samples at stage 18. * shows statistically significance compared with control samples (Student's *t*-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). *Gapdh* was used as a reference gene.

4. Conclusions

We have developed a novel pipeline called lincOFinder that establishes bona fide microsyntenic clusters to detect deeply conserved lincRNAs. Applying this tool to investigate the invertebrate-vertebrate transition, we have managed to identify 16 lincRNA putatively conserved between amphioxus and humans. To our knowledge, this represents the first successful identification of

homologous lincRNAs over very long evolutionary distances. We show that one of these conserved lincRNAs, *Hotairm1*, is expressed along the anterior half of the neural tube during amphioxus and *Xenopus* development. The injection of MO targeting the 3' splice junction triggers an imbalance between the spliced and unspliced form resulting in the disruption of the proximal and medial hoxa genes. This change in hoxa expression produces in a tadpole a patterning defect in the anterior neural system leading to a headless phenotype. However, further work needs to be done to elucidate the molecular mechanism underlying this severe phenotype. This nonetheless, is a reliable indicative that this lincRNA is at least to some degree conserved in amphioxus, *Xenopus* and human, allowing us to infer that it is conserved in the phylum Chordata and that regulation of the Hox cluster by lincRNAs may be traced back to the origin of chordates.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2079-7737/8/3/61/s1>, Table S1: Putatively conserved lincRNAs between *Homo sapiens* and *Branchiostoma lanceolatum*. Analysis of the genes surrounding the lincRNA focusing on three genes (upstream2, upstream1 and lincRNA); Table S2: Putatively conserved lincRNAs between *Homo sapiens* and *Branchiostoma lanceolatum*. Analysis of the genes surrounding the lincRNA focusing on three genes (lincRNA, downstream1 and downstream2); File S1: Homology_Expression_RawOut.xlsx.

Author Contributions: Conceptualization, C.H.-U., E.N.-P. and M.M.-B.; methodology, C.H.-U. and M.M.-B.; software, C.H.-U. and J.G.; validation, C.H.-U., and M.M.-B.; formal analysis, C.H.-U. and J.G.; investigation, C.H.-U., B.A.-C. and M.M.-B.; resources, J.G.-F., G.N.W.; data curation, C.H.-U. and J.G.; writing—original draft preparation, C.H.-U., M.M.-B. and E.N.-P.; writing—review and editing, C.H.-U., M.M.-B., J.G.-F., G.N.W. and E.N.-P.; visualization, C.H.-U. and M.M.-B.; supervision, J.G.-F. and G.N.W.; project administration, J.G.-F.; funding acquisition, J.G.-F. and G.N.W.

Funding: This research was funded by grant BFU2017-86152-P (Ministerio de Ciencia, Innovación y Universidades, Spanish Government) to J.G.-F. C.H.-U. holds a predoctoral FPI contract (Ministerio de Ciencia, Innovación y Universidades, Spanish Government) and People Program (Marie Curie Actions) of the European Union's Seventh Framework Program FP7 under REA Grant agreement number 607142 (DevCom) to GW.

Acknowledgments: We wish to thank Mauel Irimia for help, advice and support, and Tom Lewin for the modifications to the HCR protocol for *ish*.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Holland, P.W.H.; Garcia-Fernández, J.; Williams, N.A.; Sidow, A. Gene duplications and the origins of vertebrate development. *Development* **1994**, *1994*, 125–133.
2. Schmitz, J.F.; Zimmer, F.; Bornberg-Bauer, E. Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Res.* **2016**, *44*, 6287–6297. [[CrossRef](#)] [[PubMed](#)]
3. Morris, K.V.; Mattick, J.S. The rise of regulatory RNA. *Nat. Rev. Genet.* **2014**, *15*, 423–437. [[CrossRef](#)] [[PubMed](#)]
4. Zampetaki, A.; Albrecht, A.; Steinhofel, K. Long Non-coding RNA Structure and Function: Is There a Link? *Front. Physiol.* **2018**, *9*, 1201. [[CrossRef](#)] [[PubMed](#)]
5. Wan, Y.; Kertesz, M.; Spitale, R.C.; Segal, E.; Chang, H.Y. Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* **2011**, *12*, 641–655. [[CrossRef](#)] [[PubMed](#)]
6. Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, *136*, 629–641. [[CrossRef](#)]
7. Fico, A.; Fiorenzano, A.; Pascale, E.; Patriarca, E.J.; Minchietti, G. Long non-coding RNA in stem cell pluripotency and lineage commitment: functions and evolutionary conservation. *Cell. Mol. Life Sci.* **2019**, *76*, 1459–1471. [[CrossRef](#)]
8. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet.* **2014**, *30*, 121–123. [[CrossRef](#)]
9. Jathar, S.; Kumar, V.; Srivastava, J.; Tripathi, V. Technological developments in lncRNA biology. In *Advances in Experimental Medicine and Biology*; Springer: Singapore, 2017; Vol. 1008, pp. 283–323.
10. Rivas, E.; Clements, J.; Eddy, S.R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **2017**, *14*, 45–48. [[CrossRef](#)]

11. Neme, R.; Tautz, D. Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to de novo gene emergence. *Elife* **2016**, *5*. [[CrossRef](#)]
12. Garcia-Fernandez, J.; Benito-Gutierrez, E. It's a long way from amphioxus: descendants of the earliest chordate. *Bioessays* **2009**, *31*, 665–675. [[CrossRef](#)]
13. Paps, J.; Holland, P.W.H.; Shimeld, S.M. A genome-wide view of transcription factor gene diversity in chordate evolution: less gene loss in amphioxus? *Brief. Funct. Genomics* **2012**, *11*, 177–186. [[CrossRef](#)]
14. Putnam, N.H.; Butts, T.; Ferrier, D.E.K.; Furlong, R.F.; Hellsten, U.; Kawashima, T.; Robinson-Rechavi, M.; Shoguchi, E.; Terry, A.; Yu, J.-K.; et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **2008**, *453*, 1064–1071. [[CrossRef](#)]
15. Marlétaz, F.; Firbas, P.N.; Maeso, I.; Tena, J.J.; Bogdanovic, O.; Perry, M.; Wyatt, C.D.R.; de la Calle-Mustienes, E.; Bertrand, S.; Burguera, D.; et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **2018**, *564*, 64–70. [[CrossRef](#)]
16. Bertrand, S.; Escriva, H.; Williams, N.A.; Holland, N.D.; Holland, L.Z. Evolutionary crossroads in developmental biology: amphioxus. *Development* **2011**, *138*, 4819–4830. [[CrossRef](#)]
17. Pegueroles, C.; Iraola-Guzmán, S.; Chorostecki, U.; Ksieziopolska, E.; Saus, E.; Gabaldón, T. Transcriptomic analyses reveal groups of co-expressed, syntenic lncRNAs in four species of the genus *Caenorhabditis*. *RNA Biol.* **2019**, *16*, 320–329. [[CrossRef](#)]
18. Bush, S.J.; Muriuki, C.; McCulloch, M.E.B.; Farquhar, I.L.; Clark, E.L.; Hume, D.A. Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genet. Sel. Evol.* **2018**, *50*, 20. [[CrossRef](#)]
19. Zerbino, D.R.; Achuthan, P.; Akanni, W.; Amode, M.R.; Barrell, D.; Bhai, J.; Billis, K.; Cummins, C.; Gall, A.; Girón, C.G.; et al. Ensembl 2018. *Nucleic Acids Res.* **2018**, *46*, D754–D761. [[CrossRef](#)]
20. Emms, D.M.; Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [[CrossRef](#)]
21. Bogdanovic, O.; Alexis, M.S.; Tena, J.J.; Maeso, I.; Fernandez-Minan, A.; Fraser, H.B.; Gomez-Skarmeta, J.L.; Roy, S.W.; Irimia, M.; de la Calle-Mustienes, E. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* **2012**, *22*, 2356–2367.
22. Sokal, R.R. A statistical method for evaluating systematic relationships. *Univ. Kansas, Sci. Bull.* **1958**, *38*, 1409–1438.
23. Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006. [[CrossRef](#)]
24. Monsoro-Burq, A.H. A Rapid Protocol for Whole-Mount In Situ Hybridization on *Xenopus* Embryos. *Cold Spring Harb. Protoc.* **2007**, *2007*. [[CrossRef](#)]
25. Fuentes, M.; Benito, E.; Bertrand, S.; Paris, M.; Mignardot, A.; Godoy, L.; Jimenez-Delgado, S.; Oliveri, D.; Candiani, S.; Hirsinger, E.; et al. Insights into spawning behavior and development of the european amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. Part B Mol. Dev. Evol.* **2007**, *308B*, 484–493. [[CrossRef](#)]
26. Choi, H.M.T.; Schwarzkopf, M.; Fornace, M.E.; Acharya, A.; Artavanis, G.; Stegmaier, J.; Cunha, A.; Pierce, N.A. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **2018**, *145*, dev165753. [[CrossRef](#)]
27. Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **2016**, *17*, 601–614. [[CrossRef](#)]
28. Yu, H.; Lindsay, J.; Feng, Z.-P.; Frankenberg, S.; Hu, Y.; Carone, D.; Shaw, G.; Pask, A.J.; O'Neill, R.; Papenfuss, A.T.; et al. Evolution of coding and non-coding genes in HOX clusters of a marsupial. *BMC Genomics* **2012**, *13*, 251. [[CrossRef](#)]
29. Gardner, P.P.; Fasold, M.; Burge, S.W.; Ninova, M.; Hertel, J.; Kehr, S.; Steeves, T.E.; Griffiths-Jones, S.; Stadler, P.F. Conservation and Losses of Non-Coding RNAs in Avian Genomes. *PLoS One* **2015**, *10*, e0121797. [[CrossRef](#)]
30. Wang, X.Q.D.; Dostie, J. Reciprocal regulation of chromatin state and architecture by HOTAIRM1 contributes to temporal collinear HOXA gene activation. *Nucleic Acids Res.* **2017**, *45*, 1091–1104. [[CrossRef](#)]
31. Zhang, X.; Lian, Z.; Padden, C.; Gerstein, M.B.; Rozowsky, J.; Snyder, M.; Gingeras, T.R.; Kapranov, P.; Weissman, S.M.; Newburger, P.E. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **2009**, *113*, 2526–2534. [[CrossRef](#)]

32. Sekigami, Y.; Kobayashi, T.; Omi, A.; Nishitsuji, K.; Ikuta, T.; Fujiyama, A.; Satoh, N.; Saiga, H. Hox gene cluster of the ascidian, *Halocynthia roretzi*, reveals multiple ancient steps of cluster disintegration during ascidian evolution. *Zool. Lett.* **2017**, *3*, 17. [[CrossRef](#)]
33. Pascual-Anaya, J.; Sato, I.; Sugahara, F.; Higuchi, S.; Paps, J.; Ren, Y.; Takagi, W.; Ruiz-Villalba, A.; Ota, K.G.; Wang, W.; et al. Hagfish and lamprey Hox genes reveal conservation of temporal colinearity in vertebrates. *Nat. Ecol. Evol.* **2018**, *2*, 859–866. [[CrossRef](#)]
34. Dehal, P.; Boore, J.L. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biol.* **2005**, *3*, e314. [[CrossRef](#)]
35. Esfandi, F.; Taheri, M.; Omrani, M.D.; Shadmehr, M.B.; Arsang-Jang, S.; Shams, R.; Ghafouri-Fard, S. Expression of long non-coding RNAs (lncRNAs) has been dysregulated in non-small cell lung cancer tissues. *BMC Cancer* **2019**, *19*, 222. [[CrossRef](#)]
36. Li, Q.; Dong, C.; Cui, J.; Wang, Y.; Hong, X. Over-expressed lncRNA HOTAIRM1 promotes tumor growth and invasion through up-regulating HOXA1 and sequestering G9a/EZH2/Dnmts away from the HOXA1 gene in glioblastoma multiforme. *J. Exp. Clin. Cancer Res.* **2018**, *37*, 265. [[CrossRef](#)]
37. Song, L.; Zhang, S.; Duan, C.; Ma, S.; Hussain, S.; Wei, L.; Chu, M. Genome-wide identification of lncRNAs as novel prognosis biomarkers of glioma. *J. Cell. Biochem.* **2019**. [[CrossRef](#)]
38. Lin, M.; Pedrosa, E.; Shah, A.; Hrabovsky, A.; Maqbool, S.; Zheng, D.; Lachman, H.M. RNA-Seq of Human Neurons Derived from iPS Cells Reveals Candidate Long Non-Coding RNAs Involved in Neurogenesis and Neuropsychiatric Disorders. *PLoS One* **2011**, *6*, e23356. [[CrossRef](#)]
39. Albuixech-Crespo, B.; López-Blanch, L.; Burguera, D.; Maeso, I.; Sánchez-Arrones, L.; Moreno-Bravo, J.A.; Somorjai, I.; Pascual-Anaya, J.; Puelles, E.; Bovolenta, P.; et al. Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain. *PLoS Biol.* **2017**, *15*, e2001573. [[CrossRef](#)]
40. Albuixech-Crespo, B.; Herrera-Úbeda, C.; Marfany, G.; Irimia, M.; García-Fernández, J. Origin and evolution of the chordate central nervous system: insights from amphioxus genoarchitecture. *Int. J. Dev. Biol.* **2017**, *61*, 655–664. [[CrossRef](#)]
41. Schubert, M.; Holland, N.D.; Laudet, V.; Holland, L.Z. A retinoic acid-Hox hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. *Dev. Biol.* **2006**, *296*, 190–202. [[CrossRef](#)]
42. Zieger, E.; Candiani, S.; Garbarino, G.; Croce, J.C.; Schubert, M. Roles of Retinoic Acid Signaling in Shaping the Neuronal Architecture of the Developing Amphioxus Nervous System. *Mol. Neurobiol.* **2018**, *55*, 5210–5229. [[CrossRef](#)]
43. McNulty, C.L.; Peres, J.N.; Bardine, N.; van den Akker, W.M.R.; Durston, A.J. Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. *Development* **2005**, *132*, 2861–2871. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).