OPEN ACCESS Freely available online

PLoS GENETICS

A *cis*-Regulatory Signature for Chordate Anterior Neuroectodermal Genes

Maximilian Haeussler[◔], Yan Jaszczyszyn[◔], Lionel Christiaen[¤], Jean-Stéphane Joly*

INRA group, UPR3294, Institute of Neurosciences Alfred Fessard, CNRS, Gif-sur-Yvette, France

Abstract

One of the striking findings of comparative developmental genetics was that expression patterns of core transcription factors are extraordinarily conserved in bilaterians. However, it remains unclear whether *cis*-regulatory elements of their target genes also exhibit common signatures associated with conserved embryonic fields. To address this question, we focused on genes that are active in the anterior neuroectoderm and non-neural ectoderm of the ascidian *Ciona intestinalis*. Following the dissection of a prototypic anterior placodal enhancer, we searched all genomic conserved non-coding elements for duplicated motifs around genes showing anterior neuroectodermal expression. Strikingly, we identified an over-represented pentamer motif corresponding to the binding site of the homeodomain protein OTX, which plays a pivotal role in the anterior development of all bilaterian species. Using an *in vivo* reporter gene assay, we observed that 10 of 23 candidate *cis*-regulatory elements containing duplicated OTX motifs are active in the anterior neuroectoderm, thus showing that this *cis*-regulatory signature is predictive of neuroectodermal enhancers. These results show that a common *cis*-regulatory signature corresponding to K50-*Paired* homeodomain transcription factors is found in non-coding sequences flanking anterior neuroectodermal genes in chordate embryos. Thus, field-specific selector genes impose architectural constraints in the form of combinations of short tags on their target enhancers. This could account for the strong evolutionary conservation of the regulatory elements controlling field-specific selector genes responsible for body plan formation.

Citation: Haeussler M, Jaszczyszyn Y, Christiaen L, Joly J-S (2010) A *cis*-Regulatory Signature for Chordate Anterior Neuroectodermal Genes. PLoS Genet 6(4): e1000912. doi:10.1371/journal.pgen.1000912

Editor: Claude Desplan, New York University, United States of America

Received July 16, 2009; **Accepted** March 17, 2010; **Published** April 15, 2010

Copyright: © 2010 Haeussler et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by INRA and CNRS, the French GIS Institut de la Genomique Marine, the Marine Genomics Network of Excellence (EU-FP6 contract no. GOCE-CT-2004-505403), the ANR projects CHOREGNET and CHOREVONET, and the Plurigenes STREP project LSHG-CT-2005-018673. MH was a recipient of a Marie Curie Early Stage Research Training Fellowship (MEST-CT-2004-504854). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: LC joined the faculty of New York University after the review process had commenced.

* E-mail: joly@inaf.cnrs-gif.fr

¤ Current address: Department of Biology, New York University, New York, New York, United States of America

◔ These authors contributed equally to this work.

Introduction

The concept of "selector genes" was introduced 30 years ago by Garcia-Bellido to define genes that interpret a transient regulatory state and specify the identity of a given developmental field [1]. The question of how embryos execute distinct and unique differentiation programs using these selector genes can be tackled by focusing on how gene expression is encoded in *cis*-regulatory elements and their field-specific *trans*-acting factors (TF).

This concept was more recently extended to terminal selector genes that coordinate the expression of differentiation genes to determine a given cell type [2]. In vertebrates, examples include the *Crx* TF that interacts with another TF to control the expression of target genes in rod photoreceptors [3–5]. In vertebrates as well as in flies, *Crx* and its *Drosophila* homolog *Otd* act through a small *cis*-regulatory motif overrepresented in the elements flanking the target genes [6–10]. In addition to this evolutionary conserved network, many others in *Caenorhabditis elegans* and *Drosophila melanogaster* have shown that cell specific enhancers contain a common "tag" corresponding to a specific *cis*-regulatory motif, and that this motif is linked to one or a few terminal selector genes [11,12]. In contrast, during early development, very few studies have reported how a set of region-specific *cis*-regulatory elements responds to field-specific selector genes. In insects, one of the best characterized sets of functionally related *cis*-regulatory elements responds to the gradient of nuclearized *dorsal* TF in the early *Drosophila* embryo [13,14]. However, the regulatory mechanism of dorsal-ventral patterning is not enough conserved in chordates to allow comparative studies of the regulatory network.

A more general character of bilaterians is the tripartite organization of the nervous system along the antero-posterior axis [15]. In the posterior part (hindbrain and nerve cord), *Hox* genes are expressed in a colinear order. In the domain anterior to the *Hox* genes, several striking similarities in the relative expression patterns of other transcription factors have been noted in bilaterians [16–18]. The *OTX*-like homeobox transcription factors (*otd* in insects) are expressed in the anteriormost part of animals as diverse as cnidarians, insects, annelids, urochordates and vertebrates [19–21]. In chordates, OTX has a sustained expression in the anterior neuroectoderm and in derivatives of anterior ectoderm such as placodes, stomodeum [20,22]. In mice, null-mutants of this gene lack various head structures [23]. These

Author Summary

Regional identity in embryos is defined by a few specific transcription factors that activate a large number of target genes through binding to common tags in regulatory sequences. In chordates it is unclear if such tags can be identified in the *cis*-regulatory regions of regionally expressed genes. To address this question we focused on the anterior nervous system where *Otx* codes for a transcription factor that triggers expression of many other head-specific genes. We analyzed an element that is active in the region bordering the anterior nervous system in the marine invertebrate *Ciona intestinalis*. We found that the crucial binding sites have to be duplicated and close enough. One of the pairs is bound by OTX. We showed that anterior nervous system genes are often flanked by duplicated OTX binding sites. We confirmed by transgenic assays that about half of these genomic sequences are active and drive expression anteriorly. This study unravels a simple regulatory logic in the anterior enhancers. It indicates that although there are major changes in the organization of the binding sites at short evolutionary range, conserved expression patterns are partly generated by a duplicated organization of conserved binding sites for region-specific transcription factors.

results suggest that OTX-like proteins belong to a conserved developmental control system operating in the anterior parts of the brain, different from the one encoded by the *Hox* complexes [24].

Many homeodomain proteins bind to the core DNA sequence ATTA, but several subfamilies have longer binding specificities around this core [25,26]. OTX homeodomain proteins contain a lysine at position 50 which confers them additional specificity to guanines 5′ of the ATTA motif, resulting in a core recognition sequence of GATTA/TAATC [27]. The DNA binding domains of homeobox gene families are highly similar over large evolutionary distances and cross-species experiments have demonstrated that the OTX proteins can be exchanged between flies, mice and human without major developmental defects [28,29], and more recently between ascidians and mice [24,30].

For studies of anterior nervous system development, the ascidian *Ciona intestinalis* offers the advantage of a simple chordate body plan with the canonical tripartite brain along the antero-posterior axis [31]. In addition, the genome is small, with short intergenic regions which can be aligned with another ascidian species, thus simplifying the identification of *cis*-regulatory elements [32]. Moreover, complete expression patterns have been determined for thousands of genes and are readily available in public databases [33–35]. Therefore, *Ciona intestinalis* constitutes an ideal model system for combining whole genome bioinformatics and experimental *cis*-regulatory analyses.

Here, we first focus on one single anterior ectodermal enhancer in *Ciona intestinalis*. Its detailed analysis points to an internal tandem-like structure and underscores the key role of the selector gene *Otx*. We then examine if other duplicated putative binding sites for OTX preferentially flank anteriorly expressed genes in the genome.

Results/Discussion

D1 mediates the initiation of *Ci-pitx* expression in the anterior neural boundary (ANB)

We have previously described an enhancer sequence (called "D1", 323bp) that controls expression of the *Ciona intestinalis Pitx* gene in a sub-region overlapping the neural and the non neural ectoderm called the anterior neural boundary (ANB) [36]. For the sake of simplicity, and although ANB has a dual origin, we label it as a derivative of the neuroectoderm and call the region composed of anterior epidermis, ventro-anterior sensory vesicle and ANB, the "anterior neuroectoderm". For this study, we used a minimal 206 bp fragment of D1 that is sufficient to drive reporter gene expression in the ANB and divided it into five parts (D1a-e, Figure 1A and Figure 2A) for further analysis. Deletion of the first 16pb (D1a, Figure 2A) resulted in the D1bcde fragment (Figure 1A) and led to ectopic reporter gene expression in the anterior epidermis (ae) and ventro-anterior sensory vesicle (vasv) in addition to the expected expression in the ANB. All these elements indicate that D1 responds to neuroectodermal *trans*-activating factors that are not restricted to the ANB and that D1a contains motifs bound by a repressor factor that restricts D1 expression to the sole ANB.

We tested whether D1bcde controls the onset of *Ci-pitx* expression in the ANB. Endogenous *Ci-Pitx*-gene expression was not detected in ANB cells before the initial tailbud stage [37,38], suggesting that it starts at this stage. To test whether D1bcde recapitulates the temporal pattern of *Ci-Pitx* expression, we assayed reporter gene expression by either X-gal staining or lacZ *in situ* hybridization on the same batch of electroporated embryos fixed at successive stages. The rationale is to take advantage of the delay in β-galactosidase protein synthesis (e.g. [39]), which should produce a marked difference between X-gal and *in situ* staining shortly after the onset of reporter gene expression. We could detect neither lacZ RNAs nor β-galactosidase activity *before* the initial tailbud stage. At this stage, however, *lacZ* transcripts could be detected in 55.4% (n = 46 of N = 71) of the embryos while only 7% (n = 5 of N = 83) showed positive ANB cells after X-gal staining (Table S1). Hence, D1bcde-driven transcription starts at the same time as the endogenous *pitx* gene, which indicates that the D1bcde enhancer element triggers the initiation of *Ci-pitx* expression in ANB cells.

Short blocks of conserved nucleotides are required for D1 enhancer activity

Conservation between *Ciona intestinalis* and *savignyi* genomic sequences is not uniformly distributed throughout conserved non coding elements (CNEs) but rather concentrated in short blocks of identical nucleotides, which point to candidate transcription factor binding sites (TF-BS; Figure 1A, Figure 2A). We identified four classes of putative TF-BS based on nucleotide composition and by querying binding site databases [40,41]. One of them matches the OTX/K-50 *paired* homeodomain consensus sequence (sites O1 and O2, Figure 1A and Figure 2A). Other sites, called T (T/A-rich), G (G/C-rich) and M, bear resemblance to Forkhead, Smad and Meis family factors, respectively (Figure 1A). Some of them (P, T1, T2 were not completely conserved in the genome alignment. But each class of these candidate binding sites was represented at least twice in the minimal D1bcde element. The function of these candidate TF-BS was tested by introducing point mutations in the corresponding blocks of conserved sequences, followed by reporter gene expression assays. With the exception of mutations disrupting the "M" sites, each one of the individual modifications of O, T and G sequences reduced reporter gene expression in the anterior neuroectoderm derivatives (Figure 1B). Taken together, these observations indicate that D1 enhancer activity requires at least two copies of each one of three distinct classes of conserved putative TF-BS (Figure 1).

A tandem organization of binding sites is required for D1 activity

The aforementioned observation that the essential putative binding sites occur several times in the enhancer led us to

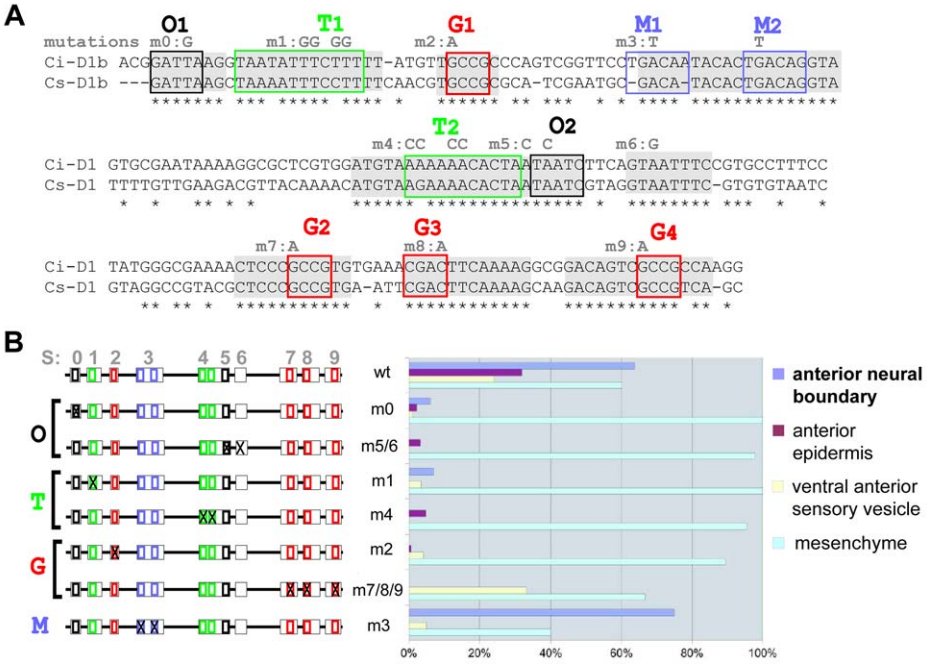


Figure 1. Mutational analysis of D1(bcde) reveals at least three binding sites necessary for its activity. (A) Alignment of D1(bcde) (195pb) between Ci/Cs. Conserved blocks in D1(bcde) with putative transcription factor binding sites (TFBS). Four classes of putative TFBS within ten blocks of conserved sequence: O1 and O2 display the characteristic GATTA/TAATC core sequence that is known to bind K50 *Paired*-related homeodomain proteins, including otx/otd, pitx and goosecoid. Putative Fox protein binding sites were identified in the blocks called T1 and T2 (T/A-rich). Blocks M1 and M2 show similarity to known binding sites of the TALE-class Meis/TGIF homeodomain proteins. Finally, putative Smad1/5/8 binding sites were observed in blocks labelled G1, G2, G3 (G/C-rich). Conserved blocks of nucleotides are highlighted in grey. Putative binding sites are represented by colored boxes. Mutations designed to disrupt the binding specificity of each sites are indicated in grey above the sequence (m0 to m9). (B) Schematic representation of D1(bcde) and series of mutations introduced into the pD1bcde:CES2 construct. On the right is shown for each construct the rate of mid-tailbud embryos expressing the lacZ reporter in the anterior neural boundary (anb), the anterior epidermis (ae), the ventro-anterior sensory vesicle (vasv) and the mesenchyme (mes) (wt n = 444; m0 n = 198; m1 n = 56; m2 n = 348; m3 n = 40; m4 n = 58; m5/6 n = 182; m7/8/9 n = 29). Most mutations led to a general reduction of the expression level in all of aforementioned domains, while mesenchyme expression remained high. Only construct m3 (block M1 and M2) retained the activity in the anterior neural boundary.
doi:10.1371/journal.pgen.1000912.g001

investigate whether the structure of D1 bears functional significance to its enhancer activity. Notably, the 54-bp D1(ab) element (Figure 2A) contains the three previously mentioned conserved motifs O, T and G in addition to a putative Pax binding site (P), but D1(ab) is not sufficient to enhance reporter gene transcription (Figure 2C). Since each of the critical sites is represented at least twice in the full length enhancer, we asked whether D1 enhancer activity relies on this tandem-like repetition of essential binding sites. We created artificial enhancers containing multiple copies of D1(ab) and found that as little as two copies of D1(ab) were sufficient to drive strong lacZ expression in the anterior neuroectoderm (88% of 167 tailbud embryos (Figure 2D and 2E)).

To test whether enhancer activity of the D1(ab) dimer relies specifically on the duplication of O, T and G sites, we introduced point mutations in the second D1(ab) copy. Each of these mutations strongly reduced enhancer activity (Figure 2D). These observations are reminiscent of the requirement for multiple copies of *bicoid* binding sites for target gene activation during *Drosophila* head development [42] and the general tendency of binding sites to occur in clusters [43]. Our results demonstrate that duplications of critical binding sites are essential for D1 enhancer activity and do not constitute mere redundancy.

We next asked whether the distance between the duplicated 54bp elements influenced the activity of the artificial D1(ab) dimer. To this aim, we designed sequences that are not predicted to bind any characterized transcription factors from the Uniprobe database (see Materials and Methods) and inserted 25, 50, 75 and 150bp spacers between the D1(ab) duplicates. Overall, enhancer activity of these constructs is reduced compared to the original D1(ab) dimer and almost completely abolished with the 75bp and 150bp spacers (Figure 2F). Similar structural constraints were reported in the *Drosophila knirps* enhancer, which was shown to require a specific arrangement of duplicated *bicoid* binding sites for activation [44,45]. Similarly, *even-skipped* enhancers contain a conserved structure of paired binding sites [46] and duplicated and relatively distant (30–200bp) TFBS are necessary for a correct activity of the SV40 enhancer [47] and the *lac* operon [48]. Taken together, our observations demonstrate that D1 enhancer activity relies on the clustering of duplicate short conserved sequences.

Ci-Otx affects D1 enhancer activity

Among D1(ab) essential putative binding sites, the GATTA/TAATC "O" sequences correspond to the consensus for K50-*Paired* homeodomain proteins. In ascidians, this family includes Goosecoid, Pitx and Otx. Only Otx, is expressed in the right time and place to account for D1 enhancer activation in the anterior neuroectoderm in *Ciona* [20] and there is only one *Otx* gene in the *Ciona intestinalis* genome.

A functional study using morpholino antisense oligonucleotides in *Halocynthia roretzi* - another ascidian species - showed that the *Hr-Otx* knockdown strongly perturbs anterior neuroectoderm development, mostly because it is required for early specification events in the gastrula [49]. To avoid this early effect, we used targeted expression of dominant-negative and hyper-active versions of the

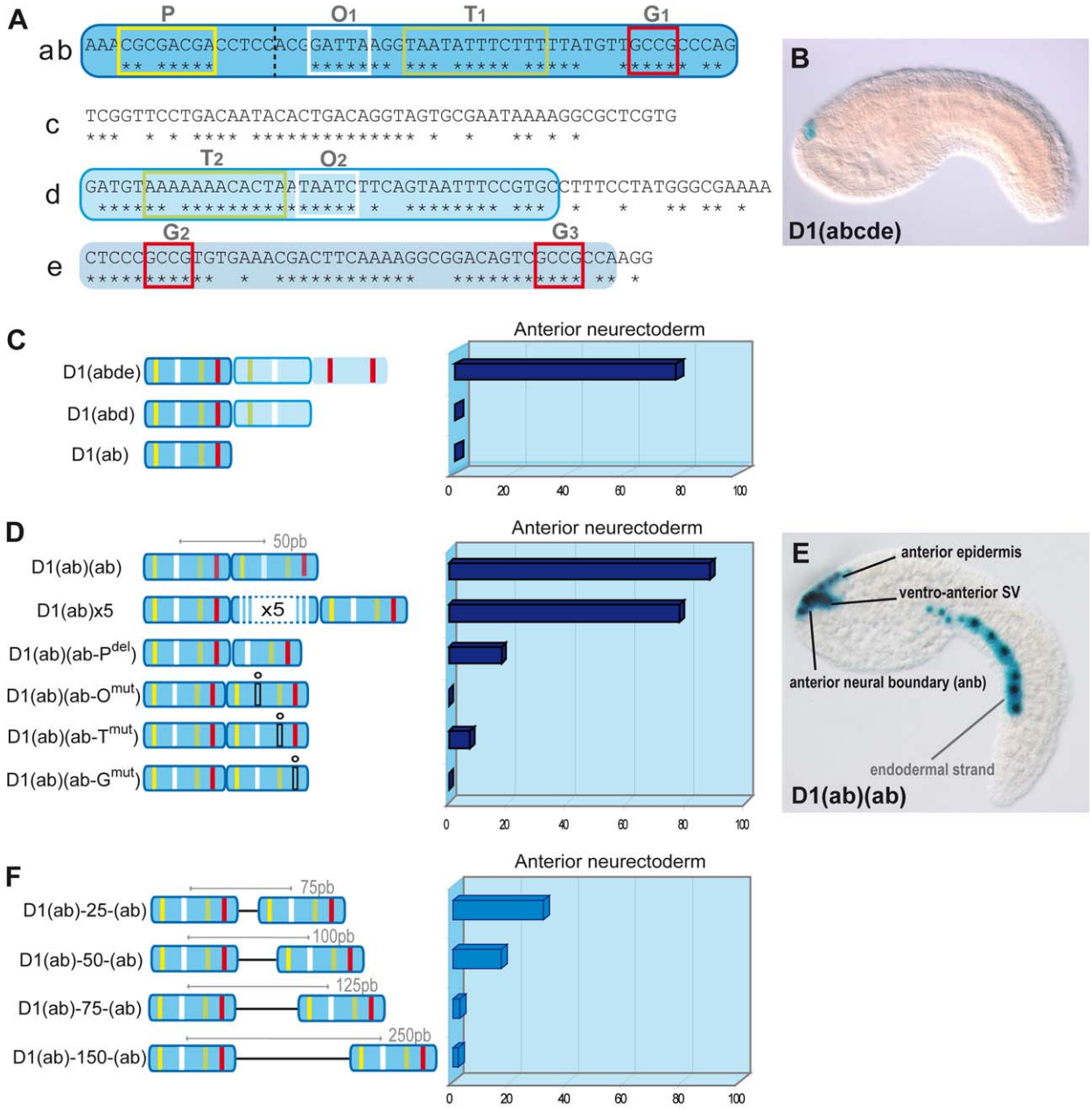


Figure 2. Artificial enhancer constructs reveal a tandem-*like* structure. (A) D1(abcde) *Ci-Pitx* enhancer. Stars show conserved positions with *Ciona savignyi*. The element has been divided into five parts (a, b, c, d, e). (ab) fragment is in dark blue, (d) in light blue with blue outline, (e) in light blue. Conserved nucleotides stretches contain putative transcription factor binding sites (TF-BS). O1 and O2 sites (white box) correspond to the BS for K-50 *Paired* homeodomain proteins and P (yellow), T1, T2 (green) and G1, G2, G3 (red) resemble Pax, Forkhead and Smad protein consensus BS respectively. (B) Side view of an early-tailbud embryo electroporated with pD1abcde:CES2:lacZ: expression in the anterior neural boundary (ANB). In some cases, ectopic expression occurs in the mesenchyme and the tail muscles (not shown). (C) Expression of artificial enhancers in the anterior neuroectoderm of mid-tailbud embryos. LacZ expression was observed after two hours of staining. The D1(abde) construct drives lacZ expression in the anterior neuroectoderm (ANB), ventro-anterior sensory vesicle (vasv) and epineural epidermis (ene) in 77.9% of developed embryos (n = 57). Deletions of (e) or (de), but not (c), abolish LacZ expression (D1(abd), D1(ab)). (D) Two (D1(ab)(ab)) or five (D1(ab)×5) copies of the 54bp D1(ab) drive expression in most of the embryos (88% (n = 167) and 77% (n = 72), respectively). Only 17% of the embryos express lacZ following the deletion of the second P site (D1(ab)(ab-P_{del}), n = 90). Mutations of O, T, and G sites in the second copy of (ab) strongly decrease lacZ expression. (D1(ab)(ab-O^{mut}): 0% (n = 137), D1(ab)(ab-T^{mut}): 7% (n = 84), D1(ab)(ab-G^{mut}): 0% (n = 118)). (E) Mid-tailbud embryo electroporated with pD1(ab)(ab):CES2:lacZ. Expression is visible in the ANB, the vasv, the ene and less frequently in the endodermal strand. (F) Introduction of spacer regions of 25/50/75/150 bp between the two D1(ab) fragments strongly decreased the activity of the tandem constructs. From 88% (D1(ab)(ab)) to 31.6% (n = 76), 16.9% (n = 154), 2.5% (n = 79) and 1.9% (n = 106), respectively. Scores were obtained after one week of LacZ revelation.
doi:10.1371/journal.pgen.1000912.g002

Ci-OTX protein to interfere with its endogenous activity specifically after gastrulation. We thus engineered protein chimeras between the *Ci*-OTX homeodomain and the *Drosophila* engrailed repressor peptide or the VP16 *trans*-activation domain to create dominant-negative (OTX:EnR) or hyper-active (OTX:VP16) forms, respectively. We then used the *Ci-Six3 cis*-regulatory

DNA to drive expression of these fusion proteins in a region that encompasses the ANB (Figure S1). These constructs were co-electroporated with the *Ci*-Distal-*Pitx* reporter plasmid, which contains the D1 enhancer with the two essential O1 and O2 K50-*Paired* binding sites [36], and the number of anterior neuroectodermal cells expressing the reporter gene was scored at the mid-tailbud stage (Figure 3). In control embryos expressing a *Ci-Six3*:Venus construct, an average of 2.78 anterior neuroectodermal cells per embryo activated the *Ci-Pitx* reporter construct, which can be accounted for by the mosaic incorporation of the transgene in the four ANB cells (Figure 3A and 3C). In contrast, targeted expression of *Ci*-OTX fusion proteins significantly altered *Ci-Pitx* reporter gene expression in the anterior neuroectoderm: the engrailed fusion inhibited ANB expression, while OTX:VP16 produced ectopic activation in surrounding neuroectodermal cells (Figure 3B–3D). Notably, OTX:VP16 also boosts expression of the ab dimer construct, and is not sufficient to induce overexpression when coelectroporated with one dimer construct bearing one O mutation (data not shown). This indicates that OTX:VP16 indeed binds to the GATTA binding sites. These observations strongly suggest that *Ci*-OTX trans-activating inputs are required for D1 enhancer activity in the anterior neuroectoderm. In addition, widespread expression of *Ci-Otx* in the anterior neuroectoderm contributes to the broad D1 trans-activation potential that encompasses the ANB, anterior epidermis and anterior sensory vesicle and is probably defined in D1 by the conserved GATTA/TAATC duplicated sequences. We cannot exclude the possibility that endogenous *Ci-Pitx* maintains its own expression through the same GATTA/TAATC BS, which binds PITX as well as OTX proteins. However, *Otx* is the best candidate for the onset of D1 activity, which begins exactly at the same time as the onset of the endogenous *Ci-Pitx* expression.

Tandems of OTX binding sites preferentially flank anterior neuroectodermal and ectodermal genes

Of the three different duplicated BS that we identified for the ANB expression domain of *Pitx* and that we suppose to be specific to this restricted area of the anterior neuroectoderm, we concentrated our effort only on the binding sites for OTX as these are the only ones assignable to a well-characterized transcription factor. The observation that the transcriptional response to the broadly expressed head field-selector gene *Otx* is

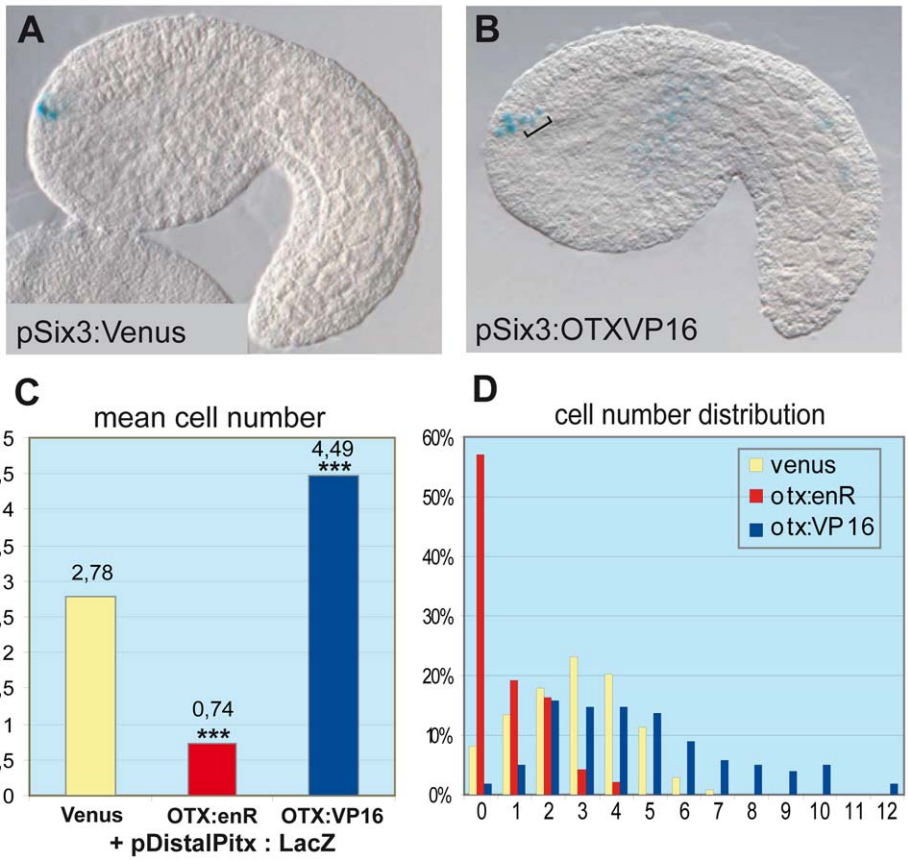


Figure 3. OTX fusions influence the activity of the *Ci-pitx cis*-regulatory element. Co-electroporation of *Pitx* full length distal region (pDistal*Pitx*:lacZ, 5.3kb, containing D1), respectively with pSix3:Venus (control), with pSix3>OTX_{HD}::enR (dominant negative OTX) and pSix3>OTX_{HD}::VP16 (hyper-active OTX). (A) Side view of an embryo co-electroporated with pDistal*Pitx*:lacZ and pSix3:Venus. Three positive cells can be detected in the ANB. (B) Co-electroporation of pDistal*Pitx*:lacZ and pSix3>OTX_{HD}::VP16. In addition to the expression in the ANB, ectopic expression is detected in the ASV cells (bracket) where OTX:VP16 is produced under the control of p*Six3*. (C) Numbers of lacZ expressing cells decrease with the OTX_{HD}::enR protein (2.78 to 0.74 cells) and increase with the OTX_{HD}::VP16 protein (2.78 to 4.49 cells) The distributions differ significantly from the control in both groups according to two *Wilcoxon–Mann–Whitney* two-sample rank-sum tests: Control/OTXenR ($U_{OTXenR} = 3891$, $n(emb)_{OTXenR} = 139$, $n(emb)_{ctrl} = 131$, $P = 2.536e\text{-}07$ two-tailed) and control/OTXVP16 ($U_{OTXVP16} = 15582.5$, $n(emb)_{OTXVP16} = 98$, $n_{ctrl} = 131$, $P < 2.2e\text{-}16$ two-tailed). (D) Distributions of cell numbers in the ANB and ASV after co-electroporation of Distal*Pitx*:lacZ and OTX fusions (yellow: control, red: enR fusion, blue: VP16 fusion; X-axis: cell numbers, Y-axis: proportions of embryos). Each experiment has been performed twice.
doi:10.1371/journal.pgen.1000912.g003

mediated by duplicated GATTA motifs led us to investigate whether this regulatory architecture was overrepresented in candidate *Otx* target genes in early tailbud embryos. At this stage, *Ci-Otx* expression extends over a broad domain referred to as in the anterior neuroectoderm, which derives from the a-line blastomeres and encompasses the ANB as well as other specific neurectodermal territories such as the anterior sensory vesicle, palps, a-line epidermis and rostral trunk epidermal neurons (RTEN). Therefore, we reasoned that candidate Otx target genes could, in principle, be expressed in all or part of the anterior neurectoderm. Hence, we asked whether duplicated GATTA motifs –the candidate signature for Otx binding- were enriched in the conserved noncoding sequences flanking genes with conspicuous expression in the anterior neurectoderm.

To this end, we obtained whole mount *in situ* hybridization data for 1518 genes showing tissue-specific expression from the model organism database ANISEED (December 2007, http://crfb.univ-mrs.fr/aniseed, see also Protocol S1). From these, we selected genes that are expressed in the central nervous system (CNS) and the ANB and classified them into different territories according to their expression along the antero-posterior axis: following previous reports [49–51], the ascidian visceral ganglion and the nerve cord were considered as "posterior" CNS whereas the whole sensory vesicle, including the ANB, constitute the "anterior" nervous system. This lead to a detailed annotation of nervous system expression patterns for 258 genes (Table S2). From this list we retained only those 100 genes that are specifically expressed in the anterior and not the posterior parts of the CNS. Finally, we obtained annotations for additional genes expressed in tissues like muscle, epidermis or notochord, from the database ANISEED. This latter set of genes was used as negative controls, which allowed for background definition for further statistical analyses. In total, our set includes annotations for 904 genes.

We then aimed at studying the distribution of duplicated short DNA motifs around these 904 genes to find those that show a bias towards genes expressed in the anterior or posterior nervous system, muscle, epidermis or notochord. We concentrated on conserved non-coding elements (CNEs), as these have been shown to be enriched in developmental enhancers [52,53]. To obtain these elements for the genome of *Ciona intestinalis*, we created a whole-genome alignment with *Ciona savignyi* [54] and removed aligned positions in transcribed regions from it. This results in 168306 CNEs with an average length of 143 bp.

Then, we searched for duplicate matches to all 512 possible pentamers within 125 bp of all CNEs in the *Ciona intestinalis* genome and subsequently calculated the number of tissue-specific neighboring genes associated to each duplicated conserved pentamer and tissue. The rationale for using consensus and not matrix-based searches was that all subclasses of homeodomain proteins have well characterized binding sites that resemble pentamer motifs without degenerate positions [25,26]. For the window size parameter, we observed from our case study that the sites had to occur in duplicates with a maximum distance of about 125bp, which was the total length of the fragment between both OTX-sites in the 75bp spacer construct. The score we chose was inspired by [55]; it does not require a sequence background model. This "motif-tissue-score" is the negative logarithm of the binomial probability to obtain a certain number of annotated genes from a given tissue by chance and therefore reflects the association of individual pentamer motifs with specific tissues.

Our first observation was that a duplicated OTX (GATTA) motif within 125 basepairs appears among the motifs with the highest score in the anterior CNS region (Table S3). For instance, genes containing duplicated GATTA motifs within 125bp in their flanking conserved genomic DNA are more likely to be expressed in the anterior nervous system than in any of the other tissues used in this analysis, including the posterior CNS (26% versus 12% or less, Table 1).

We then set out to assess the robustness of this analysis to variations of all three parameters: copy-number, window size and gene annotation. We varied the number of motif-duplicates from one to four and still obtained the highest motif-tissue scores in the anterior region with two copies. Increasing the window size from 25bp to 300bp did not change the scores to a large extent and the relative order between the anterior nervous system and other tissues always remained the same (Figure S3). The influence of errors in the manual annotation process was investigated by a simulation: we randomized 10% of all gene annotations and repeated this procedure 100 times. The 95% confidence intervals from these are small compared to the total differences between the tissues (Figure 4).

These results indicate that a biased distribution of GATTA motifs in CNEs supports the model of anterior ectodermal expression based on D1 enhancer analysis. We conclude that the presence of duplicated and conserved OTX binding sites in a cis-regulatory element is a signature for anterior neuroectoderm enhancer activity.

Duplicated GATTA–motifs identify functional anterior ectoderm enhancers

We then sought to test whether conserved sequences containing duplicated GATTA motifs act as enhancers in the anterior neuroectoderm. Out of all 53 CNEs with at least two conserved GATTAs in a 125 bp window that flank genes expressed in the anterior nervous system, we selected 30 CNEs. We succeeded in cloning 23 of them into a lacZ expression vector. After electroporation, we observed that ten of them are active enhancers in various domains of the anterior neuroectoderm derivatives, where *Otx* is expressed at the tailbud stage (Figure 5, Figure S2, and Table S4). The remaining non-coding regions were inactive or drove non-specific expression in the mesenchyme, as is often observed in electroporated ascidian embryos [56,57]. This ratio of positive elements is high compared to a previously published enhancer screen of random DNA fragments (5 active enhancers out of 138 tested fragments) [57] and similar to a prediction based

Table 1. Antero-posterior distribution of enhancers with 2×GATTA tags.

Tissue	Genes in this category	Genes flanked by 2×GATTA/125 bp in a conserved non-coding alignment	Percentage
anterior nervous system (specific)	100	26	0.26 *
posterior nervous system (specific)	58	7	0.12 *
notochord	346	18	0.05
epidermis	523	26	0.04
muscle	143	4	0.02

The percentage of positive anterior nervous system and positive posterior nervous system genes flanked by two GATTAs are significantly different ($P = 0.043$) according to a Fisher Exact two-tailed test.
doi:10.1371/journal.pgen.1000912.t001

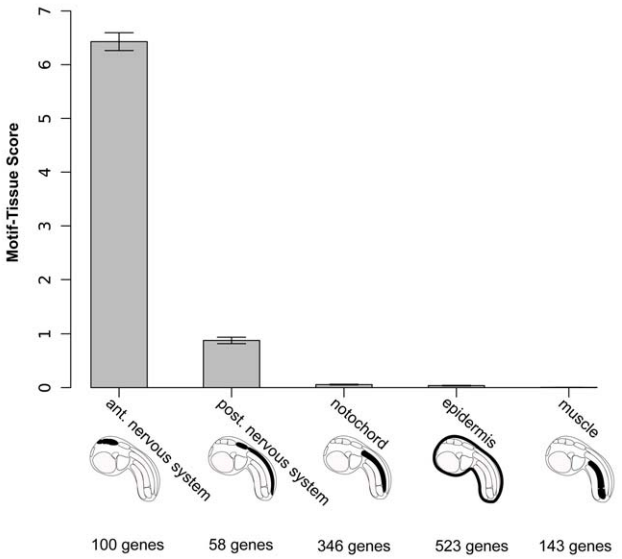


Figure 4. Motif-tissue scores for the motif 2×GATTA/125bp against genes expressed in various tissues. These territories are also visualized on schematic representation of an ascidian tailbud embryo. The number of genes is indicated for each category. To illustrate that changes in gene annotation are very unlikely to affect the overall ranking, we shuffled 10% of the gene-tissue assignments, repeated the procedure 100 times and plotted 95%-confidence intervals with error bars.
doi:10.1371/journal.pgen.1000912.g004

on binding site occurrences in *Drosophila* muscle founder cells (6 out of 12 tested elements) [58].

We were unable to identify additional motifs that would be predictive of enhancer activity in the anterior neurectoderm (Figure S4). However, additional motifs are required in natural enhancers, as we showed that pentamers of GATTA alone were unable to drive reporter gene activity in the Otx expression domain (data not shown). The diversity of expression patterns obtained with the ten active enhancers rather suggests that different transcription factors, each specific for a subdomain of the anterior neuroectoderm, might be implied in the activity of these elements.

Thus, while there might be additional motifs necessary for anterior neuroectoderm expression, this study shows the importance of the duplicated GATTA regulatory architecture as a predictive tag for the identification of anterior enhancers in chordates.

Could a signature based on GATTA-sites also be predictive in vertebrates? [52] reported that GATTA is over-represented in forebrain enhancers and used it as one of six motifs to predict forebrain enhancers in the mouse genome. We also found other overrepresented motifs in anteriorly expressed genes (see Table S3). Therefore, as determined experimentally with the D1 element, additional complexity must supplement the duplicated GATTA sites to achieve a cell-specific expression. Similar approaches performed in *Drosophila* and *Caenorhabditis* have identified several binding sites, which correspond to factors that specify a particular fate or behaviour in a combinatorial fashion, such as the myogenic factors [58,59]. However, our study identifies for the first time a *cis*-regulatory signature that determines the transcriptional response to a "master" homeobox gene in a simple chordate and establishes a model for genome-wide predictions of tissue-specific enhancers.

Materials and Methods

Animals

Adult *Ciona intestinalis* were purchased at the Station de Biologie Marine de Roscoff (France) and maintained in artificial sea water at 15°C under constant illumination. Eggs and sperm were collected from dissected gonads and used in cross fertilizations. Electroporations, using 70 µg of DNA, and LacZ stainings were performed as previously described [36]. Embryo staging at 13°C were done according to [60,61]. Images were taken on a Leica DMR microscope.

Mutations in D1bcde

For the mutational analysis of the enhancer D1bcde (Figure 1), we omitted the first 16 bp (AAACGCGACGACCTCC) of D1abcde that were not conserved between *Ciona intestinalis* and *savignyi*. Each of the mutations was designed to perturb DNA-binding of the candidate *trans*-acting factors following various reports in the literature. Mutations were performed using the Stratagene Quick-Change Kit. Seven new constructs called m0, m1, m2, m3, m4, m5/6, m7/8/9 were generated. After each electroporation, we observed LacZ expression in the tissues of the anterior neural boundary, anterior epidermis, ventro-anterior sensory vesicle and mesenchyme. We obtained a semi-quantitative estimation of the promoter activity by calculating the percentage of positive embryos.

Artificial enhancers

Plasmids with artificial enhancers were designed by cloning inserts into the pCES2::lacZ vector that contains the basal *Ci-Fkh/FoxA* promoter [57]. Insert D1(ab) was generated by cloning two long complementary primers with XhoI/XbaI cohesive ends into pCES2. Inserts (abde), (abd), (ab)(ab-P^{del}), (ab)(ab-O^{mut}), (ab)(ab-T^{mut}), (ab)(ab-G^{mut}) were generated by cloning a second insert consisting of another couple of long complementary primers into the XbaI/BamHI site of D1(ab). The insert of D1(ab)×5 was designed *in silico*, synthetized by Genecust Europe (Luxembourg) and cloned into pCES2::LacZ between XhoI and BamHI. To obtain D1(ab)(ab), we cut out the first two parts of D1(ab)×5 with SalI/XhoI and ligated them into pCES2.

The spacer sequence between both (ab) parts of D1(ab)-xx-(ab) constructs was created *in silico* by avoiding all octamers bound by homeodomain factors from a large-scale DNA-protein binding assay [25]. We recursively added random nucleotides to an unbound sequence and backtracked if the new sequence contained an octamer with PBM enrichment score >0.3 from the UniProbe database [62]. These constructs, D1(ab)-xx-(ab) are also derived from D1(ab), but the insert was synthesized by GeneScript Corporation (Piscatway, NJ, USA). We amplified spacers of the appropriate length by PCR from the longer fragment and cloned them between the two duplicated (ab) fragment by restriction/ligation.

OTX fusions

A pSix3:Venus plasmid was digested by BamHI/EcoRI to eliminate the Venus/YFP reporter.

VP16 fusion. the OTX_{HD} fragment was amplified by PCR from tailbud *Ciona* cDNA using OTX_{HD}-F (CGGGATCCACAA-TGGTATACAGTTCGTCTAGAAAA) and OTX_{HD}-R (AA-ACCATGGGTTGTTGCACTTGTTGGCGACA) oligos and digested by BamHI/NcoI. The VP16 domain was amplified with VP16-F (AAGATATCGACAAACCATGGTGCAGCTG-GCACCACCGACCGATGTCAG) and VP16-R (AACAGCT-GGAATTCTTAGATATCCCCACCGTACTCGTCAATTC) oligos, and digested by NcoI/EcoRI. Both resulting fragments

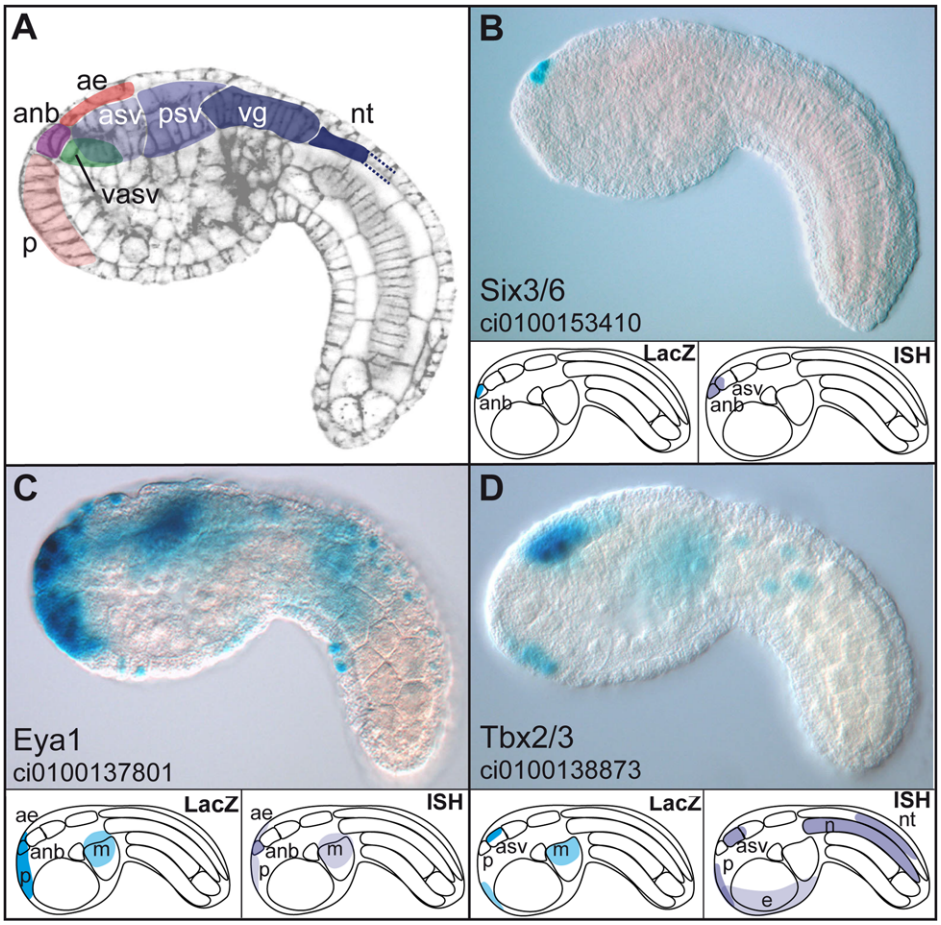


Figure 5. Enhancers with duplicated GATTA are active in the anterior region of the ascidian embryo. (A) Schematic representation of the main regions of gene expression in a mid-tailbud *Ciona intestinalis* embryo. Cell cortices are stained with Alexa-phalloidin (Christiaen et al. 2007). (B–D): expression domains of three enhancers, respectively from *Ci-Six3/6*, *Ci-Eya1* and *Ci-Tbx3* after electroporation and X-Gal staining at mid-tailbud stage. Lower panels show a schematic representation of the LacZ expression driven by the enhancer (left) and endogenous gene expression as assayed by *in situ* hybridization (ISH) (right). Enhancers can be subdivided into different classes following their expression domains: very restricted expression only in the ANB while the gene expression domain is slightly larger (Six3, (B)); broad anterior expression recapitulating more or less the endogenous expression pattern (Eya, (C)); only the most anterior expression domains are driven by the enhancer (Tbx3 (D)). anb: anterior neural boundary, asv: anterior sensory vesicle, psv: posterior sensory vesicle, vg: visceral ganglion, nt: neural tube, ae: anterior epidermis (or epineural epidermis), p: palps (precursors), m: mesenchyme, n: notochord. Lateral views, anterior to the left.
doi:10.1371/journal.pgen.1000912.g005

were ligated into the linearized pSix3 driver to obtain the pSix3:OTX_{HD}:VP16 construct.

EnR fusion. the OTX_{HD} fragment was amplified by PCR from tailbud cDNA using OTX_{HD}-F (CGGGATCCACA-ATGGTATACAGTTCGTCTAGAAAA) and OTX_{HD}-R (AA-ACCATGGGTTGTTGCACTTGTTGGCGACA) oligos. The enR repressor domain was amplified with enR-F (CTCGAG-GCCCTGGAGGATCGC) and enR-R (CGAATTCTATAC-GTTCAGGTCCT) oligos. Both fragments were fused by additionnal rounds of PCR using oligos that overlap the 3′ part of OTX_{HD} and the 5′ part of enR (enR(OTX)F:TGTCG-CCAACAAGTGCAACAACTCGAGGCCCTGGAGGATCG-C, OTX_{HD}(enR) R: GCGATCCTCCAGGGCCTCGAGTT-GTTGCACTTGTTGGCGACA). The resulting product was digested by BamHI/EcoRI and ligated into into the digested pSix3 driver to obtain the pSix3: OTX_{HD}:enR construct.

Constructs for the enhancer screen

Plasmids containing non-coding elements were created with the Gateway Technology System (Invitrogen Carlsbad, CA, USA). We cloned an AttR3/AttR4 Gateway Cassette from [63] into the XhoI/XbaI-site of pCES2 and called the resulting construct AttR3R4-pCES2. Predicted fragments were first amplified by primers including part of the flanking AttB3/AttB4-sequences and then extended by a subsequent PCR to the full length sequences of AttB3/AttB4. These fragments were recombined with BP clonase into the P3/P4-donor Vector [63] and the resulting entry vectors recombined with LR clonase into AttR3R4-pCES2 producing expression vectors.

In silico methods

Computational methods are described in Protocol S1. Programs that were used for whole-genome analyses are accessible at http://genome.ciona.cnrs-gif.fr/scripts/.

Supporting Information

Figure S1 The pSix3 driver. (A) Vista plot of the *Ci-Six3/6* locus, showing conserved sequences between *Ciona intestinalis* and *Ciona savignyi*: exons in blue and non-coding sequences in pink. The pSix3 construct (blue line) encompasses the 2 kb upstream the

initiator codon of the SIX3 protein. (B) *Ci-Six3/6* gene is expressed at the early neurula stage in the most anterior neural plate cell row (C) it continues to be expressed at the tailbud stage in the ANB and the anterior sensory vesicle. (D, E) The pSix3 construct drives reporter gene expression in a similar way to the endogenous gene at neurula and tailbud stages. It was used to drive expression of dominant-negative and hyper-active forms of the *Ci*-OTX protein in ANB precursor cells (Figure 3).
Found at: doi:10.1371/journal.pgen.1000912.s001 (3.20 MB PDF)

Figure S2 Images of LacZ-stained embryos and genomic context of the CNEs. Pictures of mid-tailbud stage embryos in side view electroporated with enhancer:pCES2:lacZ constructs and position of the enhancer on the *Ciona* genome. When more than one GATTA-rich enhancer has been found in a locus, we chose to clone one or two of them, according to the following criteria: the conservation on the vista plot (higher conservation and unique peak were preferentially chosen), the absolute number of conserved GATTA, the proximity and relative position to the transcription start (intronic and 5′ were preferred to 3′). Conserved regions with duplicated GATTAs on the first track. Positive enhancers are represented in blue boxes, negative regions in empty blue boxes, other regions in grey. The number of conserved GATTAs in the region is indicated above the boxes. When available, the best matching human gene model (BLAST), the JGI1 gene model number and VISTA conservation are shown below. (A, B) Six3/6: expression of the lacZ reporter in the ANB. (C, D) Eya1: expression in the ANB, anterior head and palp epidermis, mesenchyme, and weak expression in the tail epidermis. (E, F) Tbx3: expression in the dorsal part of the ASV and weaker in the palp epidermis. (G, H) FoxF: expression in the head epidermis surrounding the ASV. (I, J) Ets1: expression in the dorsal ASV, palps epidermis and tail tip epidermis less frequently. (K,L) Efnb3: broad expression in the anterior head epidermis. (M, N) Zswim: expression in the ASV, the PSV and weaker expression in the notochord and the tail epidermis. (O, P) Otx: expression in the most posterior part of the ASV and in the PSV. (Q, R) ci0100154565: expression in the ANB, the posterior sensory vesicle and weak in the mesenchyme. (S, T) Hes1: expression in the ASV and broad expression in the mesenchyme and tail muscles and epidermis. Endogenous expressions can be seen in Table S2.
Found at: doi:10.1371/journal.pgen.1000912.s002 (8.16 MB PDF)

Figure S3 Motif-tissue scores of the 2×GATTA *cis*-regulatory signature depending on window size. Window size between two GATTA motifs has been varied from 25 to 300 bps. It can be seen that the window size parameter does not to a large extent influence the results and never changes the order of the tissues; the anterior nervous system is always the highest ranking tissue. Procedures are described in Protocol S1.
Found at: doi:10.1371/journal.pgen.1000912.s003 (0.50 MB PDF)

Figure S4 Motif occurrences in predicted enhancers. The number of putative binding site motifs from Figure 1 (T and G motifs) and the two topscoring motifs from Table S3 are represented by colors according to the legend on the right. The name of the constructs refer to the regions electroporated and described in Table S4. (−) indicates inactive and (+) indicates active construct (grouped in the red box). All constructs contain at least two GATTA motifs as this was the main criteria for their identification. As can be seen from this diagram, no motif occurs preferentially in the positive constructs.
Found at: doi:10.1371/journal.pgen.1000912.s004 (0.24 MB PDF)

Table S1 Timing of D1bcde-driven reporter gene expression.
Found at: doi:10.1371/journal.pgen.1000912.s005 (0.40 MB PDF)

Table S2 Annotation of expression patterns for 258 genes at early-/mid-tailbud stages.
Found at: doi:10.1371/journal.pgen.1000912.s006 (1.40 MB PDF)

Table S3 Motif-Tissue-Scores of the ten highest-scoring motif duplicates in the anterior and posterior nervous system.
Found at: doi:10.1371/journal.pgen.1000912.s007 (0.30 MB PDF)

Table S4 Overview of enhancer screen CNEs and stained territories after electroporation.
Found at: doi:10.1371/journal.pgen.1000912.s008 (0.43 MB PDF)

Protocol S1 *In silico* protocols.
Found at: doi:10.1371/journal.pgen.1000912.s009 (0.03 MB PDF)

Acknowledgments

We wish to thank Laurent Legendre, Aurélie Heuzé, and Charlotte Bureau for technical assistance; Fabrice Daian and Patrick Lemaire for *in situ* hybridization annotations; Martha Bulyk for advice on the generation of neutral spacer sequences with Uniprobe; Hiram Clawson, Angie Hinrichs, Olivier Mirabeau, and Matthieu Defrance for advice on computational methods and mathematical analysis; and Casey Bergman and Helene Auger for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: MH YJ LC JSJ. Performed the experiments: MH YJ LC. Analyzed the data: MH YJ LC JSJ. Wrote the paper: MH YJ LC JSJ.

References

1. Garcia-Bellido A (1975) Genetic control of wing disc development in Drosophila. Ciba Found Symp 0: 161–182.
2. Hobert O (2008) Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. Proc Natl Acad Sci U S A 105: 20067–20071.
3. Blackshaw S, Fraioli RE, Furukawa T, Cepko CL (2001) Comprehensive analysis of photoreceptor gene expression and the identification of candidate retinal disease genes. Cell 107: 579–589.
4. Chen S, Wang QL, Nie Z, Sun H, Lennon G, et al. (1997) Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. Neuron 19: 1017–1030.
5. Hsiau TH, Diaconu C, Myers CA, Lee J, Cepko CL, et al. (2007) The cis-regulatory logic of the mammalian photoreceptor transcriptional network. PLoS ONE 2: e643. doi:10.1371/journal.pone.0000643.
6. Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8: 450–461.
7. Koike C, Nishida A, Ueno S, Saito H, Sanuki R, et al. (2007) Functional roles of Otx2 transcription factor in postnatal mouse retinal development. Mol Cell Biol 27: 8318–8329.
8. Nishida A, Furukawa A, Koike C, Tano Y, Aizawa S, et al. (2003) Otx2 homeobox gene controls retinal photoreceptor cell fate and pineal gland development. Nat Neurosci 6: 1255–1263.
9. Ranade SS, Yang-Zhou D, Kong SW, McDonald EC, Cook TA, et al. (2008) Analysis of the Otd-dependent transcriptome supports the evolutionary conservation of CRX/OTX/OTD functions in flies and vertebrates. Dev Biol 315: 521–534.
10. Tahayato A, Sonneville R, Pichaud F, Wernet MF, Papatsenko D, et al. (2003) Otd/Crx, a dual regulator for the specification of ommatidia subtypes in the Drosophila retina. Dev Cell 5: 391–402.
11. McDonald JA, Fujioka M, Odden JP, Jaynes JB, Doe CQ (2003) Specification of motoneuron fate in Drosophila: integration of positive and negative transcription factor inputs by a minimal eve enhancer. J Neurobiol 57: 193–203.
12. Wenick AS, Hobert O (2004) Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in C. elegans. Dev Cell 6: 757–770.
13. Hong JW, Hendrix DA, Papatsenko D, Levine MS (2008) How the Dorsal gradient works: insights from postgenome technologies. Proc Natl Acad Sci U S A 105: 20072–20076.
14. Zinzen RP, Cande J, Ronshaugen M, Papatsenko D, Levine M (2006) Evolution of the ventral midline in insect embryos. Dev Cell 11: 895–902.
15. Denes AS, Jekely G, Steinmetz PR, Raible F, Snyman H, et al. (2007) Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. Cell 129: 277–288.

16. Chiori R, Jager M, Denker E, Wincker P, Da Silva C, et al. (2009) Are Hox genes ancestrally involved in axial patterning? Evidence from the hydrozoan Clytia hemisphaerica (Cnidaria). PLoS ONE 4: e4231. doi:10.1371/journal.pone.0004231.
17. Davidson E (2006) The regulatory genome. San Diego: Academic.
18. Lowe CJ, Wu M, Salic A, Evans L, Lander E, et al. (2003) Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. Cell 113: 853–865.
19. Bruce AE, Shankland M (1998) Expression of the head gene Lox22-Otx in the leech Helobdella and the origin of the bilaterian body plan. Developmental biology 201: 101–112.
20. Hudson C, Lemaire P (2001) Induction of anterior neural fates in the ascidian Ciona intestinalis. Mechanisms of development 100: 189–203.
21. Williams N, Holland P (1996) Old head on young shoulders. Nature 383: 490–490.
22. Schlosser G (2006) Induction and specification of cranial placodes. Developmental Biology 294: 303–351.
23. Acampora D, Mazan S, Lallemand Y, Avantaggiato V, Maury M, et al. (1995) Forebrain and midbrain regions are deleted in Otx2−/− mutants due to a defective anterior neuroectoderm specification during gastrulation. Development (Cambridge, England) 121: 3279–3290.
24. Acampora D, Gulisano M, Broccoli V, Simeone A (2001) Otx genes in brain morphogenesis. Progress in neurobiology 64: 69–95.
25. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, et al. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell 133: 1266–1276.
26. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, et al. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell 133: 1277–1289.
27. Hanes SD, Brent R (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. Science (New York, NY) 251: 426–430.
28. Acampora D, Avantaggiato V, Tuorto F, Barone P, Perera M, et al. (1999) Differential transcriptional control as the major molecular event in generating Otx1−/− and Otx2−/− divergent phenotypes. Development (Cambridge, England) 126: 1417–1426.
29. Acampora D, Avantaggiato V, Tuorto F, Barone P, Reichert H, et al. (1998) Murine Otx1 and Drosophila otd genes share conserved genetic functions required in invertebrate and vertebrate brain development. Development (Cambridge, England) 125: 1691–1702.
30. Adachi Y, Nagao T, Saiga H, Furukubo-Tokunaga K (2001) Cross-phylum regulatory potential of the ascidian Otx gene in brain development in Drosophila melanogaster. Dev Genes Evol 211: 269–280.
31. Wada H, Saiga H, Satoh N, Holland PW (1998) Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian Pax-2/5/8, Hox and Otx genes. Development 125: 1113–1122.
32. Satoh N, Levine M (2005) Surfing with the tunicates into the post-genome era. Genes Dev 19: 2407–2411.
33. Imai KS, Hino K, Yagi K, Satoh N, Satou Y (2004) Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131: 4047–4058.
34. Satou Y, Takatori N, Yamada L, Mochizuki Y, Hamaguchi M, et al. (2001) Gene expression profiles in Ciona intestinalis tailbud embryos. Development 128: 2893–2904.
35. Tassy O, Daian F, Hudson C, Bertrand V, Lemaire P (2006) A quantitative approach to the study of cell shapes and interactions during early chordate embryogenesis. Curr Biol 16: 345–358.
36. Christiaen L, Bourrat F, Joly JS (2005) A modular cis-regulatory system controls isoform-specific pitx expression in ascidian stomodaeum. Dev Biol 277: 557–566.
37. Boorman CJ, Shimeld SM (2002) Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol Dev 4: 354–365.
38. Christiaen L, Burighel P, Smith WC, Vernier P, Bourrat F, et al. (2002) Pitx genes in Tunicates provide new molecular insight into the evolutionary origin of pituitary. Gene 287: 107–113.
39. Bertrand V, Hudson C, Caillol D, Popovici C, Lemaire P (2003) Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. Cell 115: 615–627.
40. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, et al. (2008) JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res 36: D102–106.
41. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31: 374–378.
42. Lebrecht D, Foehr M, Smith E, Lopes FJ, Vanario-Alonso CE, et al. (2005) Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 102: 13176–13181.
43. Arnone MI, Davidson EH (1997) The hardwiring of development: organization and function of genomic regulatory systems. Development 124: 1851–1864.
44. Fu D, Zhao C, Ma J (2003) Enhancer sequences influence the role of the amino-terminal domain of bicoid in transcription. Molecular and cellular biology 23: 4439–4448.
45. Ma X, Yuan D, Diepold K, Scarborough T, Ma J (1996) The Drosophila morphogenetic protein Bicoid binds DNA cooperatively. Development 122: 1195–1206.
46. Hare EE, Peterson BK, Eisen MB (2008) A careful look at binding site reorganization in the even-skipped enhancers of Drosophila and sepsids. PLoS Genet 4: e1000268. doi:10.1371/journal.pgen.1000268.
47. Ondek B, Gloss L, Herr W (1988) The SV40 enhancer contains two distinct levels of organization. Nature 333: 40–45.
48. Friedman AM, Fischmann TO, Steitz TA (1995) Crystal structure of lac repressor core tetramer and its implications for DNA looping. Science 268: 1721–1727.
49. Wada S, Sudou N, Saiga H (2004) Roles of Hroth, the ascidian otx gene, in the differentiation of the brain (sensory vesicle) and anterior trunk epidermis in the larval development of Halocynthia roretzi. Mechanisms of development 121: 463–474.
50. Dufour HD, Chettouh Z, Deyts C, de Rosa R, Goridis C, et al. (2006) Precraniate origin of cranial motoneurons. Proc Natl Acad Sci U S A 103: 8727–8732.
51. Imai KS, Satoh N, Satou Y (2002) Region specific gene expressions in the central nervous system of the ascidian embryo. Mech Dev 119 Suppl 1: S275–277.
52. Pennacchio L, Ahituv N, Moses A, Prabhakar S, Nobrega M, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature.
53. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 3: e7. doi:10.1371/journal.pbio.0030007.
54. Kent WJ, 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 U S A 100: 11484–11489.
55. Yoseph Barash GB, Nir Friedman. Algorithms in Bioinformatics, First International Workshop, WABI 2001, Aarhus, Denmark, August 28–31, 2001, Proceedings. In: Moret OGaBME, ed. Lecture Notes in Computer Science; 2001 2001; Aarhus, Denmark. Springer. pp 278–293.
56. Corbo JC, Levine M, Zeller RW (1997) Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. Development 124: 589–602.
57. Harafuji N, Keys DN, Levine M (2002) Genome-wide identification of tissue-specific enhancers in the Ciona tadpole. Proc Natl Acad Sci U S A 99: 6802–6805.
58. Philippakis A, Busser B, Gisselbrecht S, He F, Estrada B, et al. (2006) Expression-Guided In Silico Evaluation of Candidate Cis Regulatory Codes for Drosophila Muscle Founder Cells. PLoS Comput Biol 2: e53. doi:10.1371/journal.pcbi.0020053.
59. Halfon MS, Grad Y, Church GM, Michelson AM (2002) Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. Genome Res 12: 1019–1028.
60. Christiaen L, Jaszczyszyn Y, Kerfant M, Kano S, Thermes V, et al. (2007) Evolutionary modification of mouth position in deuterostomes. Semin Cell Dev Biol 18: 502–511.
61. Hotta K, Mitsuhara K, Takahashi H, Inaba K, Oka K, et al. (2007) A web-based interactive developmental table for the ascidian Ciona intestinalis, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. Dev Dyn 236: 1790–1805.
62. Newburger DE, Bulyk ML (2009) UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res 37: D77–82.
63. Roure A, Rothbacher U, Robin F, Kalmar E, Ferone G, et al. (2007) A multicassette Gateway vector set for high throughput and comparative analyses in ciona and vertebrate embryos. PLoS ONE 2: e916. doi:10.1371/journal.pone.0000916.