bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Evolution of embryonic *cis*-regulatory landscapes between divergent *Phallusia* and *Ciona* ascidians.

Alicia Madgwick[1], Marta Silvia Magri[2$], Christelle Dantec[1$], Damien Gailly[1], Ulla-Maj Fiuza[1*], Léo Guignard[3], Sabrina Hettinger[1], Jose Luis Gomez-Skarmeta[2], and Patrick Lemaire[1#]

1) Centre de Recherche en Biologie cellulaire de Montpellier (CRBM), Université de Montpellier, CNRS, Montpellier, France.

2) Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide/Junta de Andalucía, Sevilla, Spain

3) Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix drive, Ashburn, VA, USA

*: Current address: European Molecular Biology Laboratory, Cell Biology and Biophysics Unit,

$: Equal contribution

#: Author for correspondence

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ABSTRACT

Ascidian species of the *Phallusia* and *Ciona* genera are distantly related, their last common ancestor dating several hundred million years ago. Although their genome sequences have extensively diverged since this radiation, *Phallusia* and *Ciona* species share almost identical early morphogenesis and stereotyped cell lineages.

Here, we explored the evolution of transcriptional control between *P. mammillata* and *C. robusta*. We combined genome-wide mapping of open chromatin regions in both species with a comparative analysis of the regulatory sequences of a test set of 10 pairs of orthologous early regulatory genes with conserved expression patterns.

We find that ascidian chromatin accessibility landscapes obey similar rules as in other metazoa. Open-chromatin regions are short, highly conserved within each genus and cluster around regulatory genes. In most cases, chromatin accessibility is not temporally regulated during early embryogenesis. Open-chromatin regions are highly enriched in *cis*-regulatory elements: 73% of 49 open chromatin regions around our test genes behaved as either distal enhancers or proximal enhancer/promoters following electroporation in *Phallusia* eggs. Analysis of this datasets suggests a pervasive use in ascidians of "shadow" enhancers with partially overlapping activities. Cross-species electroporations point to a deep conservation of the *trans*-regulatory logic between these distantly-related ascidians and. Finally, we found that the relative order and approximate distance to the transcription start site of open chromatin regions can be conserved between *Ciona* and *Phallusia* species despite extensive sequence divergence, a property that can be used to identify orthologous enhancers, whose regulatory activity can partially diverge.

KEYWORDS

Ascidians; embryo; chromatin; enhancer; transcription; evolution

HIGHLIGHTS

- Open chromatin regions are highly enriched for *cis*-regulatory elements in ascidian early embryos.
- Ascidian Clusters of Open Regulatory Elements (COREs) are found next to regulatory genes.
- Regulatory gene expression, the *trans*-regulatory code, is conserved between distantly-related *Ciona* and *Phallusia* ascidians.
- Ascidians show pervasive use of "shadow" enhancers
- The position of homologous enhancers can be conserved, despite extensive genome divergence.

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

INTRODUCTION

Development and morphogenesis are precisely orchestrated by complex Gene Regulatory Networks. These networks combine transcription factors, and the *cis*-regulatory sequences to which they bind to control gene expression. Morphological change during evolution is frequently caused by variations in the transcriptional programme, and in particular in *cis*-regulatory sequences (Stern and Orgogozo, 2008). The converse is, however, not always true. There are examples of divergent gene regulatory networks producing very similar phenotypic outputs, a scenario referred to as Developmental Systems Drift or DSD (True and Haag, 2001), exemplified by vulva specification in nematodes (Félix and Barkoulas, 2012; Sommer, 2012) or heart morphogenesis in ascidians (Stolfi et al., 2014). Our understanding of the complex relationships between genotype and phenotype currently remains too fragmentary to predict the phenotypic outcome of a regulatory mutation.

Ascidians are marine invertebrate chordates belonging to the vertebrate sister group, the tunicates (Lemaire, 2011). Like vertebrates, their embryos develop through a tadpole larval stage. Unlike vertebrates, however, ascidian embryonic development is highly stereotyped, and proceeds with invariant cell lineages. Each cell can be individually named and found across all embryos of a given species and the number, names and location of its progeny is also precisely defined (Guignard et al., 2018). Strikingly, even distantly related species, which diverged more than 300 million years ago share very similar or identical cell lineages. This provides a rigorous framework to compare, with cellular resolution, the evolution of developmental programmes between distantly species. In light of this exceptional morphogenetic conservation, ascidian genomes are surprisingly divergent (Brozovic et al., 2016; Stolfi et al., 2014; Tsagkogeorga et al., 2012). Both coding and non-coding sequences evolve rapidly within the group. As a result, while non-coding sequence conservation within a genus (e. g. between *Ciona robusta* and *Ciona savignyi*, or between *Phallusia mammillata* and *Phallusia fumigata*) can be used as hallmark of functional importance, non-coding sequences across genera cannot be aligned. In addition, gene orders are frequently scrambled along chromosomes and even highly conserved gene clusters, such as the *Hox* cluster, are exploded (Duboule, 2007). Ascidians thus constitute a very interesting taxon to study DSD.

Cis-regulatory sequences are short (<1kb) non-coding DNA segments, which act as binding platforms for transcription factors, and control gene expression (Yáñez-Cuna et al., 2013). These sequences are so divergent between distant ascidian species that the sequences of homologous elements can generally not be aligned anymore (Oda-Ishii et al., 2005; Roure et al., 2014; Stolfi et al., 2014). In some cases, highly divergent regulatory sequences have retained the ability to respond to the same transcription factor combination (Oda-Ishii et al., 2005; Roure et al., 2014), the accumulation of *cis*-regulatory

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

mutations thus leaving the GRN architecture intact. In other cases, however, GRN architecture have changed sufficiently for functional *cis*-regulatory sequences from one species to become "unintelligible" to another species (Lowe and Stolfi, 2018; Stolfi et al., 2014; Takahashi et al., 1999). The conservation of the function of *cis*-regulatory sequences between distantly related ascidians has been assayed in too few cases to estimate the relative frequency of conservation vs divergence of *cis*-regulatory activity. The relative contributions of local TF binding site turn over versus more global GRN rewiring between ascidian species remains unknown.

In this study, we first used ATAC-seq (Buenrostro et al., 2013) to characterize chromatin accessibility landscapes in two distant phlebobranch species that diverged around 300 million years ago (Figure 3A) (Delsuc et al., 2018), *Phallusia mammillata* and *Ciona robusta*. We then used this dataset to identify in both species the early embryonic *cis*-regulatory sequences controlling the expression of ten orthologous pairs of regulatory genes. These analyses brought to light a surprising level of conservation of chromatin accessibility landscapes and *cis*-regulatory logic between these species despite extreme sequence divergence.

MATERIAL AND METHODS

Animal origin and *in vitro* fertilization of *Phallusia* and *Ciona*

Phallusia mammillata and *Ciona intestinalis* (previously *Ciona intestinalis* type B; Caputi et al., 2007) were provided by the Centre de Ressources Biologiques Marines in Roscoff. *Ciona intestinalis* dechorionation was done on fertilised eggs as previously described (Christiaen et al., 2009a). *Phallusia mammillata* unfertilised eggs were dechorionated in 0,1% trypsin in ASW with gentle shaking for 2 hours (Sardet et al., 2011). All live experiments were performed with dechorionated embryos.

Gene models (names, unique gene IDs and genome assemblies)

The *Ciona robusta* genes analysed in this project were from the KH assembly (Satou et al., 2008), represented by the following KH2012 gene models (unique ANISEED gene IDs): *Foxa.a,* KH.C11.313 (Cirobu.g00002136)*; Foxd,* KH.C8.396 (Cirobu.g00009025) and KH.C8.890 (Cirobu.g00009568); *Nodal,* KH.L106.16 (Cirobu.g00010576); *Nkx2-1/4,* KH.C10.338 (Cirobu.g00001550); *Lhx3/4* KH.S215.4 (Cirobu.g00014215); *Otx,* KH.C4.84 (Cirobu.g00006940); *Eph1,* KH.C1.404 (Cirobu.g00000642); *Tolloid,* KH.C12.156 (Cirobu.g00002684); *Lefty,* KH.C3.411 (Cirobu.g00005590); *Fgf9/16/*20, KH.C2.125 (Cirobu.g00004295).

For *Phallusia mammillata*, the gene models (unique ANISEED gene IDs) were from the MTP2014 assembly (Brozovic et al., 2016):

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Foxa.a Phmamm.CG.MTP2014.S50.g01891 (Phmamm.g00001891),
Foxd Phmamm.CG.MTP2014.S253.g06179 (Phmamm.g00006179),
Nodal Phmamm.CG.MTP2014.S1711.g15500 (Phmamm.g00015500),
Nkx2-1/4 Phmamm.CG.MTP2014.S597.g10419 (Phmamm.g00010419),
Lhx3/4 Phmamm.CG.MTP2014.S2332.g16546 (Phmamm.g00016546),
Otx Phmamm.CG.MTP2014.S166.g04580 (Phmamm.g00004580),
Eph1 Phmamm.CG.MTP2014.S160.g04451 (Phmamm.g00004451),
Tolloid Phmamm.CG.MTP2014.S14.g00600 (Phmamm.g00000600),
Lefty Phmamm.CG.MTP2014.S362.g07801 (Phmamm.g00007801),
*Fgf9/16/*20 Phmamm.CG.MTP2014.S128.g03805 (Phmamm.g00003805),

In situ hybridisation

Wholemount *in situ* hybridisation was performed as previously described (Christiaen et al., 2009b). Dig-labelled probes for *Phallusia mammillata* were synthesised from the following cDNAs: *Otx* (AHC0AAA196YH07), *Lefty* (AHC0AAA31YI04), *Nkx2-1/4* (AHC0AAA267YK08), *Tolloid* (AHC0AAA14YA03), *Fgf9/16/20* (AHC0AAA22YD05), *Foxa.a* (AHC0AAA74YF05), *Foxd* (AHC0AAA228YM13), Lhx3/4 (AHC0AAA183YA13). The *Phallusia Nodal* and *Eph1* probes were generated by PCR from cDNA using the following primers: *Nodal*, Forward: 5'-CTATGGATATGACACAAGTATCGTTCTGC-3', Reverse: 5'-GATCCTAATACGACTCACTATAGGGTTATCGACATCCACATTCT-3'; *Eph1*, Forward: 5'-CCAACGTTGCGACTCCACTTTCACC-3', Reverse: 5'-ACGATAATACGACTCACTATAGGGGTCTTGGTTAGACATCTCCCAG-3';

ATAC-seq assays, sequence pre-processing, alignment and peak calling

ATAC-seq assays were carried out essentially as described (Buenrostro et al., 2013b) starting with 50-70 thousand embryonic cells. The embryos were not dissociated prior to nucleus extraction. Sequencing was done using an Illumina HiSeq4000 sequencer (BGI) and around 20 million sequence reads were obtained per sample in each species. Illumina adapters were trimmed by Cutadapt (Martin, 2011). Alignment on the *Phallusia mammillata* and *Ciona robusta* genomes was performed with bowtie2, (very-sensitive parameter) (Langmead and Salzberg, 2012). Only a minority of sequences aligned to the nuclear genome (*Ciona* ~30%; *Phallusia* ~17%). Significant Peaks were extracted with MACS2 (Zhang et al., 2008) with the following parameters for *Phallusia mammillata*: gsize=187138625 (genome size = genome mappable size excluding repeated sequences) --nomodel -f BAMPE -q 0.01. For *Ciona*, peaks were extracted with MACS2 with the following parameters for *Ciona*: gsize= 99000000 --nomodel -- -f BAMPE -q 0.01. *Ciona* and *Phallusia* raw ATAC-seq sequences were submitted to the Short Read Archive

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(Bioproject IDs PRJNA474750 and PRJNA474983). Peaks and short read coverage can be seen on the ANISEED WashU browser: https://www.aniseed.cnrs.fr/browser/

Gateway cloning and electroporation of the regulatory sequences

Constructs to test the activity of the regulatory sequences from both species were generated by Gateway cloning into pSP1.72.RfB::NLS-LacZ destination vector (Lamy et al., 2006), or in modified versions, in which the *Foxa.a* minimal promoter from *Ciona robusta* (ANISEED ID: REG00001035) was placed upstream of NLS-LacZ to generate pSP1.72.RfB.pFoxa.a::NLS-LacZ, or into pSP1.72.RfB.pFog::NLS-LacZ as indicated in Table S1. The primers for each of the non-coding sequences, and their spatio-temporal patterns of activity and reference in the ANISEED database are listed in Table S1. The genomic DNA used to amplify by PCR the genomic fragments from *Ciona robusta* was a kind gift from Dr. A. Spagnuolo (Naples). *Phallusia mammillata* genomic DNA was extracted from Roscoff animals.

Electroporation was done as previously described (Bertrand et al., 2003), with a single pulse of 16ms at 50V for *Ciona intestinalis* and two pulses of 10ms at 40V for *Phallusia mammillata*. X-gal staining was performed as previously described (Bertrand et al., 2003). Sequences that drove activity of the LacZ reporter gene in fewer than 10% of electroporated embryos or larvae, or only in the mesenchyme (known to be a tissue displaying ectopic staining for many sequences) were considered inactive. More than 100 embryos were counted for each construct and experiment.

RESULTS AND DISCUSSION

Regions of open chromatin in ascidian early embryos are enriched in *cis*-regulatory sequences.

To characterize the *cis*-regulatory landscapes during early ascidian embryogenesis, we first mapped open chromatin regions using ATAC-seq (Buenrostro et al., 2013b) on whole *Ciona* and *Phallusia* embryos at the 64- and 112-cell stages and identified regions of open chromatin using the MACS2 peak caller (Zhang et al., 2008). Comparison of the biological replicas indicated that the process was highly reproducible (Figure 1A).

In *Ciona*, MACS2 identified 24574 (median length 119 bp) and 19397 (median length 189 bp) non-coding ATAC-seq peaks at the 64- and 112-cell stages, respectively, covering 3,32% and 4,38% of the genome sequences. As expected, most (70%) open chromatin regions were located in regions of high (>60%) sequence identity between *C. robusta* and *C. savignyi* estimated to have diverged around 120 million years ago (Delsuc et al., 2018) (Figure 1A, B). Conversely, only around 10% of non-coding sequences with more than 60% of identity between *C. robusta* and *C. savignyi* overlaped with open

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

chromatin regions during early development (Figure 1B). The majority of non-coding conserved sequences may thus correspond to *cis*-regulatory sequences acting much later during development or in adults. In *Phallusia*, similarly, MACS2 identified 10632 (median length 219 bp) and 14240 (median length 123 bp) non-coding ATAC-seq peaks, covering 1.27% and 0.93% of the assembled genome (Brozovic et al., 2016) at the 64- and 112-cell stages, respectively. In both *C. robusta* and *P. mammillata*, the GC content of ATAC-seq peaks was higher than that of other non-coding sequences and exonic sequences (Figure 1C), consistent with previous findings on *cis*-regulatory sequences in other model organisms (Li et al., 2007).

We conclude that, as in other animal taxa, open chromatin regions detected by ATAC-seq during early ascidian embryogenesis cover a small part of the non-coding genome, enriched in GC and highly-conserved within the *Ciona* genus. These ATAC-seq datasets can be explored through a public hub in the ANISEED WashU browser (https://www.aniseed.cnrs.fr/browser) and have been deposited in the NCBI Short Read Archive (Bioprojects PRJNA474750 and PRJNA474983)

Temporal dynamics of chromatin accessibility in *Phallusia mammillata* and its relationship to gene expression.

We next assessed the temporal dynamics of chromatin accessibility in *Phallusia mammillata*. We extended our *Phallusia* dataset to encompass the 16-cell, 32-cell, 64-cell, 112-cell, mid-gastrula and mid-neurula stages. We merged overlapping peaks at the different stages to create a non-redundant reference set of 11302 non-coding peaks; without overlap with annotated repeated sequences. Only 23% of these peaks showed a strong (>10-fold change) dynamics between the 32-cell and mid-neurula stages (13% monotonically increasing, 10% decreasing). Half of the peaks showed less than a 3-fold variation and were considered constitutive. These results are in keeping with *Drosophila*, *C. elegans* and vertebrates in which a minority of loci show a temporally dynamic chromatin accessibility in embryos (Cusanovich et al., 2018; Daugherty et al., 2017; Hagey et al., 2016; Liu et al., 2018; Preissl et al., 2018; Thomas et al., 2011; Wu et al., 2016).

To estimate whether the temporal dynamics of chromatin accessibility was correlated to the temporal expression control of neighbouring genes, we used published time course of RNA-seq expression levels (Brozovic et al., 2018) to survey the expression of genes around the highest 50 ATAC-seq peaks at the mid-neurula stage that also showed a strong (>10 fold) increase in chromatin accessibility between the 32-cell and mid-neurula stages. 74% of these peaks were located in an intron or in the intergenic regions of a gene with dynamic upregulation during early embryogenesis, as exemplified on Figure 1D for a Matrilin-related gene. Conversely, of the 85 most-expressed genes at the mid-neurula stage, with

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

at least a 10-fold higher expression compared to early gastrulae, only 38 were associated to temporally dynamic chromatin accessibility peaks (not shown).

There is, thus, only a modest correlation between the temporal dynamics of chromatin accessibility and gene expression. This suggests that chromatin accessibility is often a permissive feature and therefore that specific temporal and spatial gene expression depends on the expression of the entire set of TFs activating the enhancer rather than being determined by the physical accessibility of enhancer DNA.

Ascidian regulatory genes are controlled by clusters of *cis*-regulatory elements (COREs).

In vertebrates, open regulatory elements are unevenly distributed along the chromosomes. They show a higher density around developmental regulators, defining Clusters of Open *cis*-Regulatory Elements (COREs) (Gaulton et al., 2010), which have been associated to the super-enhancers controlling the expression of genes controlling cell fate decisions (Pott and Lieb, 2015).

We found likewise that non-coding ATAC-seq peaks were unevenly distributed along the genomes of both ascidian species, with clustered peaks separated by regions of much lower density (Figure 2A). 33% and 65% of *Phallusia* and *Ciona* ATAC-seq peaks belonged to 821 and 2099 clusters of at least 5 ATAC-seq peaks within 10kb, respectively. Genes overlapping with or immediately adjacent to such clusters were highly enriched for molecular functions linked to transcription and signalling and for biological processes linked to development, regulation, transcription and signalling (Figure 2B).

Thus, ascidian embryos, despite their simpler anatomy, reduced cell numbers and compact genomes share with vertebrates the presence of conserved clusters of open chromatin regulatory elements (COREs) around developmental regulatory genes. In the following sections, we will define and use a set of ten pairs of orthologous early regulatory genes to assess in more detail the *cis*-regulatory potential of open chromatin regions and their evolutionary conservation.

Conservation of early regulatory gene expression between *Ciona robusta* and *Phallusia mammillata* up to the early gastrula stage

To define a test set of orthologous regulatory genes active in *Ciona* and *Phallusia* early embryos, we chose to focus on genes expressed in endodermal precursors up to the 112-cell stage, by which point the endoderm tissue fate has been determined and gastrulation is initiated.

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ciona robusta (formerly *Ciona intestinalis* type A; Caputi et al., 2007) expression profiles of most transcription factors and signalling molecules have previously been characterised by *in situ* hybridisation (Imai et al., 2006, 2004). We mined the ANISEED database (Brozovic et al., 2016), which catalogues expression data from published work, for regulatory genes expressed zygotically throughout all *Ciona* endoderm precursor cells for at least one of the following stages: 16-, 32-, 64- or 112-cell stages. These genes include five known regulators of *Ciona* endoderm development *Foxa-a*, *Foxd* and *Fgf9/16/20, Lhx3/4* and *Nkx2-1/4* (Hudson et al., 2016; Ristoratore et al., 1999; Satou et al., 2001) and 5 regulatory genes expressed in endoderm precursors but whose function in endoderm development remains hypothetical, *Nodal, Otx*, *Tolloid*, *Lefty*, and *Eph1*.

Using the ANISEED orthology section, we identified the *Phallusia mammillata* orthologues of these genes. Unambiguous 1-to-1 orthologues for nine of these genes can be identified in *Phallusia* (*Foxa.a*, *Nodal*, *Nkx2-1/4*, *Lhx3/4*, *Otx*, *Tolloid*, *Lefty*, *FGF9/16/20*, *Eph1*) (Brozovic et al., 2018). The tenth gene, *Foxd*, is a unique gene in *Phallusia mammillata* and has undergone a recent duplication in *Ciona* (Brozovic et al., 2018). The two *Ciona* paralogs are however so similar (96% identity at the nucleotide level) that the published *in situ* hybridisation profiles may reflect the combined expression of both genes.

Next, we described by *in situ* hybridization the spatial and temporal expression patterns of the ten *Phallusia* orthologs up to the early gastrula stage (Figure 3C) and compared them to their individual *Ciona* orthologs at the equivalent stages (Figure 3D). Thanks to the conservation of early cell lineages between *Ciona* and *Phallusia*, the expression patterns of orthologs could be compared with a cellular level of resolution.

This set of *in situ* hybridisation experiments may overestimate divergence as their interpretation depends on the duration of the alkaline phosphatase enzymatic staining (Figure 3B). For instance, at the 64-cell stage, a short staining time with an *Fgf9/16/20* probe in *Phallusia mammillata* detects expression in the cells where the *Ciona Fgf9/16/20* ortholog was previously described (A7.4, A7.8 and B7.4) (Imai et al., 2004), while a longer staining detected additional weak *Phallusia Fgf9/16/20* expression in A7.1, A7.2, A7.5 and A7.6, in which no expression was reported in *Ciona*. A similar situation was also found in the case of *Lefty* at the 64-cell stage.

Despite this potential caveat, conservation of gene expression between the two species was overall high. It varied between perfect cell-to-cell conservation of the expression pattern (*Nkx2-1/4* and *Foxa.a*), minor divergence in up to 15% of expressing cells (*Otx*, *Tolloid*, *Nodal*, *Lhx3/4*, *Foxd*), and more significant divergence in 35 to 65% of expressing cells (*Eph1*, *FGF9/16/20*, *Lefty*), which may in part reflect the caveat mentioned above (Figure 3D).

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We conclude that the majority of early expression patterns of our ten test genes are conserved between *Ciona* and *Phallusia*, some of the observed differences possibly reflecting limitations of *in situ* hybridisation. Future work based on single-cell RNA seq in ascidians (Sharma et al., 2018; Treen et al., 2018) may be required to discriminate technical issues from *bona fide* evolutionary changes. In the following sections, we sought to identify the *cis*-regulatory sequences driving the conserved expression of these genes and to compare their activity in both species.

Most open chromatin regions have embryonic *cis*-regulatory activity

To identify the *cis*-regulatory sequences driving the expression of the ten ortholog pairs in both species, we first surveyed the chromatin state of known *Ciona* elements controlling the expression of four of these ten genes. Figure 4A illustrates that open chromatin regions around the *Ciona Otx* gene overlapped with previously characterized *cis*-regulatory regions and with ChIP-seq peaks for *GATAa,* a major regulator of *Otx* at the 32-cell stage. Four previously identified *Ciona robusta cis*-regulatory sequences controlling the *FGF9/16/20* (ANISEED REG00001080, Figure 4B, Oda-Ishii et al., 2016), *Foxd* (ANISEED REG00001115, Oda-Ishii et al., 2016), *Lefty* (ANISEED REG00001118, Oda-Ishii et al., 2016), *Nkx2-1/4* (ANISEED REG00000321, Fanelli et al., 2003) and *Otx* (ANISEED REG00000010, Bertrand et al., 2003a) genes in early mesendodermal territories also mapped to open chromatin regions. Our ATAC-seq dataset may thus constitute a valuable resource to identify early embryonic *cis*-regulatory sequences.

For each of our test genes, we selected 2 to 7 candidate *cis*-regulatory regions ranging from 209bp to 1316bp (average 525bp) and located up to 4kb upstream from the transcription start site of the gene, within the 5' UTR, in introns, or immediately downstream of the transcription start site (Figure 5). In most cases, selected regions corresponded to a single peak. Candidate *cis*-regulatory sequences were cloned upstream of a minimal promoter (Table S1) driving NLS-LacZ, to test for enhancer activity. In addition, sequences located immediately upstream of the transcription start site, often coinciding with the H3K4me3 mark, were also cloned upstream of the NLS-LacZ reporter, without exogenous minimal promoter, to test for combined enhancer/promoter activity. LacZ activity was monitored at three stages of development: early gastrula (EG), late gastrula (LG) and larva.

In total, we tested the *cis*-regulatory activity of 37 *Phallusia* and 12 *Ciona* candidate open regulatory regions, as well as 11 *Ciona* and *Phallusia* sequences, which were not in an open chromatin state at the 64- and 112-cell stages (Table S2). As an example, the analysis of the *cis*-regulatory sequences controlling the *Phallusia Nodal* gene is presented in Figure 4C. Out of the 49 candidate regulatory sequences we electroporated, 73% drove LacZ activity by the larval stage or earlier (Figure 5, blue,

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

green and pink squares), including 16 sequences active at or before the gastrula (Figure 5, blue, green and pink squares outlined by a black line). By contrast, only 3 of 11 closed chromatin sequences (Figure 5, squares outlined with a red line) were active by the larval stage and none of these were active by the gastrula stage. The percentage of open-chromatin regions with *cis*-regulatory activity is thus much larger than obtained in previous *cis*-regulatory sequence screens, in which larger (mean 1.7 kb) sequences were chosen at random (Harafuji et al., 2002; Keys et al., 2005).

Interestingly, the timing of onset of activity of an element was related to its position with respect to the transcription start site of the regulated gene. In each of our test genes in both species, an active sequence was located within 500 bp of the transcription start site, identified either by H3K4me3 promoter mark or RNA-seq signal. In at least 15 out of 19 cases, these proximal sequences were active by the early gastrula stage (outlined with a black line on Figure 5). By contrast, only one intronic and 5 distal enhancers, out of a total of 21 active intronic, distal or downstream elements tested at the early gastrula stage, had an early onset of activity (Figure 5).

We conclude that as in other animal taxa, open chromatin regions detected by ATAC-seq in ascidians are strongly enriched for embryonic *cis*-regulatory elements. The resource we generated should thus be of general interest to the community. We also found that ascidian elements with early onset of activity are biased towards a proximal position with respect to the transcription start site and behave as promoter/enhancer elements with spatially restricted activity. It will be interesting to survey whether similar positional bias is also observed in the early embryonic enhancers of other animal species.

Overlapping temporal and spatial enhancer activity contributes to regulatory landscape complexity.

Considering the small number of embryonic territories in ascidian embryos, the *cis*-regulatory landscape in the vicinity of regulatory genes, marked by the presence of COREs, is surprisingly complex. Recent studies in *Drosophila* and vertebrates have revealed that regulatory gene expression is frequently controlled by multiple "shadow" *cis*-regulatory elements with overlapping spatial activities (Barolo, 2012; Cannavò et al., 2016). This partial redundancy was shown to confer robustness of transcriptional regulation to genetic and environmental variations (Frankel et al., 2010; Perry et al., 2010). The *Ciona* genus includes some of the most polymorphic metazoan species known (Leffler et al., 2012) and *Phallusia* and *Ciona* both live in shallow coastal environments subject to strong variations in temperature and salinity. Spatially overlapping enhancer activity may thus be beneficial to ascidians and underlie the unexpected complexity of their early *cis*-regulatory landscapes.

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Although a few "shadow" enhancers have been previously identified in *Ciona* (Farley et al., 2016; Fujiwara and Cañestro, 2018; Matsumoto et al., 2008, 2007), the prevalence of enhancers with overlapping activity is unknown. We thus used our collection of active *cis*-regulatory sequences to investigate the presence of regulatory elements with overlapping patterns of activity in the vicinity of our test gene set.

Of the ten genes, five (*Foxa.a*, *Foxd*, *FGF9/16/20*, *Otx* and *Eph1*) were controlled during *Phallusia mammillata* embryogenesis by 2 or 3 *cis*-regulatory sequences with partially overlapping spatial activity (Figure 6), a proportion in keeping with the *Drosophila* situation (Cannavò et al., 2016). By the larval stage (Figure 6A, B), three *Foxaa cis*-regulatory sequences (*Foxa.a*Pm01, *Foxa.a*Pm03 and *Foxa.a*Pm05) drove activity in the head endoderm and endodermal strand, three *Foxd* sequences (*Foxd*Pm03, *Foxd*Pm05 and *Foxd*Pm06) drove expression in the tail muscles and two *Fgf9/16/20* (*Fgf*Pm03, *Fgf*Pm05) sequences were active in tail muscle cells, B-line notochord and mesenchyme. Analysis of the 12 *Phallusia cis*-regulatory sequences with a pre-gastrula onset of activity identified three additional pairs of regulatory sequences with overlapping spatial and temporal activity (Figure 6C). *Fgf*Pm04 and *Fgf*Pm05 drive early expression in nerve cord precursors. *Otx*Pm01 and *Otx*Pm05 drive early expression in tail muscle and b-line neural tissue. Finally, *Eph1*Pm02 and *Eph1*Pm03 are both active in early endoderm precursors. Based on these examples, ascidian shadow enhancers can be separated by another enhancer (*Otx*Pm01 and OtxPm05) or even by a coding region (*Eph1*Pm02 and *Eph1*Pm03; *Foxd*Pm06 and *Foxd*Pm05).

These results suggest that ascidian embryos make frequent use of partially overlapping regulatory sequences, or shadow enhancers, to control their regulatory genes, which may contribute to the complexity of the open chromatin landscape we identified.

Gastrula *trans*-regulatory activity is temporally and spatially conserved across species.

Having extensively characterized the enhancers driving the expression of our ten test genes, we next addressed the evolutionary conservation of their activity between the two species. The evolutionary conservation of gene expression patterns can result from a global conservation of gene regulatory network (GRN) architecture, that is of the combination of transcription factors binding to each *cis*-regulatory sequence operating within the network. This scenario predicts conservation of the activity of the *cis*-regulatory sequences despite extensive sequence divergence, as observed in insects (Hare et al., 2008), and in a few *Ciona* and *Phallusia* enhancers (Roure et al., 2014). Alternatively, GRNs can be extensively rewired without major impact on regulatory gene expression (Ciliberti et al., 2007). In this scenario, distinct combinations of transcription factors bind to orthologous *cis*-regulatory sequences

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

causing them to functionally diverge beyond cross-species intelligibility, a situation observed between *Molgula* and *Ciona* species (Lowe and Stolfi, 2018; Stolfi et al., 2014) or between distantly-related echinoderms (Hinman and Davidson, 2007).

To determine which of these two scenarios accounts for the conservation of the early expression of our ten test genes, we individually electroporated 14 *Phallusia* and 3 *Ciona* early *cis*-regulatory sequences active by the gastrula stage into the eggs of the other species and compared the activity of these sequences in the two species at the gastrula stage. 12 of these 17 sequences drove the same patterns of LacZ activity in both species, with an additional 2 elements driving very similar activity (Figure 7A, B) in spite of having diverged beyond alignability between *Phallusia* and *Ciona*. Although we cannot formally exclude that conserved activity results from the binding of different transcription factors in the two species, a parsimonious explanation is that the expression patterns and biochemical activity of the TFs driving the activity of these sequences are conserved between *Ciona* and *Phallusia.* Substantial changes in the activity patterns of the remaining 3 sequences (Figure 7B, *Eph1*Pm02, *Lhx3/4*Ci01, and *Otx*Pm05) could reflect divergence in the specific *trans*-regulatory code driving the activity of these elements between the two species. We note, however, that 11/12 tissue differences in activity across the 17 regulatory sequences are due to the detection of additional activity in *Ciona* (Figure 7B). As electroporation in *Ciona* is more efficient than in *Phallusia* (not shown), these differences could in part reflect an experimental bias, rather than a true evolutionary divergence.

We conclude that while ascidian *trans*-regulatory landscapes may diverge significantly between the more divergent *Molgula* and *Ciona* ascidian species (Lowe and Stolfi, 2018; Stolfi et al., 2014), early vegetal *trans*-regulatory landscapes appear to be well conserved between the *Phallusia* and *Ciona* species we compared here.

Collinearity of *Phallusia* and *Ciona cis*-regulatory landscapes around the *FGF9/16/20* and *Otx* genes.

The previous section assayed whether the *Ciona* and *Phallusia trans*-regulatory codes are sufficiently close to interpret similarly a given enhancer. In this final section, we conversely addressed whether, despite their sequence divergence, orthologous *Ciona* and *Phallusia* enhancers drive similar activities in a given species, that is the extent of *cis*-regulatory conservation.

For this, we first needed to identify orthologous *Phallusia* and *Ciona* enhancers, a task made difficult by the non-alignability of intergenic and intronic sequences between these two genera. Previous work in insects suggested that the relative position of enhancers for regulatory genes may be under

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

evolutionary constraint, and may thus guide the identification of orthologous enhancers (Cande et al., 2009).

To address whether the global organization of *cis*-regulatory sequences between the two species was conserved, we analysed the genomic architecture of the 9 genes with 1-to1 orthology relationships between the two species. Global gene structure and intron length of orthologs was conserved but local syntenic relationships were frequently broken. Only two genes have conserved both upstream and downstream neighbour genes, *Fgf9/16/20* and *Tolloid*. Two more have retained either their upstream (*Otx*) or downstream (*Foxa.a*) neighbour. The remaining five genes (*Nkx2-1/4*, *Nodal*, *Lhx3/4*, *Lefty* and *Eph1*) have non-orthologous upstream and downstream neighbours in the two species.

Although the number of regulatory peaks around *Ciona* and *Phallusia* orthologs was remarkably well conserved (Figure 8A), the position of open chromatin regions was for most genes not directly comparable. Strikingly, however, the upstream and intronic open chromatin landscapes were remarkably conserved between *Ciona* and *Phallusia* for *FGF9/16/20* and *Otx* (Figure 8B), suggesting that orthologous enhancers can be identified despite the lack of sequence conservation between the two genera. To test this idea further, we compared the activity of three *Phallusia* elements to that of known *Ciona* regulatory sequences (Figure 8B). At the early gastrula stage, the activity of the proximal *Otx*Pm01 element was detected in the posterior mesoderm and b6.5 cell pair progeny. That of the distal *Otx*Pm04 element was detected in the endoderm. The other distal element, *Otx*Pm05, was active in most posterior mesendodermal lineages and in the progeny of the b6.5 and a6.5 cell pairs. These activities were consistent with the reported activities of matching ATAC-seq peaks in *Ciona* (Bertrand et al., 2003a). *Otx*Ci01, matching the position of *Otx*Pm01, drove activity in posterior mesoderm, posterior endoderm and b6.5 cell pair progeny. The *Ciona* region matching the position of *Otx*Pm04 was necessary for endoderm activity. Finally, the *Ciona* a-element (REG0000010), corresponding to part of the ATAC-seq peak matching *Otx*Pm05, drove activity in the progeny of the a6.5 and b6.5. Thus, open chromatin regions located in matching positions in the *Ciona* and *Phallusia* intergenic sequences of *Otx* appear to identify orthologous *cis*-regulatory elements.

This orthology did not preclude the evolution of the activity of the enhancers. For example, *Otx*Ci01 drives posterior endodermal expression in both *Ciona* and *Phallusia*, while *Otx*Pm01 does not show this activity in either species (Figure 8C). Thus, orthologous *FGF9/16/20* and *Otx* early enhancers have retained their relative order and approximate distance to the transcription start site, yet some at least have undergone *cis*-regulatory changes, which partially affected their spatial patterns of activity.

CONCLUSIONS

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In this study, we addressed how *Ciona* and *Phallusia* species, despite their extremely divergent genomes can undergo almost identical early developments. We established ATAC-seq in both species and showed that open chromatin landscapes obey similar rules as in other metazoa (Figures 1, 2): Non-coding open chromatin regions are enriched in GC, highly conserved within each genus, frequently show reduced temporal dynamics, cluster around developmental regulatory genes, mostly transcriptional regulators and are highly enriched in *cis*-regulatory sequences. This publicly-accessible ATAC-seq dataset will be of general interest within the ascidian community.

To go beyond a global analysis of regulatory landscapes, we selected ten early *Ciona robusta* regulatory genes with early endodermal expression. We identified the *Phallusia* orthologs and showed that their expression profiles up to the gastrula stages were highly consistent with the *Ciona* data (Figure 3). We identified 39 *cis*-regulatory sequences for these genes, including 17 with activity by the gastrula stage (Figures 4, 5), thanks to which we shed light on ascidian transcriptional regulation and its evolution.

First, in half of the genes we studied, we found 2 or more enhancers with overlapping spatial activity (Figure 6). This suggests that ascidians make frequent use of shadow enhancers with partially overlapping activity. Given the proposed function of shadow enhancers in ensuring developmental robustness to environmental and genetic stress (Frankel et al., 2010; Perry et al., 2010), it is not surprising that ascidians, which cope with intense environmental variations and a high level of polymorphism, would use this mechanism. The pervasive use of shadow enhancers in the control of regulatory genes may also contribute to the clustering of open chromatin regions in the vicinity of these genes.

Second, both the cross-genera conservation of the expression patterns of the transcriptional regulators we studied and of the similar activity of individual early *cis*-regulatory elements in cross-species electroporation assays argue for a conservation of *trans*-regulatory information between *Ciona* and *Phallusia*, that is a conservation of the combination of transcription factors active in individual embryonic cells (Figure 7). Our results imply that the ascidian *cis*-regulatory code is sufficiently flexible to allow extensive, phenotypically neutral, non-coding sequence divergence within *Phlebobranchia*, thereby limiting the need for compensatory *trans*-regulatory mutations causing Developmental System Drift (True and Haag, 2001). This contrasts with the divergence of *trans*-regulatory information between *Ciona* and the more divergent stolidobranchia *Molgula* (Lowe and Stolfi, 2018; Stolfi et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Finally, we presented evidence tha *cis*-regulatory landscapes can be conserved despite extensive sequence divergence (Figure 8). We described that the overall organisation of open chromatin regions at the *Otx* and *FGF9/16/20* loci is conserved between genera, despite the loss of non-coding sequence alignability. Analysis of the *cis*-regulatory activity of individual open-chromatin regions from both species around *Otx*, revealed that this conserved organization reflects the presence of orthologous enhancers whose activity, though overall conserved, can slightly diverge. The *cis*-regulatory redundancy ensured by the presence of enhancers with overlapping activity may facilitate this divergence, as the case of divergence we observed involved a posterior endoderm activity redundantly found in two other enhancers at the same locus. The availability of our embryonic ATAC-seq dataset will allow testing on larger datasets whether the evolutionary shuffling of individual redundant regulatory activities between enhancers at a given locus is a general feature of ascidians transcriptional regulation.

REFERENCES:

Barolo, S., 2012. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays 34, 135–141. https://doi.org/10.1002/bies.201100121

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. https://doi.org/10.1016/S0092-8674(03)00928-0

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., Guéroult-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, D718–D725. https://doi.org/10.1093/nar/gkx1108

Brozovic, M., Martin, C., Dantec, C., Dauga, D., Mendez, M., Simion, P., Percher, M., Laporte, B., Scornavacca, C., Di Gregorio, A., Fujiwara, S., Gineste, M., Lowe, E.K., Piette, J., Racioppi, C., Ristoratore, F., Sasakura, Y., Takatori, N., Brown, T.C., Delsuc, F., Douzery, E., Gissi, C., McDougall, A., Nishida, H., Sawada, H., Swalla, B.J., Yasuo, H., Lemaire, P., 2016. ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. Nucleic Acids Res 44, D808–D818. https://doi.org/10.1093/nar/gkv966

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods 10, 1213–1218. https://doi.org/10.1038/nmeth.2688

Cande, J., Goltsev, Y., Levine, M.S., 2009. Conservation of enhancer location in divergent insects. PNAS 106, 14414–14419. https://doi.org/10.1073/pnas.0905754106

Cannavò, E., Khoueiry, P., Garfield, D.A., Geeleher, P., Zichner, T., Gustafson, E.H., Ciglar, L., Korbel, J.O., Furlong, E.E.M., 2016. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. Current Biology 26, 38–51. https://doi.org/10.1016/j.cub.2015.11.034

Caputi, L., Andreakis, N., Mastrototaro, F., Cirino, P., Vassillo, M., Sordino, P., 2007. Cryptic speciation in a model invertebrate chordate. PNAS 104, 9364–9369. https://doi.org/10.1073/pnas.0610158104

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Christiaen, L., Wagner, E., Shi, W., Levine, M., 2009a. Isolation of sea squirt (Ciona) gametes, fertilization, dechorionation, and development. Cold Spring Harb Protoc 2009, pdb.prot5344. https://doi.org/10.1101/pdb.prot5344

Christiaen, L., Wagner, E., Shi, W., Levine, M., 2009b. Whole-mount in situ hybridization on sea squirt (Ciona intestinalis) embryos. Cold Spring Harb Protoc 2009, pdb.prot5348. https://doi.org/10.1101/pdb.prot5348

Ciliberti, S., Martin, O.C., Wagner, A., 2007. Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Comput. Biol 3, e15. https://doi.org/10.1371/journal.pcbi.0030015

Cusanovich, D.A., Reddington, J.P., Garfield, D.A., Daza, R.M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H.A., Christiansen, L., Qiu, X., Steemers, F.J., Trapnell, C., Shendure, J., Furlong, E.E.M., 2018. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 555, 538–542. https://doi.org/10.1038/nature25981

Daugherty, A.C., Yeo, R.W., Buenrostro, J.D., Greenleaf, W.J., Kundaje, A., Brunet, A., 2017. Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. Genome Res. 27, 2096–2107. https://doi.org/10.1101/gr.226233.117

Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.-K., Turon, X., López-Legentil, S., Piette, J., Lemaire, P., Douzery, E.J.P., 2018. A phylogenomic framework and timescale for comparative studies of tunicates. BMC Biology 16, 39. https://doi.org/10.1186/s12915-018-0499-2

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 Genetics 9, e1003904. https://doi.org/10.1371/journal.pgen.1003904

Duboule, D., 2007. The rise and fall of Hox gene clusters. Development 134, 2549–2560. https://doi.org/10.1242/dev.001065

Fanelli, A., Lania, G., Spagnuolo, A., Di Lauro, R., 2003. Interplay of negative and positive signals controls endoderm-specific expression of the ascidian Cititf1 gene promoter. Developmental Biology 263, 12–23. https://doi.org/10.1016/S0012-1606(03)00397-X

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., Levine, M.S., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. PNAS 113, 6508–6513. https://doi.org/10.1073/pnas.1605085113

Félix, M.-A., Barkoulas, M., 2012. Robustness and flexibility in nematode vulva development. Trends in Genetics 28, 185–195. https://doi.org/10.1016/j.tig.2012.01.002

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., Stern, D.L., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466, 490–493. https://doi.org/10.1038/nature09158

Fujiwara, S., Cañestro, C., 2018. Reporter Analyses Reveal Redundant Enhancers that Confer Robustness on <Emphasis Type="Italic">Cis</Emphasis>-Regulatory Mechanisms, in: Transgenic Ascidians, Advances in Experimental Medicine and Biology. Springer, Singapore, pp. 69–79. https://doi.org/10.1007/978-981-10-7545-2_7

Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., Berney, T., Montanya, E., Mohlke, K.L., Lieb, J.D., Ferrer, J., 2010. A map of open chromatin in human pancreatic islets. Nature Genetics 42, 255–259. https://doi.org/10.1038/ng.530

Guignard, L., Fiuza, U.-M., Leggio, B., Faure, E., Laussu, J., Hufnagel, L., Malandain, G., Godin, C., Lemaire, P., 2018. Contact-dependent cell communications drive morphological invariance during ascidian embryogenesis. bioRxiv 238741. https://doi.org/10.1101/238741

Hagey, D.W., Zaouter, C., Combeau, G., Lendahl, M.A., Andersson, O., Huss, M., Muhr, J., 2016. Distinct transcription factor complexes act on a permissive chromatin landscape to establish regionalized gene expression in CNS stem cells. Genome Res. 26, 908–917. https://doi.org/10.1101/gr.203513.115

Harafuji, N., Keys, D.N., Levine, M., 2002. Genome-wide identification of tissue-specific enhancers in the Ciona tadpole. PNAS 99, 6802–6805. https://doi.org/10.1073/pnas.052024999

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., Eisen, M.B., 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS Genet 4, e1000106. https://doi.org/10.1371/journal.pgen.1000106

Hinman, V.F., Davidson, E.H., 2007. Evolutionary plasticity of developmental gene regulatory network architecture. Proc. Natl. Acad. Sci. U.S.A 104, 19404–19409. https://doi.org/10.1073/pnas.0709994104

Hudson, C., Sirour, C., Yasuo, H., 2016. Co-expression of Foxa.a, Foxd and Fgf9/16/20 defines a transient mesendoderm regulatory state in ascidian embryos. eLife Sciences 5, e14692. https://doi.org/10.7554/eLife.14692

Imai, K.S., 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, 4047–58. https://doi.org/10.1242/dev.01270

Imai, K.S., Levine, M., Satoh, N., Satou, Y., 2006. Regulatory blueprint for a chordate embryo. Science 312, 1183–1187. https://doi.org/10.1126/science.1123404

Keys, D.N., Lee, B., Di Gregorio, A., Harafuji, N., Detter, J.C., Wang, M., Kahsai, O., Ahn, S., Zhang, C., Doyle, S.A., Satoh, N., Satou, Y., Saiga, H., Christian, A.T., Rokhsar, D.S., Hawkins, T.L., Levine, M., Richardson, P.M., 2005. A saturation screen for cis-acting regulatory DNA in the Hox genes of Ciona intestinalis. Proc Natl Acad Sci U S A 102, 679–83. https://doi.org/10.1073/pnas.0408952102

Lamy, C., Rothbächer, U., Caillol, D., Lemaire, P., 2006. Ci-FoxA-a is the earliest zygotic determinant of the ascidian anterior ectoderm and directly activates Ci-sFRP1/5. Development 133, 2835–44. https://doi.org/10.1242/dev.02448

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., Przeworski, M., 2012. Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? PLOS Biology 10, e1001388. https://doi.org/10.1371/journal.pbio.1001388

Lemaire, P., 2011. Evolutionary crossroads in developmental biology: the tunicates. Development 138, 2143–2152. https://doi.org/10.1242/dev.048975

Li, L., Zhu, Q., He, X., Sinha, S., Halfon, M.S., 2007. Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. Genome Biology 8, R101. https://doi.org/10.1186/gb-2007-8-6-r101

Liu, G., Wang, W., Hu, S., Wang, X., Zhang, Y., 2018. Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. Genome Res. https://doi.org/10.1101/gr.228833.117

Lowe, E.K., Stolfi, A., 2018. Developmental system drift in motor ganglion patterning between distantly related tunicates. bioRxiv 320804. https://doi.org/10.1101/320804

Maere, S., Heymans, K., Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448–3449. https://doi.org/10.1093/bioinformatics/bti551

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Matsumoto, J., Katsuyama, Y., Ohtsuka, Y., Lemaire, P., Okamura, Y., 2008. Functional analysis of synaptotagmin gene regulatory regions in two distantly related ascidian species. Development, Growth & Differentiation 50, 543–552. https://doi.org/10.1111/j.1440-169X.2008.01049.x

Matsumoto, J., Kumano, G., Nishida, H., 2007. Direct activation by Ets and Zic is required for initial expression of the Brachyury gene in the ascidian notochord. Developmental Biology 306, 870–882. https://doi.org/10.1016/j.ydbio.2007.03.034

Oda-Ishii, I., Bertrand, V., Matsuo, I., Lemaire, P., Saiga, H., 2005. Making very similar embryos with divergent genomes: conservation of regulatory mechanisms of Otx between the ascidians Halocynthia roretzi and Ciona intestinalis. Development 132, 1663–74. https://doi.org/10.1242/dev.01707

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Oda-Ishii, I., Kubo, A., Kari, W., Suzuki, N., Rothbächer, U., Satou, Y., 2016. A Maternal System Initiating the Zygotic Developmental Program through Combinatorial Repression in the Ascidian Embryo. PLOS Genetics 12, e1006045. https://doi.org/10.1371/journal.pgen.1006045

Perry, M.W., Boettiger, A.N., Bothma, J.P., Levine, M., 2010. Shadow Enhancers Foster Robustness of Drosophila Gastrulation. Current Biology 20, 1562–1567. https://doi.org/10.1016/j.cub.2010.07.043

Pott, S., Lieb, J.D., 2015. What are super-enhancers? Nature Genetics 47, 8–12. https://doi.org/10.1038/ng.3167

Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E., Kuan, S., Visel, A., Pennacchio, L.A., Zhang, K., Ren, B., 2018. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nature Neuroscience 21, 432–439. https://doi.org/10.1038/s41593-018-0079-3

Ristoratore, F., Spagnuolo, A., Aniello, F., Branno, M., Fabbrini, F., Lauro, R.D., 1999. Expression and functional analysis of Cititf1, an ascidian NK-2 class gene, suggest its role in endoderm development. Development 126, 5149–5159.

Roure, A., Lemaire, P., Darras, S., 2014. An Otx/Nodal Regulatory Signature for Posterior Neural Development in Ascidians. PLOS Genetics 10, e1004548. https://doi.org/10.1371/journal.pgen.1004548

Sardet, C., McDougall, A., Yasuo, H., Chenevert, J., Pruliere, G., Dumollard, R., Hudson, C., Hebras, C., Le Nguyen, N., Paix, A., 2011. Embryological methods in ascidians: the Villefranche-sur-Mer protocols. Methods Mol. Biol. 770, 365–400. https://doi.org/10.1007/978-1-61779-210-6_14

Satou, Y., Imai, K.S., Satoh, N., 2001. Early embryonic expression of a LIM-homeobox gene Cs-lhx3 is downstream of beta-catenin and responsible for the endoderm differentiation in Ciona savignyi embryos. Development 128, 3559–70.

Satou, Y., Mineta, K., Ogasawara, M., Sasakura, Y., Shoguchi, E., Ueno, K., Yamada, L., Matsumoto, J., Wasserscheid, J., Dewar, K., Wiley, G., Macmil, S., Roe, B., Zeller, R., Hastings, K., Lemaire, P., Lindquist, E., Endo, T., Hotta, K., Inaba, K., 2008. Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. Genome Biol 9, R152. https://doi.org/10.1186/gb-2008-9-10-r152

Sharma, S., Wang, W., Stolfi, A., 2018. Single-cell transcriptome profiling of the Ciona larval brain. bioRxiv 319327. https://doi.org/10.1101/319327

Sommer, R.J., 2012. Evolution of Regulatory Networks: Nematode Vulva Induction as an Example of Developmental Systems Drift, in: Evolutionary Systems Biology, Advances in Experimental Medicine and Biology. Springer, New York, NY, pp. 79–91. https://doi.org/10.1007/978-1-4614-3567-9_4

Stern, D.L., Orgogozo, V., 2008. The Loci of Evolution: How Predictable Is Genetic Evolution? Evolution 62, 2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x

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 Sciences 3, e03728. https://doi.org/10.7554/eLife.03728

Takahashi, H., Mitani, Y., Satoh, G., Satoh, N., 1999. Evolutionary alterations of the minimal promoter for notochord-specific Brachyury expression in ascidian embryos. Development 126, 3725–3734.

Thomas, S., Li, X.-Y., Sabo, P.J., Sandstrom, R., Thurman, R.E., Canfield, T.K., Giste, E., Fisher, W., Hammonds, A., Celniker, S.E., Biggin, M.D., Stamatoyannopoulos, J.A., 2011. Dynamic reprogramming of chromatin accessibility during Drosophilaembryo development. Genome Biology 12, R43. https://doi.org/10.1186/gb-2011-12-5-r43

Treen, N., Heist, T., Wang, W., Levine, M., 2018. Depletion of Maternal Cyclin B3 Contributes to Zygotic Genome Activation in the Ciona Embryo. Current Biology 28, 1150-1156.e4. https://doi.org/10.1016/j.cub.2018.02.046

True, J.R., Haag, E.S., 2001. Developmental system drift and flexibility in evolutionary trajectories. Evol. Dev. 3, 109–119.

Tsagkogeorga, G., Cahais, V., Galtier, N., 2012. The Population Genomics of a Fast Evolver: High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate Ciona intestinalis. Genome Biol Evol 4, 852–861. https://doi.org/10.1093/gbe/evs054

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., Li, W., Li, Y., Ma, J., Peng, X., Zheng, H., Ming, J., Zhang, W., Zhang, J., Tian, G., Xu, F., Chang, Z., Na, J., Yang, X., Xie, W., 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature 534, 652–657. https://doi.org/10.1038/nature18606

Yáñez-Cuna, J.O., Kvon, E.Z., Stark, A., 2013. Deciphering the transcriptional cis-regulatory code. Trends in Genetics 29, 11–22. https://doi.org/10.1016/j.tig.2012.09.007

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

ACKNOWLEDGEMENTS:

We thank Dr. A Spagnuolo (Stazione Zoologica Anton Dohrn, Naples, Italy) for the kind gift of *Ciona robusta* genomic DNA, Marc Plays and Philippe Richard for excellent animal husbandry, members of our groups for many discussions and suggestions. PL's lab was supported by grants from the Agence Nationale pour la Recherche (TED project, ANR-13-BSV2-0011). JLG-S's lab was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No's ERC-AdG-LS8), the Spanish Ministerio de Economía y Competitividad (BFU2016-74961-P and BFU2014-55738), the 'Unidad de Excelencia María de Maetzu 2017-2021'(MDM-2016-0687) to the Department of Gene Regulation and Morphogenesis of CABD), and the Andalusian Government (BIO-396). AM was a PhD student of the University of Montpellier, supported successively by the EpiGenMed Laboratory of excellence and FRM (FDT20160435698). MSM was a PhD student from the University Pablo de Olavide supported by the Spanish Ministerio de Economía y Competitividad. DG and SH were supported by TED. UMF and LG were post-doctoral fellows supported by the FRM (SPF20120523969) and HHMI, respectively. PL and CD are members of CNRS and JLG-S is a member of CSIC.

AUTHORS CONTRIBUTION:

PL and JLG-S initiated the project. MSM and AM performed the ATAC-seq experiments, which were computationally processed and analysed by CD. AM performed most *in situ* hybridizations, with contributions from UMF. AM, SH and DG performed all cloning and electroporation experiments. LG participated to the analysis of the data and to construction the figures. AM and PL wrote the manuscript. All authors commented on the manuscript.

FIGURE LEGENDS:

Figure 1: A global analysis of the results of ATAC-seq experiments in *Ciona robusta* and *Phallusia mammillata.* A) Bird's eye view of the open chromatin landscape over 500 kb of *Ciona robusta* chromosome 12 visualized with the ANISEED WashU browser. The stages at which the ATAC-seq experiments were performed and the biological replica are indicated. Two rows of peaks, extracted from the merged reads of each stage, are shown: top, 64-cell stage; bottom, 112-cell stage. Conservation

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

between *Ciona robusta* and *Ciona savignyi* is indicated (range 60-80 %). **B)** Overlap between chromatin accessibility (ATAC-seq peaks) and sequence conservation between *Phallusia mammillata* and *Phallusia fumigata* within non-coding sequences. The left ATAC-seq vs HCNE bar shows the fraction of ATAC-seq peaks at the 112-cell stage overlapping a Highly-Conserved Non-coding Element: a region of LastZ alignment with more than 60% local identity. Conversely, the right HCNE vs ATAC-seq bar shows the fraction of HCNE overlapping one or more ATAC-seq peaks at the 112-cell stage. **C)** GC content analysis in *Ciona robusta* (yellow) and *Phallusia mammillata* (red) in either exons or non-coding sequences (introns and intergenic sequences) as a function of their chromatin accessibility at the 112-cell stage (open = in an extracted ATAC-seq peak). **D)** Dynamic opening of two chromatin regions, one immediately upstream, the other within the second intron of the Phmamm.CG.MTP2004.S242.g060002 gene. H3K4me3 and RNA-seq data from (Brozovic et al., 2018). ATAC, RNA: ATAC-seq and RNA-seq signals at the indicated stage (64-cell, mid (RNA) or late (ATAC) gastrula, mid-neurula).

Figure 2: Conservation of Clusters of Open Regulatory Elements (COREs). A) A bird's eye view of 150kb of *Ciona robusta* chromosome 1 showing the clustering of ATAC-seq signal around gene KH.C1.32, a C2H2 zinc finger gene. The ATAC-seq signal at the 64-cell and 112-cell stages and the conservation between *Ciona robusta* and *Ciona savignyi* (identity >60%) are shown. **B)** Most-enriched GO terms (P-value smaller than 10^{-02}, Terms describing Molecular Function or Biological Processes only) in genes overlapping or immediate neighbours to COREs (≥5 non-coding peaks within 10kbs) in both species. Each enriched term was manually assigned to one of the 5 shown categories. Note the poorer enrichment of "Non-regulatory" terms. GO term enrichments were computed from the Bingo tool of Cytoscape (Maere et al., 2005).

Figure 3: Conserved temporal and spatial early expression profiles between *Ciona* and *Phallusia* regulatory genes. A) A cladogram with mean divergence dates and associated standard deviation (in million years from now) of the *Phallusia* and *Ciona* species discussed in this study (Delsuc *et al*., 2018; LN CAT-GTR + Γ_4 model). **B)** Example *in situ* hybridizations for two genes, *Lefty* and *Fgf9/16/20* at the 64-cell stage in the published literature (*Ciona robusta*, Imai et al., 2004) and in our *Phallusia mammillata* experiments, showing the impact of the short or long duration of the NBT/BCIP staining reaction on the interpretation of expression patterns. **C)** *In situ* hybridisation in *Phallusia mammillata* of the indicated genes at 16-, 32-, 64- and 112-cell stages. All of the images are vegetal views except for *Foxa.a* and *Lefty* at 16-cell stage. *FGF9/16/20* is abbreviated as *FGF9*. **D)** Comparison table of ortholog gene expression at the cellular level at the 16-, 32-, 64- and 112-cell stages. Cell identities are indicated on top of each column. Endoderm precursors identities are in bold. Blue cells: shared expression between *Ciona robusta* and *Phallusia mammillata*; red cells: *Phallusia*-specific expression; yellow cells: *Ciona*-specific expression.

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4: Relationship between chromatin accessibility and *cis*-regulatory activity in *Ciona* and *Phallusia*. A) Position of previously published *Ciona robusta* regulatory sequences for *Otx* (Bertrand et al., 2003a) with respect to ATAC-seq signal (orange) at the 64-cell stage, of Gata.a ChIP-seq signal (yellow) at the 32-cell stage (Oda-Ishii et al., 2016) and of H3K4me3 ChIP-seq signal (green) at the early gastrula stage (Brozovic et al., 2018). Published pictures of representative embryos are shown next to the tested constructs. The profile of *Ciona robusta/Ciona savignyi* sequence identity (>60%) is indicated in blue. The ANISEED identifier of each *cis*-regulatory sequence is indicated. Green: active sequence; red: inactive one. **B)** Similar analysis and representation for the *Ciona robusta FGF9/16/20* gene, except that the Tcf7 ChIP-seq signal (Oda-Ishii et al., 2016) is shown in yellow. The picture showing the activity of the regulatory sequence at the 16-cell stage is from (Oda-Ishii et al., 2016). **C)** Analysis of chromatin accessibility at the 64-cell stages upstream of the *Phallusia mammillata Nodal* gene. The regions tested by electroporation are indicated (green: active by the larval stage; red: inactive up to the larval stage). Note that *Nodal*Pm04, which is evolutionary conserved between *P. mammillata* and *P. fumigata* but not in an open chromatin configuration is inactive by electroporation. **D)** Representative activities at the larval and early gastrula stages of the *Nodal*Pm01, *Nodal*Pm02 and *Nodal*Pm03 sequences. The percentage range of embryos showing staining upon electroporation is indicated. Larval *Nodal*Pm03 staining is found in the progeny of cells stained at the early gastrula stage (b6.5 lineage). Larval *Nodal*Pm01 staining is found both in the progeny of cells stained at the early gastrula stage (TLC, anterior endoderm), as well as in other territories (endodermal strand, muscle).

Figure 5: Functional characterisation of the complex regulatory landscapes of early endodermal regulatory genes. Summary of the position and activity of 60 tested sequences and 4 published known regulatory sequences, corresponding to open chromatin regions around orthologous *Phallusia* and *Ciona* genes; the colours represent the spatial tissue specificity of tested regulatory sequences. Stars indicate the published known regulatory sequences: *FGF9/16/20* (ANISEED REG00001080, Oda-Ishii et al., 2016), *Foxd* (ANISEED REG00001115, Oda-Ishii et al., 2016), *Lefty* (ANISEED REG00001118, Oda-Ishii et al., 2016), *Nkx2-1/4* (ANISEED REG00000321 and REG00001175, Fanelli et al., 2003; Doglio et al., 2013) and *Otx* (ANISEED REG00000010, Bertrand et al., 2003a). The schema respects the order of the tested sequences on the chromosome but is not drawn to scale. Only the first and last exons are represented in the gene body.

Figure 6: Overlapping temporal and spatial enhancer activity contributes to regulatory landscape complexity. A) A schematic of tissues in the *Phallusia mammillata* embryos at 112-cell stage (on animal and vegetal side) and larva (under the epidermal cells). **B and C)** Schematic of the active regulatory sequences driving *Phallusia* genes at the larval and early gastrula stages, respectively. The

bioRxiv preprint doi: https://doi.org/10.1101/371005; this version posted July 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

coloured rounded rectangles are putative shadow enhancers at the stage considered and the grey rounded rectangles are other tested sequences see Fig 5. The colour of the regulatory sequence indicates to which image it corresponds. Next to the images of the electroporation results are the schematics summarising the overall main activity found across the embryos for each construct.

Figure 7: Gastrula trans-regulatory activity is temporally and spatially conserved cross-species. A) Early gastrula LacZ activity (except *Lhx3/4*Ci01, 64-cell stage) driven in *Phallusia mammillata* and *Ciona intestinalis* by three *Ciona* and one *Phallusia* sequences with conserved activity across the two species. The central schema (left vegetal view, right animal view) summarise on the left half summarises the *Phallusia* activity (pink) and the right half the activity in *Ciona* (yellow). **B)** Summary table of the activity in *Phallusia* and *Ciona* main tissue precursors of 17 sequences as detected by LacZ staining in early or late gastrulae. Blue, red and yellow indicate conserved activity, activity only detected in *Phallusia* and activity only in *Ciona*, respectively. The top box presents a bar graph of the fraction of conserved vs divergent activities of each sequence across all tissues. The side box is a bar graph of the frequency of conserved or divergent activity in each tissue across all sequences.

Figure 8: Collinearity of *Phallusia* and *Ciona Otx* and *FGF9/16/20 cis*-regulatory elements driving similar qualitative activity. A) Graph of the number of open chromatin regions (called ATAC-seq peaks) neighbouring the 10 pairs of orthologous genes studied here. The *x* axis is the number of peaks around the *Phallusia* gene and the *y* axis is the number of peaks around the *Ciona* gene. The regression line is shown in red. The R^2 value is the coefficient of determination. **B)** Comparison of the open chromatin landscapes of the *Ciona* and *Phallusia Otx* (KH.C4 and MTP2014.S166) and *FGF9/16/20* (KH.C2 and MTP2014.S128) loci. In the case of *Otx*, the activity of the corresponding peaks tested by electroporation at the early gastrula stage is indicated by a representative picture (which may show mosaic expression) and a schematic representation of the complete pattern of activity. The activity shown represents either the activity of the sequence itself (boxed in red) or an inference of its activity obtained by comparing two constructs differing by this sequence (boxed in green with the identity of the two ANISEED cis-regulatory sequences compared). **C)** Early gastrula stage LacZ activity driven in *Phallusia mammillata* and *Ciona intestinalis* by two orthologous *Otx cis*-regulatory elements. The central schema follows the same logic as Figure 7A.

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available un
aCC-BY-NC 4.0 International license.

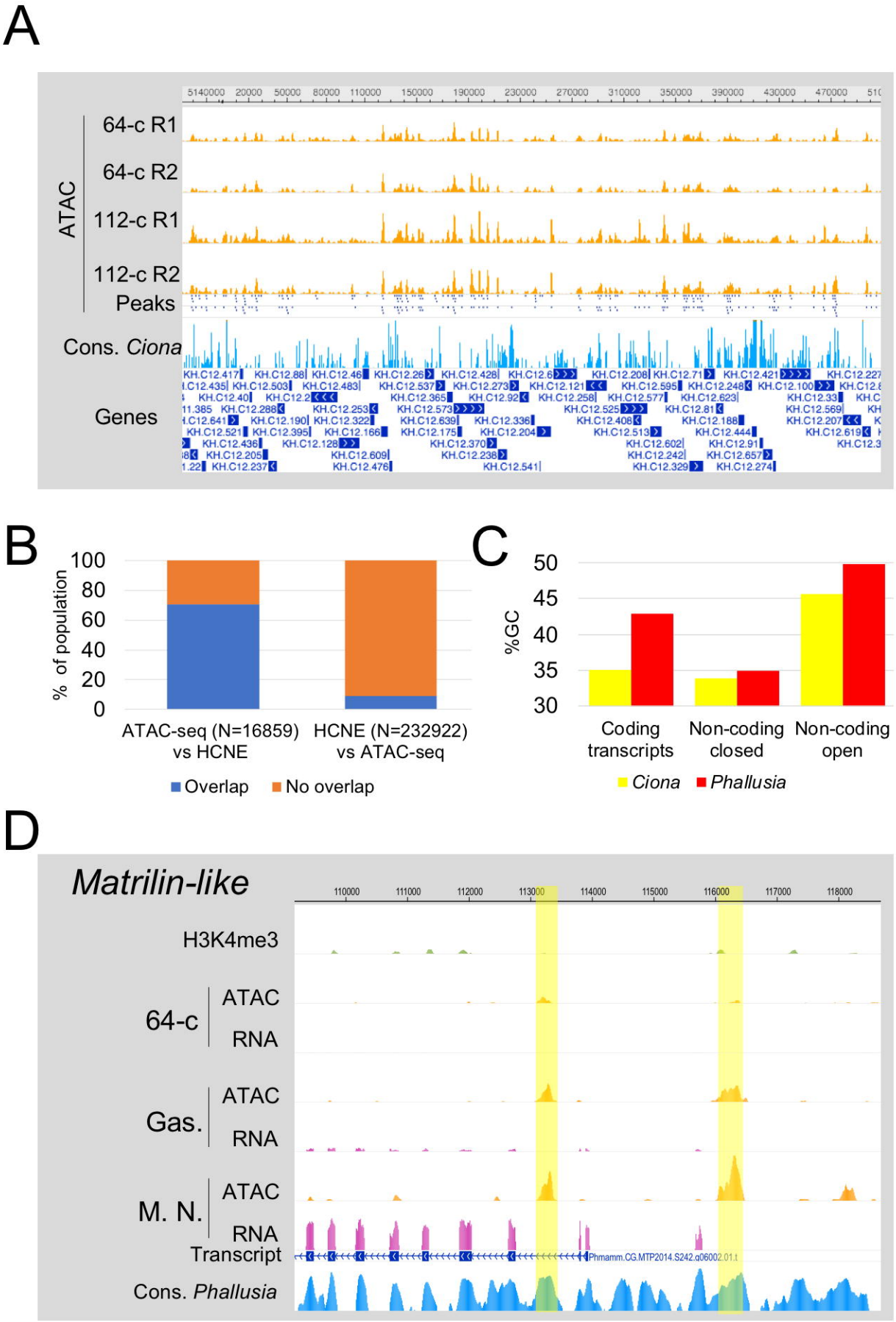
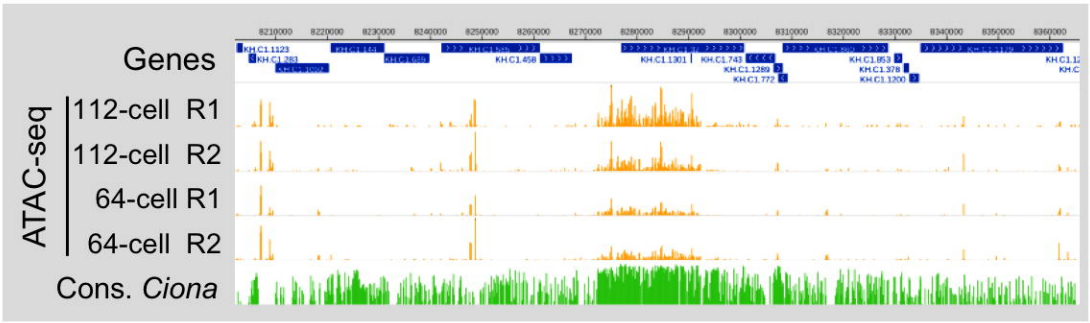


Figure 1: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available un
aCC-BY-NC 4.0 International license.

A



B

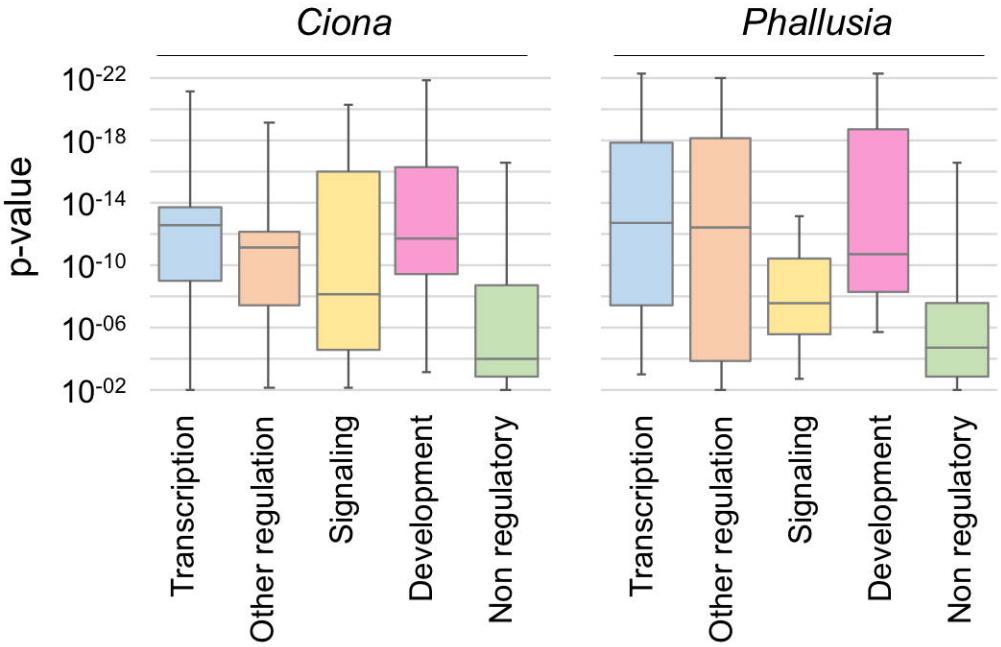


Figure 2: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC 4.0 International license.

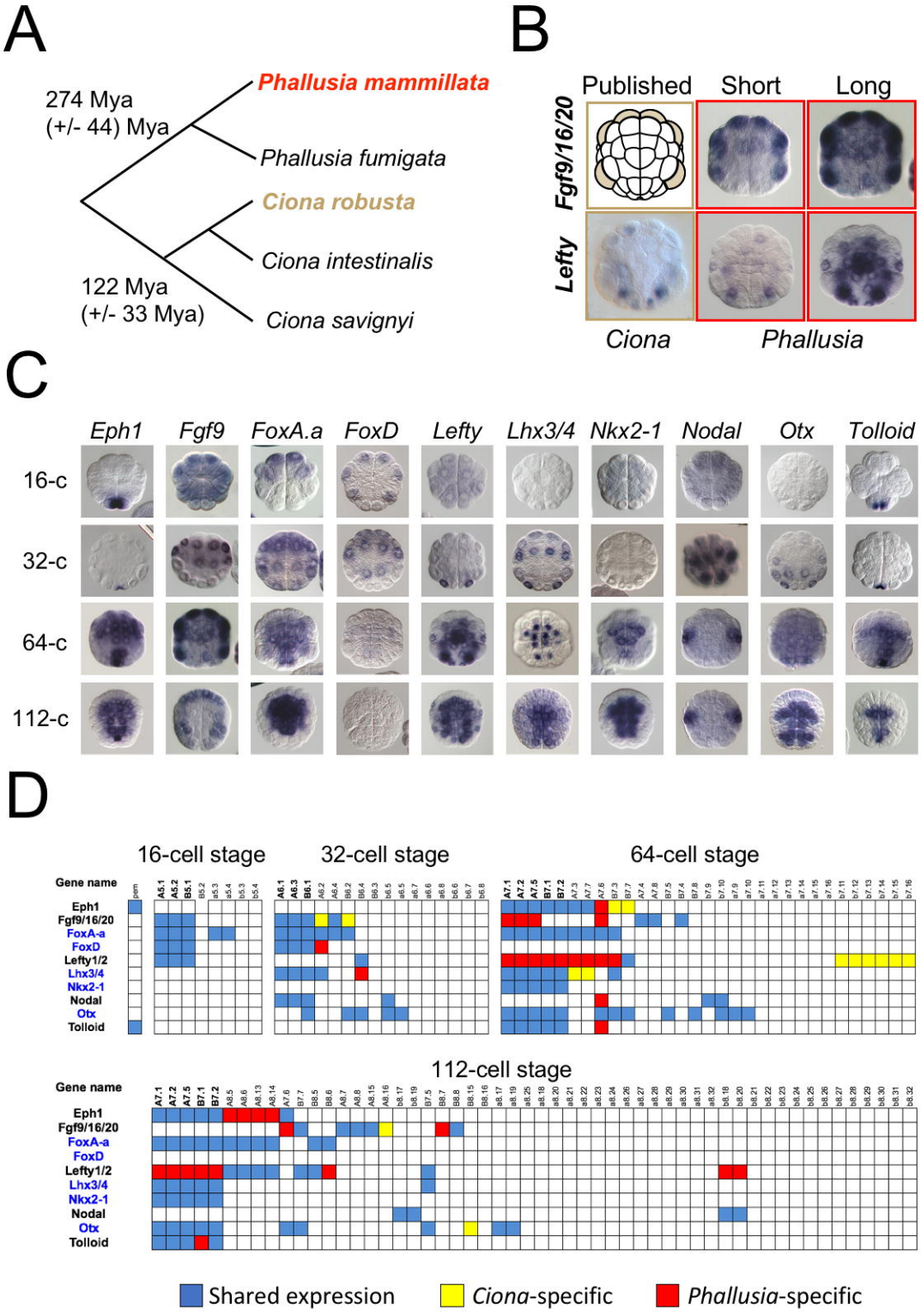


Figure 3: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC 4.0 International license.

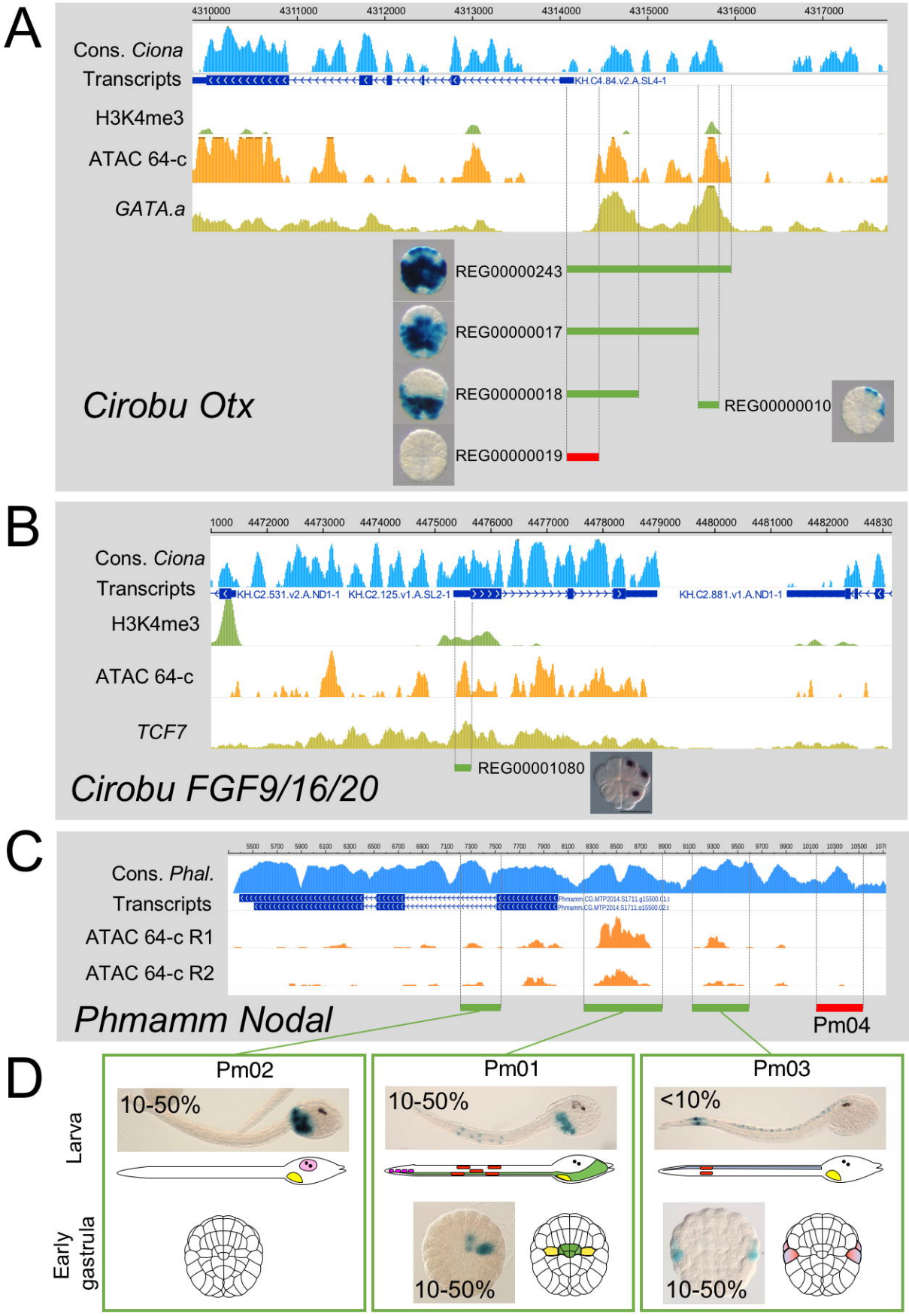


Figure 4: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC 4.0 International license.

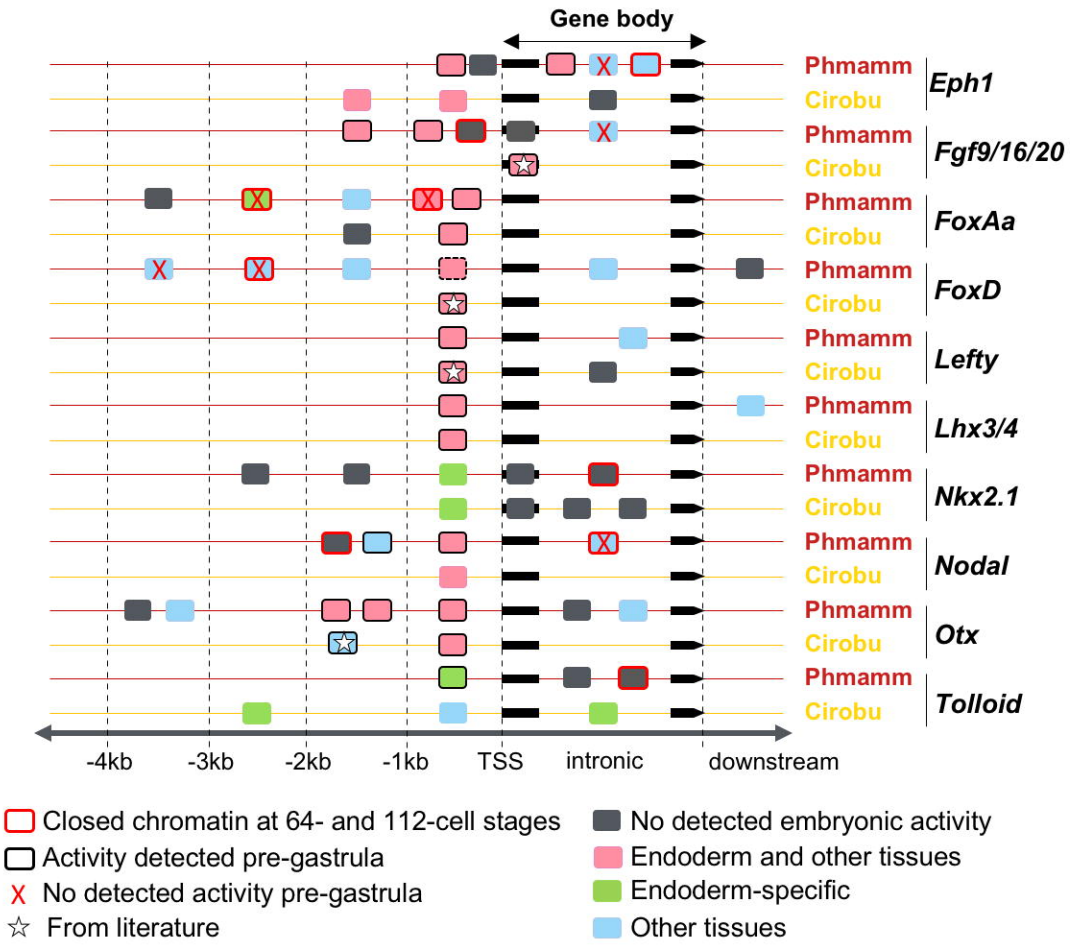


Figure 5: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC 4.0 International license.

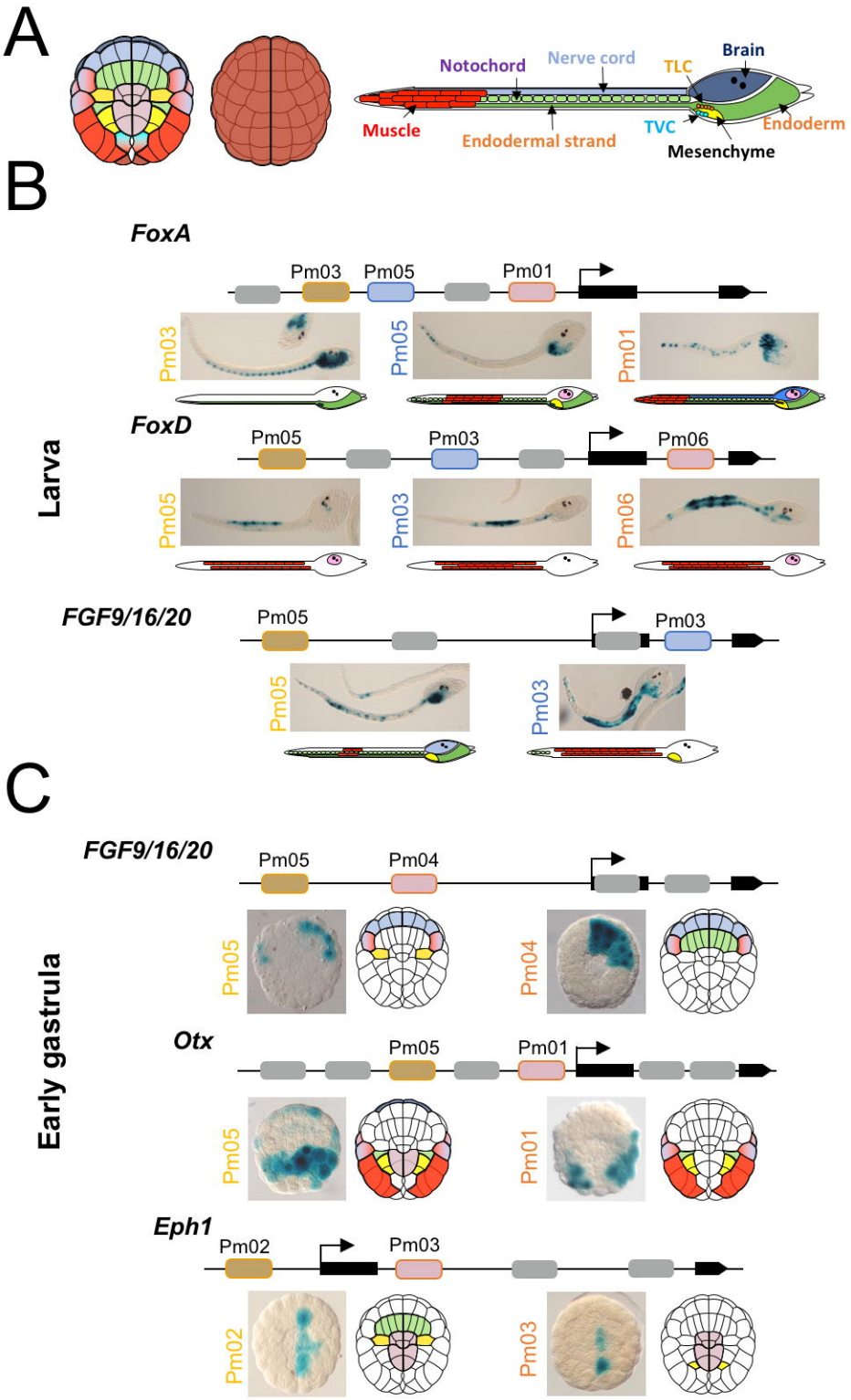


Figure 6: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC 4.0 International license.

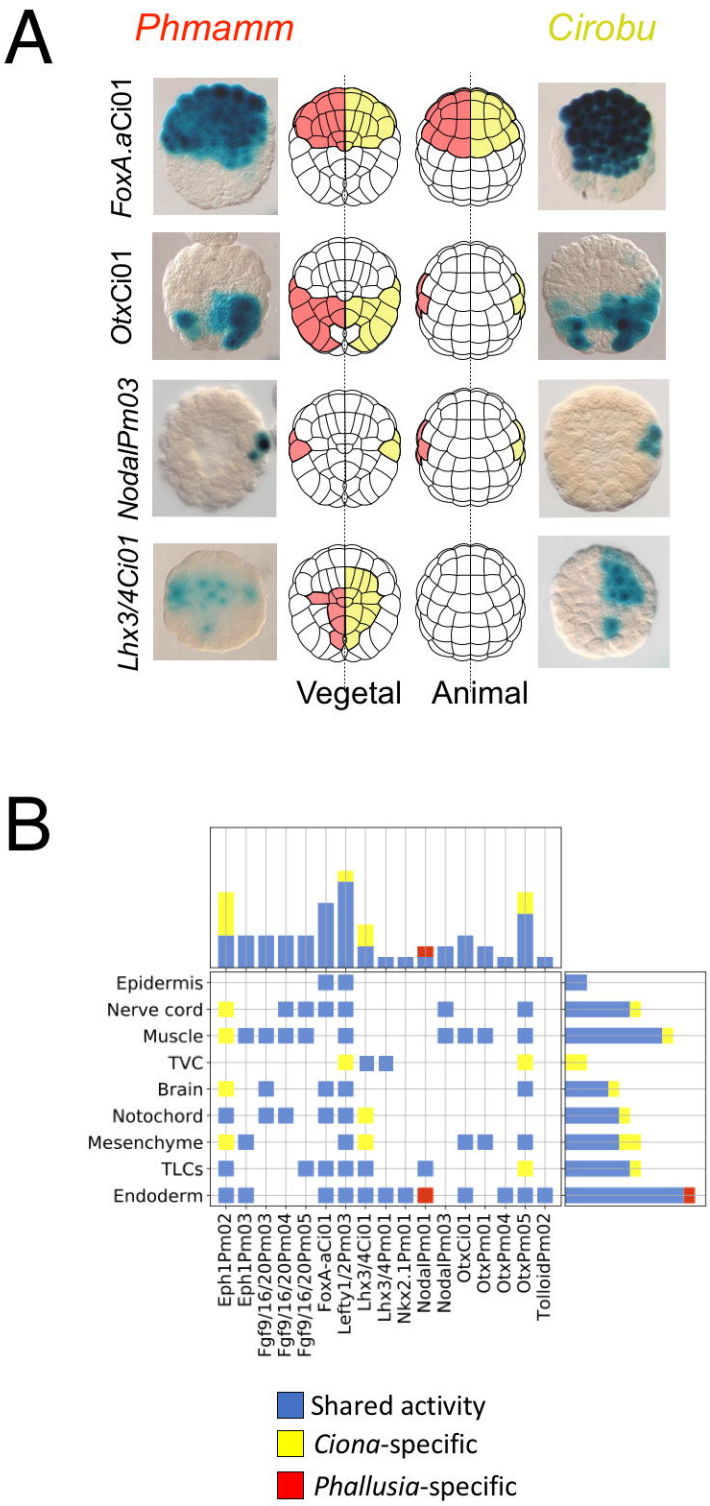


Figure 7: Madgwick et al., 2018

certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available u
aCC-BY-NC 4.0 International license.

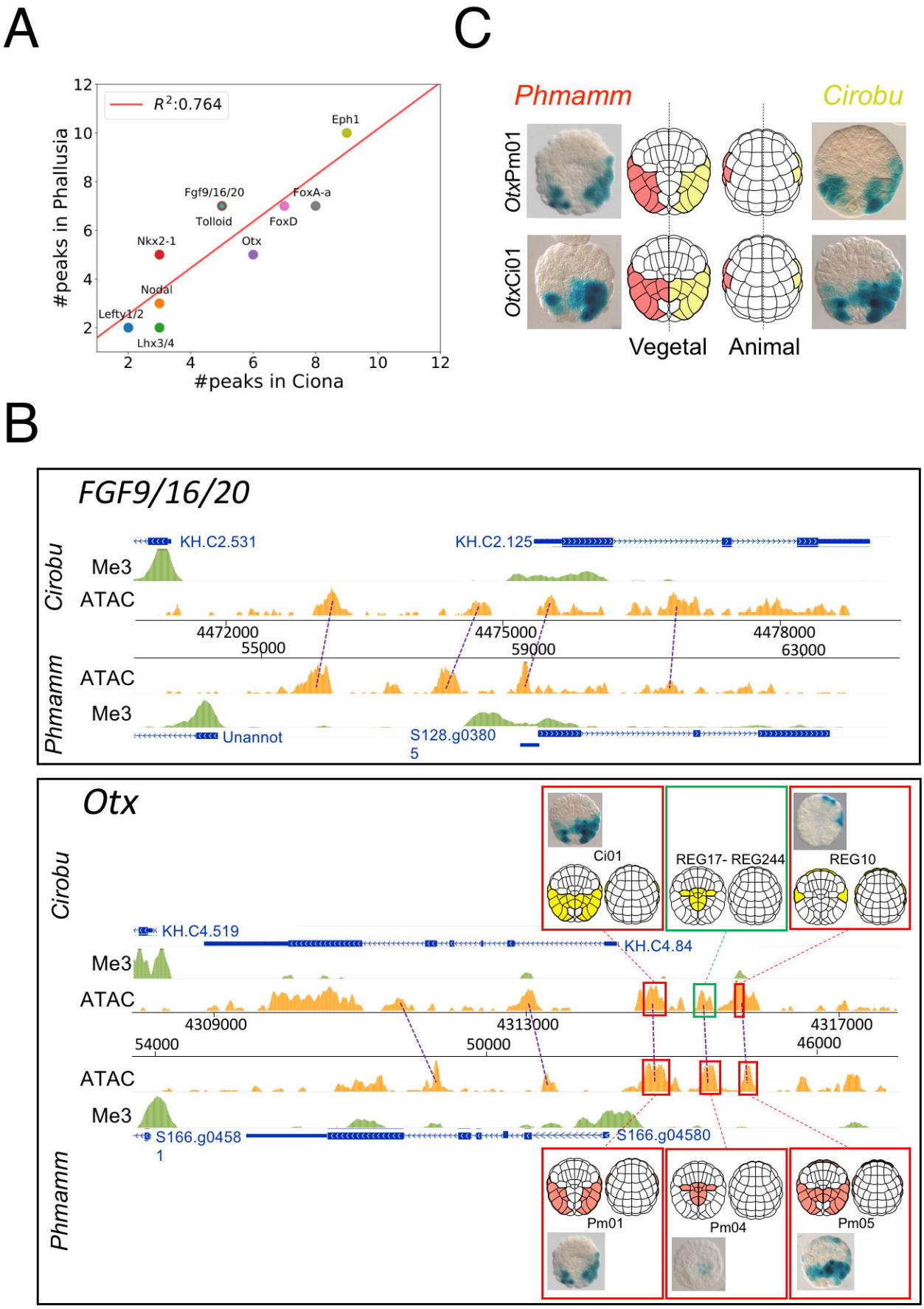


Figure 8: Madgwick et al., 2018.