Developmental Biology 448 (2019) 101–110

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Functional conserved non-coding elements among tunicates and chordates

Luca Ambrosino[a,1], Quirino Attilio Vassalli[b,1], Ylenia D'Agostino[b], Riccardo Esposito[c], Viviana Cetrangolo[c], Luigi Caputi[d], Alessandro Amoroso[b], Francesco Aniello[e], Salvatore D'Aniello[b], Marios Chatzigeorgiou[c], Maria Luisa Chiusano[a,f], Annamaria Locascio[b,*]

[a] Department of Research Infrastructures for Marine biological Resources, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
[b] Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
[c] Sars International Centre for Marine Molecular Biology, Thormøhlensgt 55, 5006 Bergen, Norway
[d] Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
[e] Department of Biology, University of Naples Federico II, Naples, Italy
[f] Department of Agriculture, University of Naples "Federico II", Portici, Napoli, Italy

ARTICLE INFO

Keywords:
Regulatory elements
CNE
Chordate evolution
Homeobox-containing genes

ABSTRACT

Non-coding regions with dozens to several hundred base pairs of extreme conservation have been found in all metazoan genomes. The distribution of these conserved non-coding elements (CNE) within and across genomes has suggested that many of them may have roles as transcriptional regulatory elements. A combination of bioinformatics and experimental approaches can be used to identify CNEs with regulatory activity in phylogenetically distant species. Nevertheless, the high divergent rate of genomic sequences of several organisms, such as tunicates, complicates the characterization of these conserved elements and very few examples really may prove their functional activity.

We used a comparative approach to facilitate the identification of CNEs among distantly related or highly divergent species and experimentally demonstrated the functional significance of these novel CNEs.

We first experimentally tested, in *C. robusta* and *D. rerio* transgenic embryos, the regulatory activity of conserved elements associated to genes involved in developmental control among different chordates (*Homo sapiens* and *Danio rerio* for vertebrates, *Ciona robusta* and *Ciona savignyi* for tunicates and *Branchiostoma floridae* for cephalochordates). Once demonstrated the cross-species functional conservation of these CNEs, the same gene loci were used as references to locate homologous regions and possible CNEs in available tunicate genomes.

Comparison of tunicate-specific and chordate-specific CNEs revealed absence of conservation of the regulatory elements in spite of conservation of regulatory patterns, likely due to evolutionary specification of the respective developmental networks. This result highlights the importance of an integrative in-silico/in-vivo approach to CNEs investigation, encompassing both bioinformatics, essential for putative CNEs identification, and laboratory experiments, pivotal for the understanding of CNEs functionality.

1. Introduction

Comparative genomics is a key approach to understand evolution. Comparison of genomic regions from evolutionary close or distantly related species highlighted the presence of many conserved sequences located in non-coding regions (Turner and Cox, 2014). The alignment of genomic regions of vertebrate species, such as human and fish (Abnizova et al., 2007; Parveen et al., 2013), led to the identification of conserved DNA stretches in the intergenic portion of the genome and hence defined "Conserved Non-coding Elements" or CNEs (Nelson and Wardle 2013). Most of these CNEs are supposed to be cis-regulatory elements activating or repressing transcription factors and often associated to tissue-specific gene expression (Howard and Davidson, 2004; Sandelin et al., 2004; Harmston et al., 2013). Moreover, the regulatory function of some CNEs, primarily involved in vertebrates differentiation and developmental processes, has been validated by *in vivo* experiments (Nobrega et al., 2003; Pennacchio et al., 2006; Visel et al., 2008). In addition to vertebrates, these elements have been found also in tunicates (Irvine, 2013; Vassalli et al., 2015) and plants (Kritsas et al., 2012), suggesting an ancient origin of these conserved

* Corresponding author.
E-mail address: annamaria.locascio@szn.it (A. Locascio).
[1] These authors equally contributed.

https://doi.org/10.1016/j.ydbio.2018.12.012
Received 29 May 2018; Received in revised form 10 December 2018; Accepted 11 December 2018
Available online 20 December 2018
0012-1606/

DNA regulatory fragments (Harmston et al., 2013). However, the high divergent rate of the genome sequences of several species, in particular of tunicates, complicates the characterization of these conserved elements in dedicated studies. Tunicates (or urochordates), along with cephalochordates and vertebrates, constitute the Chordata phylum. Paleontological evidences suggest that ancestral chordates appeared on earth more than 550 Myr ago in the Cambrian oceans (Shu et al., 1996; Morris, 1999). Following the publication of the new chordate phylogeny (Delsuc et al., 2006), cephalochordates are considered the most basal group of chordate, with tunicates and vertebrates being sister subphyla. Tunicates have remarkable peculiarities in their genomic organization, molecular evolution and developmental patterns. In particular, they show high genomic reorganization and accelerated evolutionary rates compared to other chordates (Berna and Alvarez-Valin, 2014; Jue et al., 2016), with *Oikopleura dioica* being by far the fastest evolving species. Extreme genomic reorganization led to genome shortening, with extreme reduction of non-coding regions and partial gene loss. Adaptive evolution is thought to have played a main role in the genomic evolution of tunicates, but the evolutionary causes of the marked peculiarities of tunicate genomes are yet to be clarified (Berna and Alvarez-Valin, 2014). The disruption of the *Hox* gene cluster in tunicates has been often used to describe their evolutionary and developmental divergence within chordates (Ikuta et al., 2010; Sekigami et al., 2017). Indeed, tunicate *Hox* genes do not cluster tightly together and events of lineage-specific gene loss have been described (Seo et al., 2004; Ikuta et al., 2010; Sekigami et al., 2017).

Since the evolution of sequencing technologies has considerably increased the number of available genomes, computational approaches for genome comparisons detected putative regulatory sequences in a growing phylogenetic range of species. The combination of bioinformatics methodologies, such as sequence conservation profiling, and *in vivo* experiments to test for gene regulatory activity has been a successful approach for the detection of functional *cis*-regulatory elements (Woolfe et al., 2005; Pennacchio et al., 2006; McEwen et al., 2009; Hemberg et al., 2012). Moreover, whilst the study of these elements in species having different evolutionary relationships may help to decipher different evolutionary splits (Lang et al., 2010; Braasch et al., 2016), it is difficult to identify lineage-specific behaviors based on sequence conservation (Yue et al., 2016). Indeed, although some CNEs seem to be conserved across different phyla (Royo et al., 2011; Clarke et al., 2012), other CNEs show a high variability even between closely-related species (Meader et al., 2010; Hiller et al., 2012). An important axiom to be considered for CNEs comparisons in genomes sharing high sequence similarity, known as the “Goldilocks” principle, states that discriminating functional CNEs from non-functional ones may be complicated (Yue et al., 2016). On the contrary, in highly divergent genomes, sequence conservation may be confined only at coding regions compromising the detection of CNEs. For these reasons, studies in vertebrates, with slowly evolving genomes, led to the identification of many functional CNEs (Woolfe et al., 2005; Ishibashi et al., 2012; Parker et al., 2014; Yousaf et al., 2015; Martinez-Morales, 2016). Conversely, only few CNEs were detected in comparative studies between fast-evolving tunicates and vertebrates (Maeso et al., 2013; Sanges et al., 2013). Further, comparative analyses of two congeneric tunicate species that split 3 million years ago allowed the identification of additional CNEs in *Ciona robusta* and *Ciona savignyi* (Doglio et al., 2013).

CNEs distribution along a genome often occurs in clusters, mainly near genes that encode for developmental regulators (Bejerano et al., 2004; Sandelin et al., 2004; Plessy et al., 2005; Woolfe et al., 2005). However, there is still no evidence for peculiar features at sequence level that would facilitate their identification (Harmston et al., 2013). Walter et al. (2005) observed that CNEs have an increased adenine and thymine (AT) content at their boundaries in contrast with a pronounced AT decrease in the sequences flanking the CNEs. However, the stability of this pattern depends on the properties of the considered genome: as an example, this AT pattern is more pronounced in CNEs found in genomes with high GC content, while it is not distinguishable in GC poorer regions in mammals (Vavouri et al., 2007). In this respect, computational approaches for the prediction of CNEs fail when heuristic methods are applied to sequence comparisons (Camacho et al., 2009; Buchfink et al., 2014; Pearson, 2016). Many works (Irvine, 2013; Marcovitz et al., 2016; Yue et al., 2016) reported the use of exact algorithms (Bray et al., 2003; Brudno et al., 2003; Kent et al., 2003) to perform pairwise sequence alignments. Such approaches, combined with programs that enable the browser-like visualization of the results of comparative sequence analysis (Frazer et al., 2004), led to in-silico definition of many CNEs also in distantly-related species (Irvine, 2013; Yue et al., 2016).

In this work, we analyzed the *cis*-regulatory activity of some predicted CNEs in *C. robusta* (formerly *C. intestinalis* type A, (Pennati et al., 2015)) and zebrafish *Danio rerio*, demonstrating their cross-species regulatory activity. We, then, used the combination of sequence comparison approaches with *in vivo* experimental validation for *de novo* identification, on the same set of genes, of functional conserved non-coding elements across tunicate subphylum.

2. Materials and methods

2.1. Animals and embryos

Adult *C. robusta* were collected from the Gulf of Naples, Italy. Gametes were recovered from the gonoducts of several animals and used for transgenic experiments. Adult zebrafish (*D. rerio*) were maintained according to standard procedure on 14 h light/10 h dark lighting cycle at 28 °C, as previously described (Westerfield, 2000). Embryos were obtained by natural spawning and were staged according to hours and days post-fertilization (hpf and dpf, respectively) and to morphological criteria (Kimmel et al., 1995).

2.2. Constructs

CNEs of *Otx*, *Hox5* and *Ipf1* from *C. robusta, Molgula occulta* and *D. rerio* were obtained as previously reported in Vassalli et al. (2015). The primers used for PCR amplification of the CNE fragments are listed in Supplementary Table S. The CNEs of *Hox5* and *Otx* of *C. robusta* and *D. rerio* were cloned in ZED vector, containing the internal control cassette [cardiac Actin promoter and red fluorescent protein (RFP)] and the enhancer detection cassette [Gateway entry site, the gata2 minimal promoter and the enhanced green fluorescent protein (EGFP) reporter gene] (Bessa et al., 2009).

2.3. Transgenesis

C. robusta transgenic larvae were obtained via electroporation and LacZ staining of transgenic embryos as previously described (D'Aniello et al., 2011). Experiments were repeated at least five times for each construct. In each experiment, 200–300 embryos were electroporated and incubated at 18 °C up to larva stage as described in Pezzotti et al. (2014). Heterologous constructs showed a lower percentage of stained embryos (from 10% to 30%) with respect to the *C. robusta* constructs (about 60%). The transgenesis in zebrafish was performed according to the procedure reported in Bessa et al. (2009). 15–20 embryos were used in each experiment of microinjection and they were performed in triplicate. About 10% of microinjected embryos showed a positive signal. A Zeiss Axio Imager M1 microscope was used for embryo image capture. Pictures were edited with Adobe Photoshop CS5 and adjustments, where applied, were applied only for clarity without affecting any essential part of the image.

2.4. Data sets

Genomic data in the form of scaffolds for *C. robusta* (GCF_000224145.2), *Ciona savignyi* (GCA_000149265.1), *Oikopleura dioica* (GCA_000209535.1), *Botryllus schlosseri* (GCA_000444245.1) and *Salpa thompsoni* (GCA_001749815.1) were downloaded from the Genome partition (https://www.ncbi.nlm.nih.gov/genome/) at the Database resources of the National Center for Biotechnology Information (NCBI, 2018). Genomic scaffolds for *M. occidentalis* (Moxi: genome_v1), *M. occulta* (Mocc_genome_v1) and *M. oculata* (Mocu_genome_v12) were downloaded from the Aniseed database (https://www.aniseed.cnrs.fr/aniseed/species/) (Brozovic et al., 2018). Transcript sequences of *Hox5*, *Otx* and *Ipf1* genes were retrieved from the Ensembl Genome Browser (https://www.ensembl.org/index.html) (Zerbino et al., 2018) for the following species: *H. sapiens* and *D. rerio* (Vertebrata), *C. robusta* and *C. savignyi* (Tunicata).

2.5. Gene loci assignment in tunicate genomes

Sequence similarity searches between the transcript sequences of *Hox5*, *Otx* and *Ipf1* from *H. sapiens*, *D. rerio*, *C. robusta* and *C. savignyi* and the genomic scaffold sequences from *C. robusta*, *C. savignyi*, *M. occidentalis*, *M. occulta*, *M. oculata*, *O. dioica* and *B. schlosseri*, were performed using tBLASTx (Camacho et al., 2009) and Transcriptologs (Ambrosino and Chiusano, 2017). All the similarity searches were carried out setting an expect-value (e-value) cut-off at e^{-3}. Best similarities were detected with *C. robusta*. The resulting coordinates of these similarities are listed in Supplementary file S1 and graphically represented in Supplementary Fig. 2.

2.6. CNE prediction

Tunicate genomic regions corresponding to the *Hox5*, *Otx1* and *Ipf1* loci plus 5 kb at its 5' and 3' ends were aligned using the AVID software (Bray, 2003). Sequence comparisons were performed using mVISTA (Frazer et al., 2004). All the comparisons were performed setting a 50% identity, with a sliding windows of 20 bp and a minimum length of the conserved region of 10 bp. The coordinates of the corresponding CNEs are listed in Supplementary file S1.

Key Resources Table

Deposited Data		
C. robusta reference genome	NCBI (https://www.ncbi.nlm.nih.gov/genome/)	GCF_000224145.2
C. savignyi reference genome	NCBI (https://www.ncbi.nlm.nih.gov/genome/)	GCA_000149265.1
O. dioica reference genome	NCBI (https://www.ncbi.nlm.nih.gov/genome/)	GCA_000209535.1
B. schlosseri reference genome	NCBI (https://www.ncbi.nlm.nih.gov/genome/)	GCA_000444245.1
S. thompsoni reference genome	NCBI (https://www.ncbi.nlm.nih.gov/genome/)	GCA_001749815.1
M. occidentalis reference genome	Aniseed (https://www.aniseed.cnrs.fr/aniseed/species/)	Moxi:genome_v1
M. occidentalis reference genome	Aniseed (https://www.aniseed.cnrs.fr/aniseed/species/)	Mocc_genome_v1
M. occulta reference genome	Aniseed (https://www.aniseed.cnrs.fr/aniseed/species/)	Mocu_genome_v12
Software and Algorithms		
tBLASTx	Camacho et al. (2009)	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+ /LATEST/
Transcriptologs	Ambrosino et al. 2017	https://github.com/LucaAmbrosino/Transcriptologs
AVID	Bray (2003)	http://genome.lbl.gov/vista/mvista/download.shtml
mVISTA	Frazer et al. (2004)	http://genome.lbl.gov/vista/mvista/download.shtml
Other		

3. Results

3.1. Cross-species regulatory activity of the Hox5 and Otx chordate-specific CNE

We used *Ciona robusta* embryos to test the *cis*-regulatory activity of the chordate *Hox5* and *Otx* CNEs (Table 1) identified by Vassalli et al. (2015). The authors identified an Hox5 and two Otx functional elements conserved among various chordate species (Supplementary Fig. 1). We, here, analyzed in *C. robusta* transgenic embryos the ability of these *Hox5* and *Otx* CNE-A and CNE-B, identified in *B. floridae*, *D. rerio* and *H. sapiens*, to specifically recapitulate the expression profiles of the endogenous genes The amphioxus and human *Hox5* and *Otx* CNEs, as also the *D. rerio Otx* CNE-A, only showed a non-specific signal in the mesenchyme of the transgenic larvae (data not shown). Interestingly, we obtained positive results with the *D. rerio Hoxa5a* and *Otx* CNE-B (DrHoxa5a CNE and DrOtx CNE-B) conserved sequences. The DrHoxa5a CNE drove LacZ staining in the anterior part of the tail neural tube (55% of stained larvae), in a region with well defined anterior and posterior limits and almost corresponding to that of the endogenous *Hox5* gene and recapitulating the activity of the *Hox5* CNE from *C. robusta* (CrHox5 CNE) (Fig. 1A, B and Table 1). The DrHoxa5a CNE also showed a signal around the ocellus in the

Table 1
Chordate-specific CNEs that are functionally active in *C. robusta*.

CNE	Position	Distance from the gene	Expression in *Ciona robusta*
CrHox5 CNE	Sc HT001177.1:5433-5492 (+)	967 bp	Anterior part of caudal neural tube
CrOtx CNE-A	Chr4:4314496-4314552 (-)	332 bp	Sensory vesicle
CrOtx CNE-B	Chr4:4315070-4315102 (-)	906 bp	Sensory vesicle and tail tip
CrIpf1 CNE	Chr14:3532496-3532520 (+)	368 bp	Tail muscles
DrHoxa5a CNE	Chr19:19221951-19222010 (+)	3312 bp	Anterior part of caudal neural tube
DrOtx CNE-B	Chr15:46865799-46865841 (+)	1352 bp	Sensory vesicle

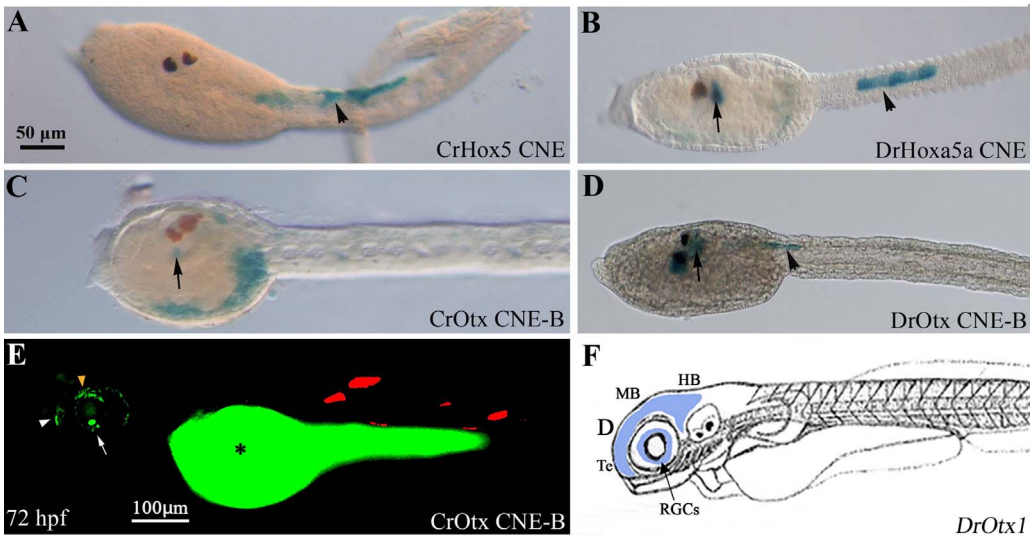

Fig. 1. Cross-species regulatory activity of the *C. robusta* and *D. rerio Hox5* and *Otx* CNEs. (A) Representation of the *C. robusta Hox5* CNE expression in the most anterior part of the tail neural tube at larva stage. (B) The *D. rerio Hoxa5a* CNE shows well defined expression profile in the anterior part of the *Ciona* neural tube in a pattern reminiscent that of the *Ciona Hox5* CNE (black arrowheads). (C) Representation of the *C. robusta Otx* CNE-B expression in the sensory vesicle(black arrow). (D) The *D. rerio Otx* CNE-B shows expression in the sensory vesicle (black arrow) and extends posteriorly along the visceral ganglion (black arrowhead). (E) *C. robusta Otx* CNE-B is active in a *D. rerio* transgenic embryo at 72 hpf. GFP signal (green) is visible in Te (white arrowhead), RGC (white arrow) and MB (orange arrowhead). As control, RFP expression (red) in muscles under the control of cardiac actin promoter. Diffuse GFP signal in the yolk is unspecific (asterisk). (F) Schematic representation of the *D. rerio Otx1* gene endogenous expression at 72 hpf. *Otx1* is expressed in RGCs, Te, D, MB and HB. Abbreviations: D, diencephalon; HB, hindbrain; MB, midbrain; RGC, retinal ganglion cells; Te, telencephalon. A-E *Ciona* embryos at larva stage. B, dorsal view; A,C-F, lateral view, anterior is on the left. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

sensory vesicle (50% of stained larvae), in a pattern partly reproducing that of the endogenous *Hoxa5a* gene in the *D. rerio* anterior CNS (Thisse and Thisse, 2005). The regulatory activity of the DrOtx CNE-B was detected in the sensory vesicle in a wider area with respect to the *C. robusta Otx* CNE-B expression (Fig. 1C, D). The signal is also visible in the axonal extension that elongates from the sensory vesicle to the anterior part of the neural tube (Fig. 1D). Comparing DrOtx CNE-B expression profile (Fig. 1D) with that of the *D. rerio Otx1* gene (Thisse et al., 2001) (Fig. 1F), this CNE activates *LacZ* reporter gene in Ciona in a small subset of the zebrafish *Otx1* mRNA endogenous territories. These *C. robusta* and *D. rerio Hox5* and *Otx* CNEs were then tested in zebrafish in order to further investigate their cross-species activity in vertebrate embryos. We assayed reporter gene activity in Tol2 transgenic embryos at "bona fide" comparable stages of development as 24, 48 and 72 hpf. We obtained positive results with the *Ciona Otx* CNE-B (CrOtx CNE-B) in zebrafish F0 transgenic embryos (Fig. 1E), while the *D. rerio Otx* and *Hox5* CNEs and the *C. robusta Hox5* CNE did not show any positive signal (data not shown). We observed the appearance of a faint signal in the eye of zebrafish embryos at 24hpf and in the anterior CNS at 48hpf (data not shown). The same signals were visible and better defined at 72 hpf. In particular, the reporter GFP signal of the CrOtx CNE-B was detected in the ganglion cell layer and in the telencephalon (Fig. 1E) in 70% of the injected embryos (n = 150), partly reproducing zebrafish *Otx1* endogenous expression (Fig. 1F) and, thus, suggesting a *cis*-regulatory conservation of the *C. robusta* CNE in vertebrates. The RFP signal in muscles under the control of cardiac actin promoter was used as internal control (Fig. 1E).

3.2. Hox5, Otx and Ipf1 genes identification in tunicates

We investigated if the same computational strategy adopted to identify functional conserved elements among chordates could be useful when considering genomes with a high divergence rate such as those of tunicates. For this analysis, we selected the three homeobox genes for which CNEs gave positive results in *C. robusta*, i.e. *Hox5*, *Otx* and *Ipf1*, (Table 1) and the genome sequences of 8 species belonging to all three tunicate classes. In particular, we used the genomes of one Thaliacean (*S. thompsoni*), 6 Ascidiaceans (three Molgulidae species, *M. occidentalis*, *M. occulta* and *M. oculata*; one Styelidae, *B. schlosseri;* two Cionidae, *C. robusta* and *C. savignyi*), and one Appendicularian (*O. dioic*a). With the exception of *C. robusta* genome, which is already assembled at chromosome level, and of *C. savignyi*, for which a reliable gene annotation is currently available, the sequenced genomes of the other tunicates are still in the form of genomic scaffolds, with no official gene annotation available for the scientific community. Therefore, in order to infer the presence of CNEs localized close to the *Hox5*, *Otx* and *Ipf1* homeobox genes in these tunicate genomes, we predicted the homologous regions of the three homeobox genes in each of the tunicate genomes using the transcript sequences from reference gene annotations (i.e. *H. sapiens*, *D. rerio*, *C. robusta* and *C. savignyi*) (see "Data sets" section in Material and Methods). *Hox5*, *Otx* and *Ipf1* similarities between tunicate genomes and *C. robusta* transcripts (Supplementary Fig. 2; Supplementary file S1) highlighted homologous regions in all tunicates with the exception of *S. thompsoni*. Unfortunately, the scaffold sequences of the genome of *S. thompsoni* are still too preliminary and not well assembled and we had to eliminate this species from subsequent analyses. Furthermore, this comparative analysis did not permit to recognize a clear *Ipf1* gene region in *O. dioica* and *B. schlosseri*. Conversely, we identified *Hox5* and *Otx* homologous regions in all tunicate species considered, thus revealing a lower degree of conservation of *Ipf1* with respect to *Hox5* and *Otx* genes in tunicate evolution. When measuring the similarity of the homologous regions from the various tunicates, considering only the aligned portion of sequences, it resulted that *C. robusta* and *C. savignyi*, as expected, share longer regions compared to the other tunicates (Supplementary Fig. 2, Supplementary file S1). Regarding *M. occidentalis*, *M. occulta*, *M. oculata*, *O. dioica* and *B. schlosseri*, instead, even if they show high level of similarity (expressed in terms of "positive" percentages) when compared with *C. robusta* transcripts (in some cases up to 90%), the similarity corresponds to shorter regions of alignment (Supplementary Fig. 2, Supplementary file S1). Moreover, if we consider exclusively the homologous regions shared by the *C. robusta* transcripts and the congeneric *Molgula* species, it is evident their similarities in terms of shared aligned portions (Supplementary Fig. 2). Details about the predicted homologous regions on the genomic scaffolds, as well as identity and positive percentages, score and the e-value of the aligned sequences, are summarized in Supplementary file S1.

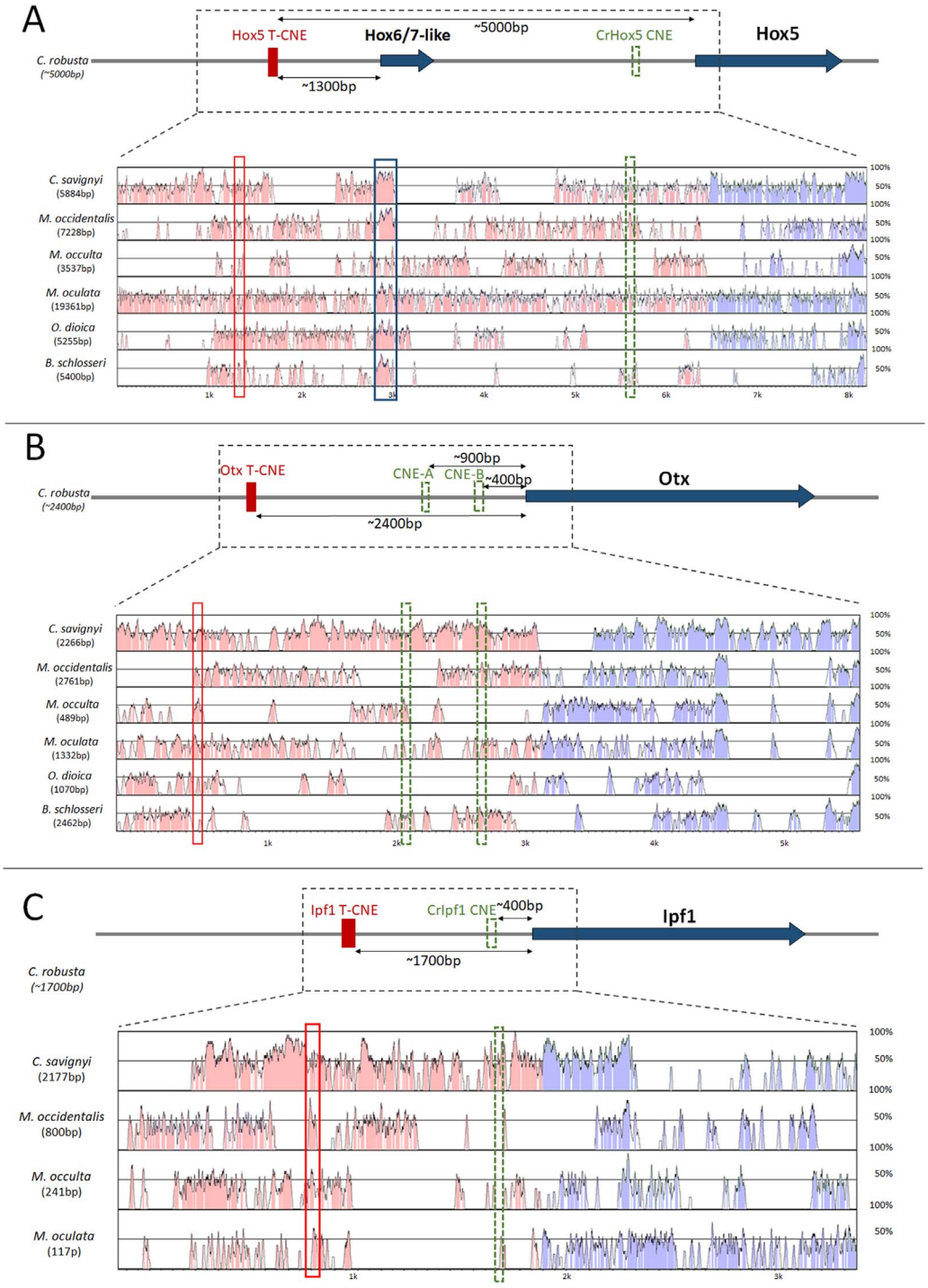

Fig. 2. Vista plots of *Hox5* (A), *Otx* (B) and *Ipf1* (C) regions across tunicates. The genomic sequences sharing at least 50% of sequence identity compared with *C. robusta* are shown for *C. savignyi*, *M. occidentalis*, *M. occulta*, *M. oculata*, *O. dioica* and *B. schlosseri*. Regions aligned to *Hox5, Otx* and *Ipf1* transcripts are shown in purple, while non-coding regions are shown in red. For each CNE sequence, the distance in base pair from the start of *C. robusta Hox5, Otx* and *Ipf1* transcripts are indicated. Regions aligned to *Hox6/7-like* transcript are shown in a blue rectangle. The predicted tunicate-specific CNEs are shown in red rectangles, while the chordate-specific CNEs are shown in green rectangles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

3.3. The Hox5, Otx and Ipf1 chordate-specific CNE in tunicates

Alignments of the predicted *Hox5*, *Otx* and *Ipf1* gene loci in tunicates were investigated to search for chordate-specific CNEs, including regions 5 kb upstream and downstream of the 5' and 3' ends of the corresponding homologous regions (Fig. 2).

The chordate-specific *Hox5* CNE is located less than 1 kb upstream of the *C. robusta Hox5* gene (Table 1; Fig. 2A). Here, the comparison of the 7 tunicate *Hox5* homologs highlighted the absence of this CNE in *M. occulta* and *O. dioica*.

The comparison of the tunicate *Otx* homologous regions again evidenced the lack of conservation of the two chordate-specific *Otx* CNEs predicted at 900 and 400 bp upstream of the *C. robusta Otx* gene (CrOtx CNE-A and CNE-B) (Fig. 2B). *O.dioica* did not retain any of the two chordate-specific CNEs, *M. occulta* did not retain one of the two conserved regions (CNE-B), while *M. occidentalis* did not retain the other (CNE-A).

The chordate-specific *Ipf1* CNE (Table 1) was located 400 bp

upstream of the *C. robusta Ipf1* gene (CrIpf1 CNE; Fig. 2C). By comparing the *Ipf1* homologous regions of *C. robusta*, *C. savignyi*, *M. occidentalis*, *M. occulta*, and *M. oculata*, it turns out that all the *Molgula* species did not retain the chordate-specific *Ipf1* CNE.

Given the lack of conservation among tunicates of all the chordate-specific *Hox5*, *Otx* and *Ipf1* CNEs, we then searched the same tunicate genomes for the presence of the 183 vertebrate-specific CNEs identified in a previous work by Sanges et al. (2013). Again, none of these CNEs was conserved among all the analyzed tunicate species, thus confirming the unusual diversification rate of the tunicate genomes with respect to other even more phylogenetically distant chordate species.

3.4. Tunicate-specific CNE prediction

Once established that none of the chordate-specific CNEs, was conserved among all the analyzed tunicate species, we then, performed a second comparative analysis looking for tunicate-specific CNEs in the *Hox5*, *Otx* and *Ipf1* gene loci. By this approach, we classified as tunicate-specific CNEs (T-CNEs) only the sequences simultaneously aligned in all the considered species. This survey led to the identification of three putative T-CNEs conserved in all the tunicate species, flanking each of the analyzed genes (Fig. 3) and different from the previously identified chordate-specific ones.

3.4.1. Hox5 T-CNE

A 21 bp T-CNE (Hox5 T-CNE) was identified about 5 kb upstream of the *C. robusta Hox5* gene (Figs. 3A and 2A). This element is present in all tunicate species, with a sequence identity percentage ranging from 33% to 62% (Figs. 3A and 2A).

Moreover, the multiple alignment that enabled to identify the tunicate-specific Hox5 CNE also showed a conserved sequence of about 300 bp, located in the proximity (~3 kb upstream) of the *Hox5 C. Robusta* gene. A dedicated analysis in this region revealed the presence of a relatively short *Hox6/7-like* gene in the ENSEMBL annotation (ENSCING00000023524). This *Hox6/7-like* gene, confirmed in ENSEMBL by EST evidences, is also present in the RefSeq annotation, while is not described in the ANISEED database gene annotation. This region seems to be present in all the considered species with the exception of *M. occulta* (Fig. 2A).

3.4.2. Otx T-CNE

A 34 bp T-CNE was identified 2,4 kb upstream of the *C. robusta Otx* gene (CrOtx T-CNE). This element is present in all the considered tunicate species, with a sequence identity percentage ranging from 29% to 53% (Figs. 3B and 2B).

3.4.3. Ipf1 T-CNE

It was not possible to detect a *Ipf1* gene locus in *O. dioica* and *B. schlosseri*, we thus removed from the multiple alignment these two more divergent species (Fig. 2C). By reducing the number of tunicate species, a 23 bp T-CNE was identified 1,7 kb upstream of the *C. robusta Ipf1* gene (CrIpf1 T-CNE). This element shows a sequence identity percentage ranging from 41% to 70% (Figs. 3C and 2C).

3.5. Activity of the T-CNE in Ciona transgenic embryos

The computational analysis among tunicate genomes revealed three putative new T-CNEs conserved in most of the considered tunicate species (Fig. 3). We tested by electroporation in *C. robusta* embryos, the activity of these T-CNEs. Hence, for this in vivo analysis we chose species representative of the class of Appendicularian (*O. dioica*) and of each family of Ascidiacea (*C. robusta* for Cionidae, *B. schlosseri* for Styelidae and *M. occulta* for Molgulidae). For *Hox5* T-CNEs, we obtained positive results only with the MoccHox5 T-CNE. In particular, MoccHox5 T-CNE drove LacZ staining in the most anterior part of the tail neural tube in more than 65% of stained larvae (Fig. 4A). Remarkably, MoccHox5 T-CNE expression is very similar to that obtained with the chordate-specific CrHox5 CNE and DrHox5 CNE (Fig. 4B,C), despite the absence of a clear sequence similarity between the mentioned CNEs. The other newly identified Hox5 T-CNEs, from *Ciona*, *Botryllus* and *Oikopleura* (CrHox5 T-CNE, BsHox5 T-CNE and OdHox5 T-CNE) did not give any consistent signal except for a non-specific staining in the mesenchyme (Supplementary Fig. 3). The in vivo analysis of the *Otx* T-CNEs evidenced different degree of conservation in their regulatory activity. The MoccOtx T-CNE, activates *LacZ* reporter expression in the ventral part of the sensory vesicle in about 50% of stained larvae (Fig. 4D). Interestingly, also in this case, the expression is very similar to that obtained with the chordate-specific CrOtx CNE-A and CNE-B (Fig. 4E,F), while the CrOtx T-CNE displayed only an ectopic signal in mesenchymal cells (Supplementary Fig. 3). The OdOtx T-CNE showed in 32% of stained larvae a specific and intriguing expression profile in the sensory vesicle and in neuronal axonal extensions of the visceral ganglion up to the anterior neural tube (Fig. 4G). This pattern perfectly resume that observed with the chordate-specific *D. rerio Otx* CNE-B (Fig. 4H). Again, as for the MoccHox5 T-CNE, this T-CNE shows a functional activity conserved with a chordate-specific CNE but not with the other tunicate-specific *Otx* T-CNEs.

Although no sequence similarity was detected, We, then, interrogated Jaspar, a transcription factor binding profiles database (Khan et al., 2018), in order to search for transcription factors possibly responsible for the regulatory activity of these CNEs and T-CNEs. This analysis identified a binding site (TAATTC/G) for the paired-related HD factor (PRRX1) common to DrOtx CNE-B and OdOtx T-CNE. These two CNEs are active in neural axons of the visceral ganglion and this factor is involved in cortical neural progenitor's differentiation in vertebrates (Li et al., 2017). Interestingly, a binding site for another paired-related HD factor of the same family (PRRX2) was common to the CrOtx CNE-A and MoccOtx T-CNE, both expressed in the sensory vesicle. Further studies will be necessary to establish their involvement in the control of the transcriptional activity of these CNE/T-CNE.

For *Ipf1* T-CNEs, because of the lack of defined *Ipf1* gene loci in *B. schlosseri* and *O. dioica*, we tested the activity of the *M. occulta* and *C. robusta* elements, but did not observe any specific activity (Supplementary Fig. 3).

4. Discussion

Tunicates show a high evolutionary plasticity and are key marine organisms to study morphological ancestral patterns of chordates evolution. In spite the largely divergent genomes, tunicates often display shared morphological features during development, a paradox, named 'Developmental system drift'. This happens when unaltered physical features between different species are controlled by altered developmental pathway not based on sequence conservation (Stolfi et al., 2014). On the other hand, regulatory elements have been found to be conserved in tunicates and between tunicates and vertebrates (Sanges et al., 2013; Vassalli et al., 2015), making even more confused our comprehension of the evo-devo dynamics in tunicates.

The *in vivo* cross-species functional conservation of the CNEs identified by bioinformatics analyses is fundamental to demonstrate the real usefulness of the computational tools. In this study, our aim was to exploit appropriate bioinformatics for further investigations about the evolution of transcriptional regulatory mechanisms among highly divergent chordates. We performed a cross-species analysis centered on the conserved elements identified by multiple sequences comparison among chordates of the *Hox5*, *Otx* and *Ipf1* gene loci (Vassalli et al., 2015). The transgenic experiments performed on both *C. robusta* and *D. rerio* embryos using the *Hox5* and *Otx* CNEs led to the identification of chordate-specific CNEs functionally active in *C. robusta* and in particular one, namely the CrOtx CNE-B, active in both *Ciona* and *D. rerio*. Our results evidenced the evolutionary conserva-

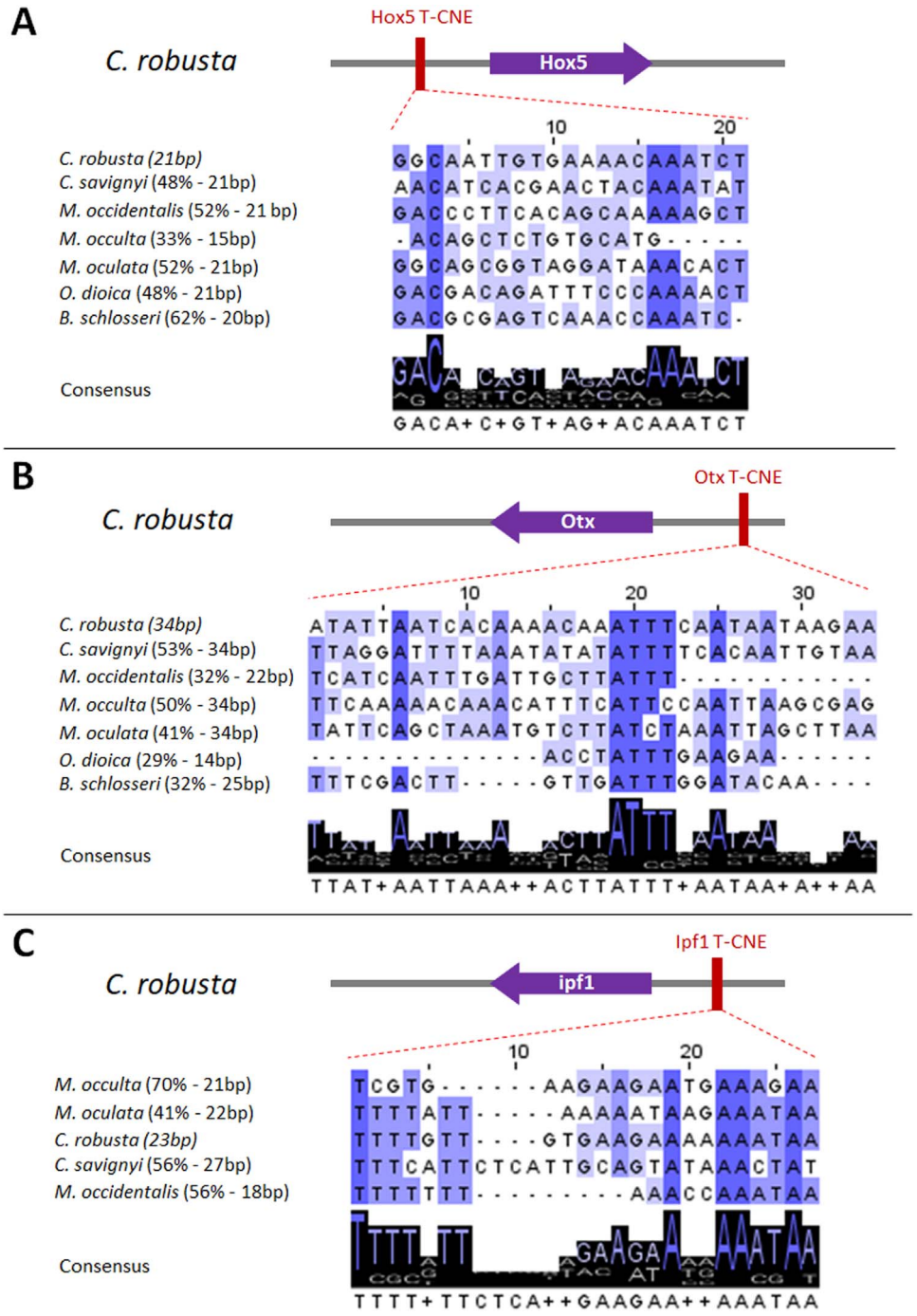

Fig. 3. Multiple alignments of *Hox5* (A), *Otx* (B) and *Ipf1* (C) CNEs. In the upper part of each multiple alignment, the position of each CNE and the related *C. robusta* homeobox gene are indicated. Based on the nucleotide conservation along the considered species, aligned nucleotides are highlighted with different intensities of blue. In the lower part of each multiple alignment, a nucleotide consensus is shown. For each CNE sequence, the length in base pair and the percentage of identity with the *C. robusta* CNE sequence are indicated.

tion of small non-coding regulatory elements that partly reproduced the expression profile of the endogenous gene. We obtained positive cross-species results for genes that maintained common expression profiles during evolution, such as *Otx1*, which shows endogenous expression in the anterior central nervous system in both *C. robusta* and *D. rerio* chordate species. On the contrary, in case of divergence and diversification of the expression territories we did not observe any functional activity in territories specific for a single species (Vassalli et al., 2015). This is the case of the *Ipf1* gene, also known as pancreatic and duodenal homeobox 1 (*Pdx1*), whose expression in the *Ciona* nervous system (Corrado et al., 2001) is not conserved in other chordates, where it is necessary for pancreatic development (Jensen, 2004).

It is interesting to note that only CNEs from *D. rerio*, a teleost species, have shown positive results. Indeed, tunicates and vertebrates, but not cephalochordates, are evolutionary strictly related, being both included in the olfactores supergroup (Delsuc et al., 2006) and conserved regulatory non-coding regions were previously identified between tunicates and vertebrates (Sanges et al., 2013).

Regarding the conservation between evolutionarily distant species, it is important to highlight the different efficiency in terms of functionality of these CNEs between *Ciona* and zebrafish transgenic embryos. We observed more CNEs being active in *Ciona* with respect to zebrafish and, even more peculiar, only the *Ciona Otx* CNE-B and not

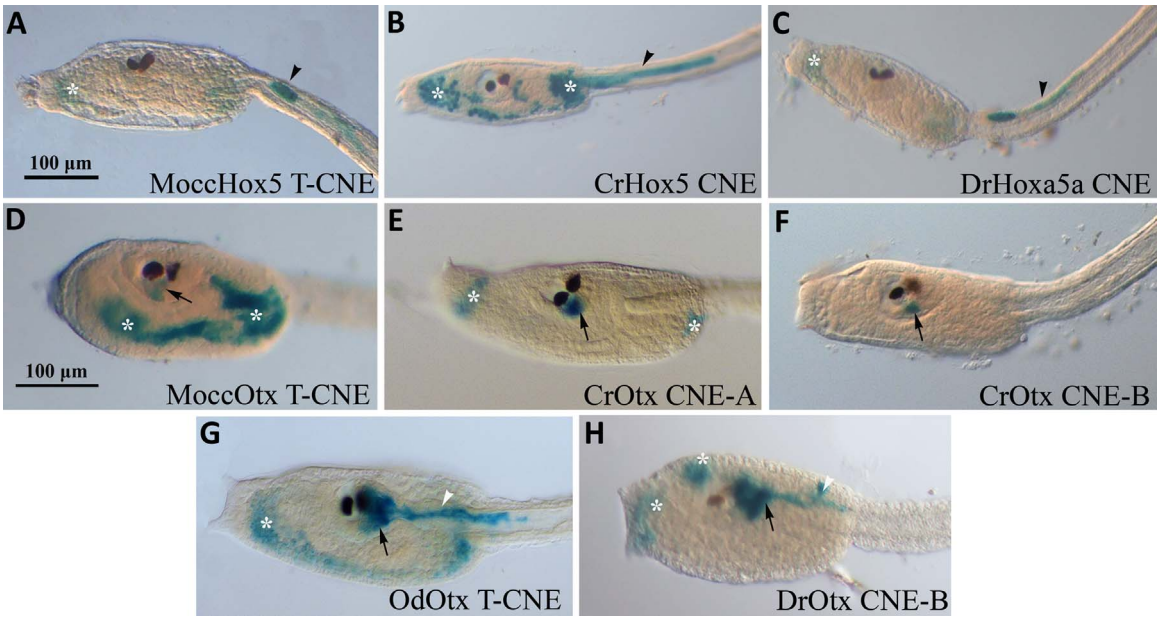

Fig. 4. Comparison of the regulatory activity of tunicate-specific and chordate-specific *Hox5* and *Otx* T-CNE/CNE. (A-C) Comparable expression of *M. occulta Hox5* T-CNE, *C. robusta* chordate-specific *Hox5* CNE and *D. rerio Hoxa5a* CNE in the most anterior part of the tail neural tube (black arrowheads). MoccHox5 T-CNE shows a more restricted expression in a subdomain of CrHox5 and DrHox5CNEs at larva stage. (D-F) The tunicate-specific *M. occulta Otx* T-CNE (D), and the chordate-specific *C. robusta Otx* CNE-A (E) and CNE-B (F) show a similar expression profile in the ventral part of the sensory vesicle (black arrows). (G,H) The tunicate-specific OdOtx T-CNE reproduces almost the same expression profile of the chordate-specific DrOtx CNE-B in the posterior sensory vesicle and in the neural axons along the visceral ganglion. Black/white arrowheads and arrows indicate the corresponding conserved territories of expression. Asterisks indicate ectopic expression in the mesenchyme. A-F,H: lateral view; G, dorsal view. Anterior is on the left.

the *D. rerio Otx* CNE-B was active in zebrafish transgenes. This most probably indicates an increase in the complexity of regulation in vertebrates (Olson, 2006) which cannot be represented by a single regulatory element. Presumably, single conserved elements are recognized and show efficient functionality in more basal organisms, while they may need more complex regulatory assemblages to become active in vertebrates. Moreover, *D. rerio*, as all teleosts, underwent an additional round of genome duplication, which led to an increased number of paralogue genes and to their consequential sub-functionalization (Sobral et al., 2009). This may partially explain the reduced efficiency of the CNEs in *D. rerio*.

Once demonstrated the cis-regulatory conservation between tunicates and vertebrates of these so-called chordate-specific CNEs, the same bioinformatics approach was applied towards the identification of tunicate-specific CNEs (T-CNE) for the same *Hox5*, *Otx* and *Ipf1* genes. We have compensated for the lack of a complete genome assembling and gene annotation for most of the tunicate genomes by sequence similarity analyses from better annotated chordate genomes (Supplementary Fig. 2). We, then, compared the regulatory activity of the previously detected chordate-specific CNEs (Vassalli et al., 2015) and of the tunicate-specific ones(Fig. 4).

None of the identified Cr T-CNEs has been found to be functionally active in *C. robusta*, while the corresponding T-CNEs from other tunicate species were found to be active (e.g. MoccOtx T-CNE and OdOtx T-CNE). One may speculate that, in *C. robusta*, the identified CNEs contain repressor elements or need to cooperate with other elements to be functionally active. The first hypothesis seems to be questionable for the activity of most of these elements in the mesenchyme. As alternative hypothesis, it is possible that they require more than one element for the proper regulatory activity. As these additional elements have not been identified as conserved elements, it is not possible at this stage to demonstrate this hypothesis.

The Hox5 T-CNE, shared by all tunicate genomes under investigation, is located quite distant (5 kb upstream) from the transcriptional start site of the *C. robusta Hox5* gene (Fig. 2A). It should be noted that the small *Hox6/7*-like gene is located ~3 kb upstream of the *Hox5* gene, suggesting that the identified CNE might play a regulatory role for *Hox6/7*-like rather than for *Hox5*. The multiple alignment provided in this work, moreover, highlighted that the *Hox6/7*-like gene is conserved in all the considered tunicates, with the exception of *M. occulta* (Fig. 2A). Interestingly, *in vivo* experiments revealed that only the *M. occulta Hox5* T-CNE was specifically active while the corresponding *C. robusta*, *B. schlosseri* and *O. dioica* were not functional. The presence of the *Hox6/7*-like gene probably interfered with this enhancer activity and led to its inactivation. A possible phenomenon of exaptation for the *M. occulta* CNE could be the consequence of specific *Hox6/7*-like gene loss in this species.

Comparing the *Otx* T-CNEs tested *in vivo*, we observed functional conservation of the *M. occulta* T-CNE with that of the two *Ciona* chordate-specific CNEs (CrOtx CNE-A and CrOtx CNE-B). Furthermore, we obtained a very similar regulatory activity between the OdOtx T-CNE and DrOtx CNE-B. It is interesting to note that in both cases we observed a functional conservation between tunicate and a chordate CNEs which came out from two separate sequence analyses and were not related by sequence similarity.

These results confirm a remarkable degree of divergence among tunicate species but, at the same time, once more assess the usefulness of interspecies experiments to identify functional CNEs in tunicates.

It is important to note that the *Ipf1* T-CNEs, which were both not active by *in vivo* assays, were identified without taking into account *Oikopleura* and *Botryllus* species. The definition of the homologous regions in these species, in fact, was limited to a restricted and inadequate portion of the *Ciona* transcript. Although a single gene example is not sufficient to statistically demonstrate its significance, this evidence seems to confirm the variability of the tunicate genomes and suggest the need of more complete tunicate genome sequences to better investigate the patterning of developmental genes. Comparative analyses among tunicates need to include more complete genome regions and more distantly related species in order to have a more representative analysis.

Investigating the evolutionary and genomic features of the tunicate-specific CNEs herein described is beyond the scope of the present work. However, the dichotomy between tunicate-specific and chordate-specific CNEs that, in a first analysis, came out from the sequence

comparisons, resulted only apparent after the *in vivo* functional activity of these CNEs and T-CNEs. Despite none of the sequences of the *Hox5* and *Otx* chordate-specific CNEs was conserved in all tunicate genomes, these elements showed in some cases a regulatory activity very similar to that of completely different sets of CNEs identified for their conservation among tunicates. This similarity cannot be detected by a simple search for conserved sequences among different species but requires the comparison of different groups of species followed by their in vivo functional analysis and finally by the search for conserved short sequence elements containing specific binding sites.

Both chordate-specific and tunicate-specific CNEs seem to represent key regulators of complex developmental processes. They underwent different processes of selection and during evolution became fine-tuning, regulators of group-specific or species-specific processes (Pennacchio et al., 2006). According to this, tunicate-specific CNEs would not be evolutionary more recent than chordate-specific CNEs (Villar et al., 2015) but, more simply, went through positive regulatory adaptation and selection at sequence level, (Villar et al., 2015).

Acknowledgments

We are indebted with Josè Luis Gomez Skarmeta for giving us Tol2 constructs and with Cristian Canestro, Loriano Ballarin and Filomena Ristoratore for Oikopleura, Botryllus and Molgula occulta genomic DNAs. We are grateful to Paolo Sordino for critical reading of the manuscript. Special thanks to Catia Pedalino, Claudia La Vecchia, Eleonora Gagliardi, Federica Berruto and Federica Salatiello for their help in constructs preparation and transgenesis experiments. We thank the Molecular Biology and Bioinformatics Unit of the Stazione Zoologica Anton Dohrn, Italy for technical support. Many thanks also to Alberto Macina, the MaRe and Meda Units of the Stazione Zoologica Anton Dohrn for animals fishing and husbandry. Contract grant sponsor: QAV and YDA were supported by SZN-Open University (UK) Ph.D. Fellowships. Research in MC's lab is funded by the Sars Centre, Norway core budget.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2018.12.012.

References

Abnizova, I., Walter, K., Te Boekhorst, R., Elgar, G., Gilks, W.R., 2007. Statistical information characterization of conserved non-coding elements in vertebrates. J. Bioinform. Comput. Biol. 5 (2b), 533–547.

Ambrosino, L., Chiusano, M.L., 2017. Transcriptologs: a transcriptome-based approach to predict orthology relationships. Bioinform Biol. Insights 11, (1177932217690136).

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304 (5675), 1321–1325.

Berna, L., Alvarez-Valin, F., 2014. Evolutionary genomics of fast evolving tunicates. Genome Biol. Evol. 6 (7), 1724–1738.

Bessa, J., Tena, J.J., de la Calle-Mustienes, E., Fernandez-Minan, A., Naranjo, S., Fernandez, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., Gomez-Skarmeta, J.L., 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Dev. Dyn. 238 (9), 2409–2417.

Braasch, I., Gehrke, A.R., Smith, J.J., Kawasaki, K., Manousaki, T., Pasquier, J., Amores, A., Desvignes, T., Batzel, P., Catchen, J., Berlin, A.M., Campbell, M.S., Barrell, D., Martin, K.J., Mulley, J.F., Ravi, V., Lee, A.P., Nakamura, T., Chalopin, D., Fan, S., Wcisel, D., Canestro, C., Sydes, J., Beaudry, F.E., Sun, Y., Hertel, J., Beam, M.J., Fasold, M., Ishiyama, M., Johnson, J., Kehr, S., Lara, M., Letaw, J.H., Litman, G.W., Litman, R.T., Mikami, M., Ota, T., Saha, N.R., Williams, L., Stadler, P.F., Wang, H., Taylor, J.S., Fontenot, Q., Ferrara, A., Searle, S.M., Aken, B., Yandell, M., Schneider, I., Yoder, J.A., Volff, J.N., Meyer, A., Amemiya, C.T., Venkatesh, B., Holland, P.W., Guiguen, Y., Bobe, J., Shubin, N.H., Di Palma, F., Alfoldi, J., Lindblad-Toh, K., Postlethwait, J.H., 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. 48 (4), 427–437.

Bray, D., 2003. Molecular networks: the top-down view. Science 301 (5641), 1864–1865.

Bray, N., Dubchak, I., Pachter, L., 2003. AVID: a global alignment program. Genome Res. 13 (1), 97–102.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K.R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Aicha, S.B., Fagotto, M., Gueroult-Bellone, M., Haeussler, M., Jacox, E., Lowe, E.K., Mendez, M., Roberge, A., Stolfi, A., Yokomori, R., Brown, C.T., Cambillau, C., Christiaen, L., Delsuc, F., Douzery, E., Dumollard, R., Kusakabe, T., Nakai, K., Nishida, H., Satou, Y., Swalla, B., Veeman, M., Volff, J.N., Lemaire, P., 2018. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res. 46 (D1), (D718-d725).

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Program, N.C.S., Green, E.D., Sidow, A., Batzoglou, S., 2003. LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 13 (4), 721–731.

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinforma. 10, 421.

Clarke, S.L., VanderMeer, J.E., Wenger, A.M., Schaar, B.T., Ahituv, N., Bejerano, G., 2012. Human developmental enhancers conserved between deuterostomes and protostomes. PLoS Genet. 8 (8), e1002852.

Corrado, M., Aniello, F., Fucci, L., Branno, M., 2001. Ci-IPF1, the pancreatic homeodomain transcription factor, is expressed in neural cells of Ciona intestinalis larva. Mech. Dev. 102 (1–2), 271–274.

D'Aniello, E., Pezzotti, M.R., Locascio, A., Branno, M., 2011. Onecut is a direct neural-specific transcriptional activator of Rx in Ciona intestinalis. Dev. Biol. 355 (2), 358–371.

Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965.

Doglio, L., Goode, D.K., Pelleri, M.C., Pauls, S., Frabetti, F., Shimeld, S.M., Vavouri, T., Elgar, G., 2013. Parallel evolution of chordate cis-regulatory code for development. PLoS Genet. 9 (11), e1003904.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32, (Web Server issue)(: W273-W279).

Harmston, N., Barešić, A., Lenhard, B., 2013. The mystery of extreme non-coding conservation. Philos. Trans. R. Soc. B Biol. Sci. 368 (1632), (20130021).

Hemberg, M., Gray, J.M., Cloonan, N., Kuersten, S., Grimmond, S., Greenberg, M.E., Kreiman, G., 2012. Integrated genome analysis suggests that most conserved non-coding sequences are regulatory factor binding sites. Nucleic Acids Res. 40 (16), 7858–7869.

Hiller, M., Schaar, B.T., Bejerano, G., 2012. Hundreds of conserved non-coding genomic regions are independently lost in mammals. Nucleic Acids Res. 40 (22), 11463–11476.

Howard, M.L., Davidson, E.H., 2004. cis-Regulatory control circuits in development. Dev. Biol. 271 (1), 109–118.

Ikuta, T., Satoh, N., Saiga, H., 2010. Limited functions of Hox genes in the larval development of the ascidian Ciona intestinalis. Development 137 (9), 1505–1513.

Irvine, S.Q., 2013. Study of cis-regulatory elements in the ascidian Ciona intestinalis. Curr. Genom. 14 (1), 56–67.

Ishibashi, M., Noda, A.O., Sakate, R., Imanishi, T., 2012. Evolutionary growth process of highly conserved sequences in vertebrate genomes. Gene 504 (1), 1–5.

Jensen, J., 2004. Gene regulatory factors in pancreatic development. Dev. Dyn. 229 (1), 176–200.

Jue, N.K., Batta-Lona, P.G., Trusiak, S., Obergfell, C., Bucklin, A., O'Neill, M.J., O'Neill, R.J., 2016. Rapid Evolutionary rates and unique genomic signatures discovered in the first reference genome for the Southern Ocean Salp, Salpa thompsoni (Urochordata, Thaliacea). Genome Biol. Evol. 8 (10), 3171–3186.

Kent, W.J., Baertsch, R., Hinrichs, A., Miller, W., Haussler, D., 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. USA 100 (20), 11484–11489.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S.R., Tan, G., Baranasic, D., Arenillas, D.J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W.W., Parcy, F., Mathelier, A., 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 46 (D1), D1284.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Dev. Dyn. 203 (3), 253–310.

Kritsas, K., Wuest, S.E., Hupalo, D., Kern, A.D., Wicker, T., Grossniklaus, U., 2012. Computational analysis and characterization of UCE-like elements (ULEs) in plant genomes. Genome Res. 22 (12), 2455–2466.

Lang, M., Hadzhiev, Y., Siegel, N., Amemiya, C.T., Parada, C., Strahle, U., Becker, M.B., Muller, F., Meyer, A., 2010. Conservation of shh cis-regulatory architecture of the coelacanth is consistent with its ancestral phylogenetic position. Evodevo 1 (1), 11.

Li, Y., Wang, W., Wang, F., Wu, Q., Li, W., Zhong, X., Tian, K., Zeng, T., Gao, L., Liu, Y., Li, S., Jiang, X., Du, G., Zhou, Y., 2017. Paired related homeobox 1 transactivates dopamine D2 receptor to maintain propagation and tumorigenicity of glioma-initiating cells. J. Mol. Cell Biol. 9 (4), 302–314.

Maeso, I., Irimia, M., Tena, J.J., Casares, F., Gomez-Skarmeta, J.L., 2013. Deep conservation of cis-regulatory elements in metazoans. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368 (1632), 20130020.

Marcovitz, A., Jia, R., Bejerano, G., 2016. Reverse genomics" predicts function of human conserved noncoding elements. Mol. Biol. Evol. 33 (5), 1358–1369.

Martinez-Morales, J.R., 2016. Toward understanding the evolution of vertebrate gene regulatory networks: comparative genomics and epigenomic approaches. Brief. Funct. Genom. 15 (4), 315–321.

McEwen, G.K., Goode, D.K., Parker, H.J., Woolfe, A., Callaway, H., Elgar, G., 2009. Early evolution of conserved regulatory sequences associated with development in vertebrates. PLoS Genet. 5 (12), e1000762.
Meader, S., Ponting, C.P., Lunter, G., 2010. Massive turnover of functional sequence in human and other mammalian genomes. Genome Res. 20 (10), 1335–1343.
Morris, S.C., 1999. The Crucible of Creation: The Burgess Shale and the Rise of Animals. Oxford University Press.
NCBI, R.C., 2018. Database resources of the national center for biotechnology information. Nucleic Acids Res. 46 (D1), D8–d13.
Nobrega, M.A., Ovcharenko, I., Afzal, V., Rubin, E.M., 2003. Scanning human gene deserts for long-range enhancers. Science 302 (5644), 413.
Olson, E.N., 2006. Gene regulatory networks in the evolution and development of the heart. Science 313 (5795), 1922.
Parker, H.J., Sauka-Spengler, T., Bronner, M., Elgar, G., 2014. A reporter assay in lamprey embryos reveals both functional conservation and elaboration of vertebrate enhancers. PLoS One 9 (1), e85492.
Parveen, N., Masood, A., Iftikhar, N., Minhas, B.F., Minhas, R., Nawaz, U., Abbasi, A.A., 2013. Comparative genomics using teleost fish helps to systematically identify target gene bodies of functionally defined human enhancers. BMC Genom. 14, 122.
Pearson, W.R., 2016. Finding protein and nucleotide similarities with FASTA. Curr. Protoc. Bioinforma. 53, (3.9.1-3.925).
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B.L., Couronne, O., Eisen, M.B., Visel, A., Rubin, E.M., 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444 (7118), 499–502.
Pennati, R., Ficetola, G.F., Brunetti, R., Caicci, F., Gasparini, F., Griggio, F., Sato, A., Stach, T., Kaul-Strehlow, S., Gissi, C., Manni, L., 2015. Morphological differences between larvae of the ciona intestinalis species complex: hints for a valid taxonomic definition of distinct species. PLoS One 10 (5), e0122879.
Pezzotti, M.R., Locascio, A., Racioppi, C., Fucci, L., Branno, M., 2014. Auto and cross regulatory elements control Onecut expression in the ascidian nervous system. Dev. Biol. 390 (2), 273–287.
Plessy, C., Dickmeis, T., Chalmel, F., Strahle, U., 2005. Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes. Trends Genet. 21 (4), 207–210.
Royo, J.L., Maeso, I., Irimia, M., Gao, F., Peter, I.S., Lopes, C.S., D'Aniello, S., Casares, F., Davidson, E.H., Garcia-Fernandez, J., Gomez-Skarmeta, J.L., 2011. Transphyletic conservation of developmental regulatory state in animal evolution. Proc. Natl. Acad. Sci. USA 108 (34), 14186–14191.
Sandelin, A., Bailey, P., Bruce, S., Engstrom, P.G., Klos, J.M., Wasserman, W.W., Ericson, J., Lenhard, B., 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genom. 5 (1), 99.
Sanges, R., Hadzhiev, Y., Gueroult-Bellone, M., Roure, A., Ferg, M., Meola, N., Amore, G., Basu, S., Brown, E.R., De Simone, M., Petrera, F., Licastro, D., Strahle, U., Banfi, S., Lemaire, P., Birney, E., Muller, F., Stupka, E., 2013. Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development. Nucleic Acids Res. 41 (6), 3600–3618.
Sekigami, Y., Kobayashi, T., Omi, A., Nishitsuji, K., Ikuta, T., Fujiyama, A., Satoh, N., Saiga, H., 2017. Hox gene cluster of the ascidian, Halocynthia roretzi, reveals multiple ancient steps of cluster disintegration during ascidian evolution. Zool. Lett. 3, 17.
Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal, M., Jensen, M.F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., Reinhardt, R., Chourrout, D., 2004. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. Nature 431 (7004), 67–71.
Shu, D.G., Morris, S.C., Zhang, X.L., 1996. A Pikaia-like chordate from the lower Cambrian of China. Nature 384, 157.
Sobral, D., Tassy, O., Lemaire, P., 2009. Highly divergent gene expression programs can lead to similar chordate larval body plans. Curr. Biol. 19 (23), 2014–2019.
Stolfi, A., Lowe, E.K., Racioppi, C., Ristoratore, F., Brown, C.T., Swalla, B.J., Christiaen, L., 2014. Divergent mechanisms regulate conserved cardiopharyngeal development and gene expression in distantly related ascidians. Elife 3, e03728.
Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X., Thisse, C., 2001. "Expression of the Zebrafish Genome during Embryogenesis (NIH R01 RR15402)." ZFIN Direct Data Submission. (⟨http://zfin.org⟩).
Thisse, B., Thisse, C., 2005. "Fast Release Clones: High Throughput Expression Analysis of ZF-Models Consortium Clones." ZFIN Direct Data Submission. (⟨http://zfin.org⟩).
Turner, E.E., Cox, T.C., 2014. Genetic evidence for conserved non-coding element function across species–the ears have it. Front. Physiol. 5, 7.
Vassalli, Q.A., Anishchenko, E., Caputi, L., Sordino, P., D'Aniello, S., Locascio, A., 2015. Regulatory elements retained during chordate evolution: coming across tunicates. Genesis 53 (1), 66–81.
Vavouri, T., Walter, K., Gilks, W.R., Lehner, B., Elgar, G., 2007. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. Genome Biol. 8 (2), R15.
Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., Turner, J.M., Bertelsen, M.F., Murchison, E.P., Flicek, P., Odom, D.T., 2015. Enhancer evolution across 20 mammalian species. Cell 160 (3), 554–566.
Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., Pennacchio, L.A., 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat. Genet. 40 (2), 158–160.
Walter, K., Abnizova, I., Elgar, G., Gilks, W.R., 2005. Striking nucleotide frequency pattern at the borders of highly conserved vertebrate non-coding sequences. Trends Genet. 21 (8), 436–440.
Westerfield, M., 2000. The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio) 4th ed.. University of Oregon Press, Eugene.
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3 (1), e7.
Yousaf, A., Sohail Raza, M., Ali Abbasi, A., 2015. The evolution of bony vertebrate enhancers at odds with their coding sequence landscape. Genome Biol. Evol. 7 (8), 2333–2343.
Yue, J.-X., Kozmikova, I., Ono, H., Nossa, C.W., Kozmik, Z., Putnam, N.H., Yu, J.-K., Holland, L.Z., 2016. Conserved noncoding elements in the most distant genera of cephalochordates: the goldilocks principle. Genome Biol. Evol. 8 (8), 2387–2405.
Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A., Flicek, P., 2018. Ensembl 2018. Nucleic Acids Res. 46 (D1), D754–D761.