Logics and properties of a genetic regulatory program that drives embryonic muscle development in an echinoderm

Carmen Andrikou[1#], Chih-Yu Pai[2], Yi-Hsien Su [2]* and Maria Ina Arnone[1]*

[1] Cellular and Developmental Biology, Stazione Zoologica Anton Dohrn, Napoli 80121, Italy

[2] Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan

[#]Present address: Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen 5008, Norway

*Corresponding authors: miarnone@szn.it, yhsu@gate.sinica.edu.tw

ABSTRACT

Evolutionary origin of muscle is a central question when discussing mesoderm evolution. Developmental mechanisms underlying somatic muscle development have mostly been studied in vertebrates and fly where multiple signals and hierarchic genetic regulatory cascades selectively specify myoblasts from a pool of naïve mesodermal progenitors. However, due to the increased organismic complexity and distant phylogenetic position of the two systems, a general mechanistic understanding of myogenesis is still lacking. Here, we propose a gene regulatory network (GRN) model that promotes myogenesis in the sea urchin embryo, an early branching deuterostome. A FGF signalling and four Forkhead transcription factors consist the central part of our model and appear to orchestrate the myogenic process. The topological properties of the network reveal dense gene interwiring and a multilevel transcriptional regulation of conserved and novel myogenic genes. Finally, the comparison of the myogenic network architecture among different animal groups highlights the evolutionary plasticity of developmental GRNs.

Keywords: Gene Regulatory Network, specification, myogenesis, FGF, Forkhead, Myogenic Regulatory Factors, Myosin Heavy Chain, perturbation, mesoderm, myoblast, *Strongylocentrotus purpuratus*

INTRODUCTION

Muscle cells are present in most animals with the characteristic of containing protein filaments that slide and produce contractions. In bilaterians, muscles develop usually from mesodermal cells in a process called myogenesis, during which the naïve mesodermal progenitors are selectively specified as myoblasts and later differentiate into

muscle cells. In vertebrate embryos, the mesoderm is subdivided into several regions from which different muscle types originated. For example, the axial skeletal muscles originate in the segmented paraxial mesoderm called somites, whereas the cardiac muscles develop from the lateral plate mesoderm. The subdivisions of different mesoderm regions require differential expression of the Forkhead (Fox) family of transcription factors, such as FoxC1 and FoxC2 for the paraxial mesoderm (1) and FoxF1 for the lateral plate mesoderm (2). Analyses of molecular mechanisms underlying myogenesis in various parts of the somites further lead to a series of complex GRNs (3-5). These networks are composed of multiple signals that drive the expression of the genes encoding basic helix-loop-helix (bHLH) domain-containing myogenic regulatory factors (MRFs), which include myogenic differentiation 1 (MyoD), myogenic factor 5 (Myf5), Myf6 and myogenin (Myog). Other factors, including the myocyte enhancer binding factor 2 (MEF2) family of MADS-box proteins, transcription factors Pitx2, Pitx3, sine oculis-related homeobox (six) family members and their cofactors eyes-absent homologs (Eya), are also involved in specification and migration of the myogenic precursor cells (5, 6). These myogenic transcriptional regulators then in turn initiate the transcription of muscle differentiation markers, such as myosin heavy chain proteins (MHC), that determine the morphological and functional identity of the differentiated muscle cells (3, 4).

In addition to the transcription factors mentioned above, myogenesis in different parts of the vertebrate somites depends on several signals including Notch, Fibroblast Growth Factor (FGF), Wnts, Bone Morphogenetic Factor 4 (BMP4) and Sonic Hedgehog (Shh) that are secreted from adjacent tissues, such as neural tube, notochord, dorsal and lateral ectoderms (7-9). Among these signals, FGF signaling pathway is of particular interest because it has a crucial and conserved role in mesoderm patterning and muscle

development in various animal models. In vertebrates, FGF induces the expression of myogenic genes in the primary myotome, controls the timing of the epithelial-mesenchymal transition (EMT) of the dermomyotome, which triggers the emergence of muscle progenitors (7) and positively regulates muscle differentiation (10, 11). In flies, an FGFR ortholog (*Heartless*) mediated pathway is essential for cell migration and fate induction of the visceral mesoderm, heart and the somatic muscle lineages (12). In ascidians, FGF signaling is required for the specification of the heart progenitor cells and secondary muscle development (13, 14). In the nematode *Caenorhabditis elegans*, the crosstalk between FGF and Wnt signals is involved in larval sex myoblast specification and migration (15, 16). In the amphioxus *Brachiostoma lanceolatum*, FGF signaling is necessary for the formation of the anterior somites that will generate the musculature (17) and in the hemichordate *Saccoglossus kowalevskii* FGF signaling is necessary for all types of mesoderm development (18).

The overall genetic regulatory cascade that drives myogenesis appears to be highly conserved between ecdysozoans (*Drosophila melanogaster* and *C. elegans*) and vertebrates (19). However, due to the high complexity that characterizes the vertebrate organogenesis and the distant phylogenetic position between ecdysozoans and vertebrates, the logics behind the genomic regulatory interactions that drive muscle development remain unclear. Therefore, elucidating the properties of the key genetic interconnections that orchestrate myogenesis in a larger set of model organisms is necessary to reveal the core myogenic circuits. Moreover, given the fact that the common origin of musculature is a highly debated topic (20-22), a larger interspecies comparison of the transcriptional myogenic networks would provide new insights into the evolution of muscle and the kernel driven hypothesis.

Echinoderms occupy a key phylogenetic position since they are early branching deuterostomes, therefore are more closely related to vertebrates compared to the other two most studied invertebrate model systems (*D. melanogaster* and *C. elegans*). Sea urchin is a powerful model system for studying regulatory events that take place during development due to their advantageous properties for GRN analysis (23). GRNs provide causal explanations of the molecular interactions occurring during dynamic developmental processes such as cell specification and differentiation (24, 25). In the last decade, the extensive amount of data collected by perturbation analyses have led to the assembly of the largest so far known endomesodermal GRN (25-27) representing most of the transcriptional interactions that take place during endomesoderm formation in the sea urchin embryo. These properties make echinoderms an excellent model to study molecular developmental mechanisms and address aspects in evolution of organogenesis.

Sea urchin larvae possess a muscular apparatus that surrounds their esophagus and produces contractile force (28, 29). These muscle cells originate from mesoderm and adapt the myogenic fate by segregating from the other three non-skeletogenic mesodermal (NSM) lineages at the early gastrula stage (30). The first appearance of myoblasts is seen at the late gastrula stage as a few cells in the oral vegetal domain of each coelomic sac at the tip of the archenteron and they express a number of muscle-specific transcription regulators (31). Few hours later, at the prism stage, these cells extend pseudopods towards the midline of the esophagus, increase in number and diameter, fuse to each other and finally form the circumesophagael contractile bands (29). Despite the existence of large amount of data concerning the mechanisms patterning the early endomesoderm segregation, little is known about the regulatory landscape of the later NSM specification. Parts of the regulatory events that underlie the specification of the three NSM mesodermal lineages (blastocoelar, pigment and coelomic pouch cell

lineages) were recently revealed (26, 32-35) and only scattered information regarding the molecular interplay that establish the muscle lineage is available in the literature. In particular, transcription factors such as Twist (36) and a sea urchin lineage-specific Fox family factor, FoxY (33), and signalling pathways such as Delta/Notch (D/N) (37, 38) and Hedgehog (Hh) (39, 40) seem to be involved in muscle development but the molecular mechanisms remain poorly understood.

In our previous work, we identified several homologues of myogenic regulators in the sea urchin embryo and revealed the regulatory state of the myoblasts and their precursor cells (31). Further functional analyses are needed to confirm the myogenic function of these transcription factors and to unravel the regulatory architecture of the muscle GRN. This study provides an explanatory mechanism on how the sea urchin muscle lineage is specified using a perturbation approach accompanied by a combination of temporal and spatial gene expression analyses. We show that FGF signalling is necessary for specifying naïve NSM cells to myoblast precursors at the very early gastrula stage. The four Fox family factors FoxY, FoxC, FoxF and FoxL1 (41) constitute the central part of our myogenic GRN model and are activated sequentially. Other conserved key factors that occupy different hierarchical levels in the myogenic GRN are members of the T-box (Tbx6) (42), bHLH (MyoD2) (31), SRY (Sox) (42), Scratch (ScratchX) (33) and Six (Six1/2) (43) family genes. These findings imply an overall high level of functional conservation of both key myogenic transcriptional regulators and signalling components but not of the myogenic GRN architecture per se. Also, they explain in a rational way the logics and properties of the regulatory interactions that drive myogenesis in the sea urchin embryo.

RESULTS

Genes encoding FGF signaling components are expressed in the putative myoblast precursors

As shown in our previous work (31), the myogenic lineage seems to segregate as early as the very early gastrula stage (30 h post fertilization). The relative position of these cells is not precise at the earlier stages but they seem to be located at the oral/lateral periphery of the vegetal plate, at the border between the blastocoelar and pigment cell precursors (Figure 1-figure supplement 1 (30)). To reveal the cause of myoblast emergence at the early gastrula stage, we searched for putative candidate signaling components that are expressed at that time in the myoblast precursors. As elsewhere demonstrated, a mesodermal Delta/Notch signal activates *FoxY,* which is required for the specification of the two mesoderm derivatives, coelomic pouches and muscles (33, 44). The segregation of these two derivatives should thus specifically rely on another signal that triggers myoblast specification at that particular developmental time.

There are one FGF ligand, FGFA (Figure 1-figure supplement 2), and two FGF Receptors (FGFRs), FGFR1 and FGFR2 (Figure 1-figure supplement 3), annotated in the sea urchin genome (45, 46). A previous study showed that inhibition of FGFR2 affects morphogenesis of the embryonic skeleton (47). We therefore focused on characterizing the expression patterns of genes encoding FGFR1 and FGFA ligand in the sea urchin *Strongylocentrotus purpuratus.* The expression profile of both genes matches with the one already described in another sea urchin species, *Paracentrotus lividus* (47). To better resolve the expression patterns of these two genes relatively to the myoblast precursors, double fluorescent *in situ* hybridizations (FISHs) with *FoxC*, the earliest myoblast precursor marker (31), were performed. At the very early gastrula stage, *FGFR1* was expressed in all *FoxC*-positive cells (Figure 1A) whilst at the late gastrula stage (48 h),

only a number of *FGFR1* transcripts overlap with *FoxC*-positive cells at the tip of the archenteron (Figure 1-figure supplement 4). Moreover, *FGFA* transcripts were observed in the ventrolateral ectodermal regions in the vicinity of *FoxC*-positive cells (Figure 1B). The spatial and temporal expression of *FGFR1* and *FGFA* suggests a putative role of FGF signaling in sea urchin myogenesis.

FGF signaling specifies the time of the myoblast emergence and acts through a MAPK/ERK cascade

To test whether FGF signaling through FGFR1 is involved in sea urchin myogenesis, we perturbed FGF signaling pathway by using a) SU5402, a FGFR inhibitor (48); b) U0126, a MEK inhibitor (49); c) an antisense morpholino oligonucleotide (MO) targeted to FGFR1; d) a dominant negative form (Dn) of FGFR1. A summary of the phenotypes observed after inhibitor treatments is in Figure 2-figure supplement 1A. 70% of the 300 pluteus larvae treated with SU5402 from 26 h showed an abnormal elongated archenteron missing the pyloric and possibly the anal sphincter constrictions, although the cardiac sphincter was still formed (Figure 2A-D) suggesting that the three distinct sphincters are differentially regulated (50). Moreover, all larvae had shorter spicules and their triradiate skeleton was never fully shaped possibly due to the inhibition of FGFR2, which is specifically expressed in the skeletogenic primary mesenchyme cells (PMCs) that produce the larval skeleton (47). Finally, the coelomic pouches were not formed and the circumesophagael muscle fibers were completely absent when tested for *MHC* gene expression by FISH at the prism stage (Figure 2D). To reveal the downstream cascade of the FGF signaling, we used the MEK inhibitor U0126. In one of our previous studies we showed that MAPK/ERK activation is required for muscle formation (51). We repeated the experiment by treating the embryos with U0126 at 26 h, as we did with the SU5402

inhibitor. The treated larvae were tested again for the presence of muscle fibers by examining the MHC protein level using immunostaining. Again, the level of MHC protein was greatly reduced in the treated larvae (Figure 2-figure supplement 1B-C). These findings show an involvement of FGF signaling through MAPK/ERK pathway in muscle development.

To reinforce the role of FGF signaling mediated through FGFR1 during myogenesis, we performed gene knockdown experiments by injecting anti-translation MO against FGFR1 at a concentration of 500 μM. The MO toxicity and efficacy were examined by using a control MO and a GFP fusion construct, respectively (Figure 2-figure supplement 2). Although in 90% of the morphant larvae a malformation of coelomic pouches (Figure 2F) and a severe down-regulation of the MHC protein level was evident (Figure 2G-H), two different phenotypes concerning the compartmentalization of the archenteron were observed. 70% of the larvae displayed a fully elongated gut, with the cardiac and anal sphincters well formed (Figure 2F) while in the rest cases both sphincters were absent or partly formed (Figure 2H). In addition, an abnormal extension of cilia in the apical ectoderm was observed (Figure 2H insert), that provides a putative additional role of FGFR1 in patterning the apical organ (Paola Oliveri, personal communication) consistent with the expression of *FGFR1* in the apical domain (Figure 1A). An evident reduction of the cilia in the gut lumen is also observed (Figure 2H), which may be related to the FGFR1 expression seen in the endoderm during gastrulation (Figure 1-figure supplement 4). Moreover, the larval skeleton and the formation of pigment cells were not affected, confirming that FGFR1 is not involved either in PMC patterning and subsequent skeleton formation or in pigment cell development (Figure 2F). An even more severe phenotype was obtained by injecting the embryos with the mRNA encoding FGFR1 Dn with the cardiac and anal sphincters being absent in 60% of the cases (Figure 2J). As expected,

MHC levels were greatly reduced whilst the same abnormal extension of cilia in the apical ectoderm as well as the reduction of the cilia in the gut lumen were observed. Taking together, these experiments not only demonstrate clearly an essential role of FGF signaling through FGFR1 in sea urchin myogenesis but also suggest its possible involvement in the formation of the ciliated gut epithelium and the ciliary ectoderm.

We then tested the expression of the myoblast marker genes *FoxY, FoxC* and *FoxF*, in the FGF signaling-perturbed embryos at the very early gastrula (28 h), mid gastrula (36 h) and late gastrula (48 h) stages, respectively. While *FoxY* expression did not change when FGF signaling was perturbed (Figure 3A-D), a severe impact was observed in *FoxC* and *FoxF* transcript levels in all treated embryos (Figure 3E-L). To check whether the specification of other oral NSM cell lineages were affected as well by the FGFR1 perturbation, we also tested the expression of *Ese*, a marker of the blastocoelar cell lineage, and found it unaffected (Figure 3A-B). These results further support the involvement of FGF signaling in myogenesis and demonstrate that *FoxC* and *FoxF* factors are acting downstream of it.

ETS family transcription factors are known effectors of the MAPK signaling pathway. In order to understand whether an ETS-domain-containing transcriptional regulators is likely to be the downstream effector of the MAPK pathway leading to *FoxC* and *FoxF* transcription in presumptive myoblasts, we searched for ETS family factors that are significantly expressed in the NSM at the very early gastrula stage. From the 11 members of the ETS gene family that are present in the sea urchin genome, only 3 are expressed in the NSM at 30 h; *Elk*, *Erg* and *Ets1/2* (52). We performed an immunostaining experiment on the phosphorylated active form of Elk (P-Elk) and double FISH of *FoxC* and *Erg* or *Ets1/2*. The spatial localization of P-Elk protein seemed to be excluded from the myogenic cells, when comparing it to the expression pattern of *FoxC* (Figure 2-figure

supplement 3A-C), suggesting that P-Elk is probably not the candidate effector of MAPK involved in myoblast specification. *Erg* expression was also not coincided with *FoxC* transcripts at 30 h (Figure 2-figure supplement 3C-D), thus indicating that this factor is also not the downstream effector. On the other hand, *Ets1/2* expression showed a clear co-localization with *FoxC* transcripts at around 30 h (Figure 2-figure supplement 3E-F) that perfectly corresponds to the onset of myogenesis, suggesting a putative role of Ets1/2 as one of the terminal effectors of MAPK pathway acting during myoblast specification. Further perturbation experiments are needed to confirm this hypothesis.

Four members of the Forkhead family of transcription factors in the myogenic GRN model

Given the interesting temporal and spatial expression profile of *FoxY, FoxC*, *FoxL1* and *FoxF* regulators during myogenesis and the decrease in *FoxC* and *FoxF* expression observed after perturbing FGF signaling, translation-blocking MOs specific for all four Fox factors were designed and tested at a concentration of 200 μM.

The injected embryos were fixed at the larval stage for immunostaining to test MHC protein levels. In 90% of the *FoxY* morphants, the archenteron was fully elongated with signs of incomplete differentiation and the coelomic pouches were lacking, as was shown previously (33, 53). The MHC level was also severely reduced (Figure 4A-B). 80% of the *FoxC* morphants showed a similar morphological alteration; the archenteron was fully invaginated and partially compartmentalized, missing completely the coelomic pouches and MHC protein (Figure 4C). In 90% of the *FoxF* morphant larvae, a fully tripartite archenteron was present with disorganized coelomic pouches and a great decrease in the MHC level (Figure 4D). 95% of the *FoxL1* MO-injected larvae possessed a differentiated archenteron with disrupted coelomic pouches and reduced MHC protein level (Figure

4E). The phenotypes caused by FoxC, FoxF and FoxL1 MOs were confirmed by injecting a second translation-blocking MO targeting to different sequences (Figure 4-figure supplement 1). These observations reinforce the crucial role of Fox family factors in sea urchin myogenesis.

MyoD2, Tbx6 and Six1/2 are necessary for proper muscle development

MOs blocking the translation of three mesodermal markers *Tbx6*, *MyoD2* and *Six1/2* that are known to have conserved roles in muscle development were included in the study. *MyoD2* and *Tbx6* are both parts of the molecular fingerprint of the myoblasts at the late gastrula stage, which makes them good candidates as myogenic regulators (31). *Six1/2* is known to be a part of the regulatory state of the aboral mesoderm, which gives rise to two other mesodermal derivatives (pigment cells and coelomic pouches) (43). At the mid gastrula stage *Six1/2* expression partially overlapped with that of *FoxC* in the oral mesodermal domain (Figure 4-figure supplement 2). This transient expression seems to be correlated with the transcription of a late isoform whose peak of expression was seen at 42 h (Figure 4-figure supplement 3A-B). Therefore, a MO blocking the translation of the late isoform was designed and tested.

85% of the *MyoD2* morphants showed an elongated archenteron, and in most of the cases with reduced sphincter constrictions. The coelomic pouches were also not formed properly and the MHC level was dramatically reduced (Figure 4F). In 80% of the *Six1/2* morphants, the archenteron was elongated but not differentiated, the coelomic pouches were absent and the MHC level was decreased (Figure 4G). Interestingly, unlike what previously shown by using an MO against the early isoform of *Six1/2,* which negatively affects pigmentation (34), embryos injected with the MO targeted to the late *Six1/2* isoform were fully pigmented (Figure 4-figure supplement 3C-D). Finally, 95% of *Tbx6*

morphants had a milder phenotype, with no signs of malformed coelomic pouches and only a partial disrupted muscle fiber assembly as shown by MHC immunostaining (Figure 4H). Similar phenotypes were observed when injecting a second MO against *Six1/2* and *Tbx6* (Figure 4-figure supplement 4). The above experiments demonstrate an important role of MyoD2 and the late isoform of Six1/2 as well as a limited role of Tbx6 in sea urchin muscle development, thus suggesting that Tbx6 acts in a redundant manner with other factors, a known property of T-box factors (54). To test this hypothesis, double knock-down assays should be conducted.

Expression of selected mesodermal genes in *FoxY, FoxC, FoxL1, FoxF, Tbx6, MyoD2* and *Six1/2* morphants

In order to identify the myogenic gene core set and unravel in detail the inter-regulatory mechanisms that take place during myogenesis, we tested the expression of genes encoding selected mesodermal factors and signaling components in the morphants by FISH (Figure 5), quantitative PCR (qPCR) (asterisks in Figure 6) and NanoString analysis (Figure 6-source data 1 and 2). We selected three different developmental stages for our analyses that correspond to three distinct steps of myogenesis: early gastrula (30-35 h) as myoblast specification stage, a broad mid gastrula (36-44 h) and late gastrula (45-48 h) stage as intermediate and later steps of myogenesis, respectively. For NanoString analyses, we used a code set containing probes for most known NSM factors expressed at these developmental stages and a number of signaling components. The MO effects on transcript levels analyzed using the three methods were mostly consistent. For quantitative analyses using QPCR and Nanostring (Figure 6), changes in the transcript level were considered significant if the effect between control and MO injected embryos is more than two folds. Epistatic interactions were evaluated when the effects were

observed in at least two significant data points.

At the early gastrula stage, *FoxY* and *FoxC* expression was strongly reduced in *FoxY* MO-injected embryos (Figure 5A-D), suggesting that *FoxY* positively regulates itself and *FoxC.* This positive regulation continued to the mid and late gastrula stages and FoxY also positive regulated the other two Fox factors, *FoxF* and *FoxL1* (Figure 6). The expression of *Nanos,* a germ cell marker (55) that is part of the molecular signature of the myoblasts precursors at that developmental time point (31), was also downregulated in *FoxY* morphants (Figure 5E-F and Figure 6). *FoxC, FoxF* and *FoxL1* were downstream of *FoxC* at the mid gastrula stage. Expression of *FoxF* and *FoxL1* required *FoxC* input, whereas *FoxC* negatively regulated itself (Figure 5G-H and Figure 6). At the late gastrula stage, *FoxY* expression domain was expanded in the *FoxF* morphants to include the oral vegetal region of the archenteron tip, where myogenesis takes place (Figure 5I-J).

In addition to the inter-regulatory properties of the Fox factors, at the mid gastrula stage, *FoxY* also positively regulated *Pitx2* (56)*, MyoR*, a gene expressed in sea urchin mesoderm but not in the myoblasts (31), *Six1/2* and *ScratchX. FoxC* also activated *Tbx6* and repressed *Not* and *Dachshund (Dachs),* an aboral NSM gene in the sea urchin embryo (32). *Six1/2* appeared to activate *FoxL1, MyoR* and *Pitx2*, as previously suggested (57). At the late gastrula stage, *FoxY* continued to activate the aforementioned downstream factors although its effect on *Six1/2* was diminished and activated in addition *SoxE* gene. *FoxL1* gave positive inputs to *FoxF, MyoR*, *ScratchX* and *Pitx2* and repressed *Dachs*, *FoxY* and *Not. Tbx6* repressed *FoxY, Dach* and *Scl* and activated *Pitx2* and *MyoR*. Positive inputs on *Tbx6* gene from *FoxC* and on *MyoR* gene from *Six1/2* remained at the late gastrula stage. These perturbation analyses revealed complex positive and negative regulatory interactions among the four Fox factors and other transcriptional regulators, and the myogenic GRN models were formulated based on these results.

A provisional GRN model driving muscle specification in the sea urchin embryo

This detailed perturbation analysis coupled with the available high resolution transcriptional profiling during sea urchin embryogenesis (58) led to the construction of a GRN model that orchestrates sea urchin myoblast specification. Based on the examined developmental stages and regulatory states, three GRN diagrams have been illustrated using the Biotapestry software (www.biotapestry.org), one for the very early and two for the mid and late gastrula stages. The three proposed GRN models are summarized in Figure 7.

At the very early gastrula stage (28-32 h; Figure 7A), FGFA is produced in the ventrolateral ectoderm and is received by FGFR1, which is expressed in all myoblast precursor cells. FGF signaling induces myoblast specification through the MAPK/ERK pathway and a downstream effector (indicated as a circle in Figure 7A), possibly Ets1/2, and results in the activation of *FoxC* transcription. *FoxC* activation needs an additional positive input from *FoxY*. Since *FoxY* expression is excluded from the other two NSM lineages (pigment and blastocoelar cells) (31), it is reasonable to consider the existence of a Non-Skeletogenic Mesoderm Repressor (NSMR) that inhibits *FoxY* transcription in these cell populations. The existence of such a repressor has also been postulated by Materna and Davidson (2012). Here we propose that in the myoblast precursors this repressive action is blocked through a double negative gate, due to the existence of another repressor (X).

At the mid gastrula stage (40-44 h; Figure 7B), *FoxY* establishes the myoblast regulatory state by activating a variety of transcriptional regulators including all four Fox factors. *FoxC* is a broad connected hub gene and gives positive inputs to a number of factors that compose the muscle gene battery such as *FoxF, FoxL1* and *Tbx6*. *FoxL1* needs also an

additional input from *Six1/2*. Finally, *SoxE* is also part of the molecular identity of the myoblasts at the mid gastrula stage since it appears to largely co-express with *FoxC* (Figure 7-figure supplement 1A); however, the upstream factor(s) controlling *SoxE* expression remain unknown.

At the late gastrula stage (44-48 h; Figure 7C), the expression of *FoxY, Nanos, Six1/2* and *SoxE* clears from the myogenic territory and confines in other NSM domains (small micromere (SM) and aboral NSM (AB NSM domain)). This occurs due to the repressive functions of *FoxF* and *Tbx6* on *FoxY,* which results in a subsequent loss of *Six1/2, SoxE* and *Nanos* expression. Similarly, *Not* (whose expression domain relative to that of *FoxC* is reported in Figure 7-figure supplement 1B), as well as *Scl* and *Dach* receive negative inputs from *FoxL1* and *Tbx6*, respectively, that together prevent their expression in the myogenic domain. The initiation of *MHC* transcription marks the terminal differentiation state (31).

Regulatory interactions in the other three NSM domains

The analyses of the perturbation data provided additional information concerning the regulatory interactions seen in the other three NSM domains. As elsewhere stated (33), *FoxY* is at the top of the hierarchy of the NSM GRN. *Tbx6* and *FoxL1* positively regulate *Pitx2* (whose expression domain relative to that of *FoxC* is reported in Figure 7-figure supplement 1C) and together with *Six1/2* activate *MyoR*. However, the fact that *Tbx6* and *FoxL1* are not expressed in the AB NSM (31) lead us to the conclusion that their inputs on *MyoR* are indirect. *FoxL1* may also activate *Pitx2* in the oral ectoderm (OR ECT). Finally, in the SM lineage, *FoxY* is upstream of *Nanos*, *FoxC* and *Pitx2*.

DISCUSSION

FGF signaling pathway triggers myoblast specification

In this study, we propose that the muscle progenitors originate from a pool of unspecified cells located at the oral/lateral periphery of the vegetal plate at the mesenchyme blastula stage and they adapt the myogenic fate by receiving an inductive signal at the moment of gastrulation. This would further imply that all developmental decisions regarding the separation of all four NSM regulatory states (pigment, blastocoelar, muscle and coelomic pouch cells) take place during the interval between the blastula and the very early gastrula stage.

A key finding of this work was the recruitment of FGF signal in myoblast specification. FGF signaling cascade is reported to trigger the emergence and/or promote the expansion of muscle lineage progenitors in different organisms and probably possessed an ancestral role in mesoderm patterning (at least in deuterostomes), along with other developmental processes. Genes encoding FGF and FGFR were already present in the last common ancestor of all metazoans but the origin of the FGF/FGFR couple appears to be an innovation specific to the Eumetazoa, potentially linked to the increase of animal complexity (59). The acquisition of diverse roles and functions of FGF signaling occurred with the expansion of FGF and FGFR families during evolution (60). For example, vertebrates are known to have the largest number of FGF signaling components due to the two whole-genome duplication events (22 ligands and 4 receptors). FGF signals are involved in various developmental contexts (e.g. somitogenesis, cancer, gastrulation, metabolism, neural induction, etc.) (61-66). Similar situations have been observed in *Drosophila* where the moderate expansions of both FGFR (*breathless* and *heartless*) and FGF (*Pyramus* (*Pyr*), *Thisbe* (*Ths*) and *Branchless* (*Bnl*)) families resulted in different biological functions such as cell–cell interactions during mesoderm layer

formation, caudal visceral muscle formation, tracheal morphogenesis and glia differentiation (67). In Echinoderms, the two paralogues FGFR1 and FGFR2 are involved in two distinct developmental processes: myoblast specification (this study) and PMC migration (47), respectively. Diverse functions of FGF signaling seen among different animal taxa suggest that this signaling system is redeployed at different levels of GRN hierarchy and acquire new functions in developmental and physiological processes. Moreover, the hierarchical position of FGF signaling seems to depend on the level of complexity of the developmental process placing it as a 'checkpoint' in a non-conserved way regarding its downstream targets, a known property of GRN 'plug-in' devices defined previously (68). The functional importance of being attachable onto any kind of circuitry reflects the need of developmental GRNs to adopt factors that act as turn on/off apparatuses, resulting in change of the undefined cell regulatory state to their specific cell lineage.

The multilevel transcriptional regulation

An interesting finding of this analysis is the hierarchical organization of the regulatory interactions in the network topology. *FoxY*, a sea urchin specific Fox family factor, lies on the top of the GRN architecture and is the key upstream regulator of sea urchin myogenesis and coelomic pouch formation. As shown elsewhere, *FoxY* is a direct target of the D/N signal, produced in the adjacent NSM. D/N signaling is known to be effective in cells being in direct contact with the ligand producing cells (69). Therefore, the positive autoregulation of *FoxY* maintains its own expression during gastrulation when cells invaginate into the blastocoel and may be away from the signaling source.

FoxY then promotes the expression of downstream positive regulators necessary for the execution of the myogenic lineage fate. The next tier of the myogenic GRN includes

FoxC, SoxE, ScratchX and *Six1/2* factors and may be regarded as the core level of GRN. *FoxC* initially needs both positive inputs from FGF signaling and *FoxY* to be activated but then acts as an autorepressor in order to stabilize its number of transcripts. All factors belonging to this group are highly interconnected and provide multiple inputs to their downstream targets. On the same level of the GRN, immediately adjacent to the core, stands the intermediate layer of the network composed of the *FoxF* and *FoxL1* transcriptional effectors. Finally, the last level of the network includes the differentiation driver and repressor *Tbx6* that triggers the terminal myoblast differentiation. *Tbx6* seems to cooperate redundantly with other factors in an OR logic (68), which also explains the mild phenotype seen in the Tbx6 perturbations. The differentiation drivers, such as *MyoD* (31), are seen in the bottom of the GRN and probably activate the differentiation gene battery set, composed of a number of structural genes (e.g *MHC*) that account for the specific function of the muscle cell.

This highly multilevel transcriptional regulation seen in myogenesis can be interpreted as an example of 'correlated evolution' where the increase of complex processes is accompanied by the expansion of the relevant regulatory systems whilst in less complex systems a shallower regulatory system can be provided.

Recruitment of the Fox factors in the myogenic GRN

Another important outcome that arose from this analysis is the sequential inter-regulatory mechanism observed among the four Fox family members FoxY, FoxC, FoxF and FoxL1. The Fox family of transcription factors is an ancient gene family and it has been proposed that its evolutionary origin occurred in a clade of unicellular organisms (70). This family has expanded over time through multiple duplication events, and sometimes through gene loss, and resulted in over 40 members in vertebrates, grouped in 23

subclasses (71). Among them, FoxC, FoxF and FoxL subfamilies are of particular interest because they are clustered in most metazoan genomes and usually involved in mesoderm specification/differentiation processes (72-74). Moreover, a linked activation or an overlapping expression has been reported in some cases, as in vertebrates, where FoxC specifies the dorsal mesoderm and derivatives, while FoxF patterns the lateral mesoderm and derivatives (13, 75-77). In the sea urchin, where a cluster of *FoxC*, *FoxF* and *FoxL1* genes is present (31), we witness a similar regulatory logic where the different Fox factors pattern in an overlapping fashion in different compartments of the coelomic pouches. This cluster was probably expressed initially in developing mesodermal tissues and further evolved in regulating the specification and compartmentalization of mesodermal derivatives. Moreover, our findings highlight the importance of the hierarchical position of the Fox family factors in the GRN. The expression patterns and sequential activation of *FoxC, FoxF* and *FoxL1* genes in time reflect the linkage properties of the retained cluster: among the three genes, *FoxC* is the first to be turned on in the myogenic lineage and is necessary for the activation of the downstream factors *FoxF* and *FoxL1,* which are inter-regulated. The outcome of such complex inter-wiring GRN may contribute to the establishment of a more robust output, able to mask putative perturbations of single nodes, as recently proposed (78).

Exclusion of alternative NSM fates

This study provides additional insights into understanding the logic of the exclusive mechanisms that occur in the sea urchin embryo during myoblast specification (79). By 24 h post fertilization, the separation of the different NSM regulatory states is defined. The vegetal plate of the embryo consists of four distinct NSM cell populations: the aboral part that occupies about two-thirds of the total cells will differentiate into the pigment

484 cells; the oral part that consists of the remaining one-third of the total cells will become
485 the blastocoelar cells; the four SMs in the center are primordial germ cells; approximately
486 2-3 cells in the oral/lateral domain of the periphery of the vegetal plate will give rise to
487 the myoblast precursors. As showed elsewhere, the specification of the distinct NSM
488 domains depends on D/N and Nodal signals (37, 80, 81). At 9 h post fertilization, Delta
489 ligand produced by the skeletogenic mesoderm activates a D/N cascade, which
490 subsequently initiates *Gcm* and *GataE* transcription in all NSM cells (81, 82). Later, at 24
491 h, the NSM GRN becomes regionalized into distinct oral and aboral NSM GRNs in
492 consequence of Nodal signaling through the immediate expression of a Nodal target, *Not*,
493 in the oral NSM. This causes the repression of the aboral GRN in the oral NSM cells, and
494 the aboral NSM cells still express *Gcm* and *Gata*E as part of the regulatory state of the
495 pigment cell lineage (83). In the oral NSM the expression of a new suite of regulatory
496 genes is taking place that belong to the blastocoelar lineage GRN (35). Our study showed
497 that in the periphery of the oral/lateral NSM, the appearance of 2-3 unspecified cells is
498 evident. These cells will soon get specified, at 28/30 h, by receiving an inductive FGF
499 signaling at the moment of gastrulation. FGF signal reception together with the double
500 negative gate caused by the repressive action of factor X on NSMR leads to the
501 transcriptional activation of *FoxY* and *FoxC* expression. Positive inputs from *FoxY* and
502 *FoxC* genes into downstream effectors promote the recruitment of the myogenic genes
503 and lock down the myoblast regulatory state. Genes such as *Scl, Dachs, Pitx2* and *Not*
504 that belong to different NSM regulatory states are excluded from the myogenic domain
505 leaving only the muscle GRN operating. Finally, spatial repression circuits generate
506 regulatory transitions in the expression of key genes such as *FoxY, Nanos, SoxE* and
507 *Six1/2,* which now are established only in the other NSM GRNs. The sea urchin
508 myogenic GRN is a nice example of how dynamic developmental processes can be

encoded in the genome and shows clearly that understanding in depth the wiring properties of a developmental GRN model can provide a comprehensive view on the relationship between the regulatory architecture and gene expression dynamics.

Rewiring the myogenic GRN over evolutionary time

The nature of the evolutionary alterations that arise from regulatory changes depends on the hierarchical positions of these changes within a GRN. One of the most striking findings of this study concerns the paradox that genes are constantly re-used in the same context, but are rewired in different networks. Despite the conserved regulatory modules found in the system, the myogenic GRN structure has diverged extensively among animal groups. As a consequence, the level of the functional importance of the homologous transcriptional regulators (e.g. *Six1/2* in sea urchin and *Ceh-34* in *C. elegans* or MyoD in vertebrates and *Nautilus* in *Drosophila*) (this study, (84-86)), as reflected from the position within the GRN, is often diversified. New, lineage specific genes are recruited (*FoxY* in sea urchin), and 'master regulatory genes' either have been lost completely (e.g. the absence of *Pax3/7* ortholog in the sea urchin) or lose their hierarchical position within the network (e.g. *Nautilus* in *Drosophila* and *MyoD* in vertebrates), with the level of complexity to be reflected in the wiring density and in the GRN organization. However, the fact that the same factors are used over and over again in such different animal systems indicates that the modular components are somehow required for keeping their myogenic activity during evolutionary time. It seems that as transcription factor families expanded and functionally diversified during evolution, the ancestral myogenic function may have been preserved in a more distant family member, rather than the homologous gene, providing the system with several regulatory alternatives, and explaining the high degree of evolutionary plasticity of developmental GRN architecture (87).

In conclusion, this large scale GRN analysis demonstrated a necessary hierarchical role for a large number of transcriptional regulators in muscle development and explained in a rational way the core gene network that is orchestrating the specification of the myogenic lineage. Moreover, it revealed the key signaling events involved in the activation of the muscle gene battery and underlined their crucial role in transforming an unspecified cell into a specific cell type with a characteristic molecular signature. Finally, this study reinforces the importance of GRN-based approach in understanding in detail complex developmental processes by assessing the causality of the regulatory mechanisms that accompany each step of the process.

MATERIALS AND METHODS

Animal husbandry and embryo cultures

Adult *Strongylocentrotus purpuratus* were obtained from Patrick Leahy (Kerckhoff Marine Laboratory, California Institute of Technology, Pasadena, CA, USA) and housed in circulating seawater aquaria in the Stazione Zoologica Anton Dohrn of Naples. Spawning was induced by vigorous shaking of animals or by intracoelomic injection of 0.5 M KCl. Embryos were cultured at 15°C in Millipore filtered Mediterranean seawater (MFSW) diluted 9:10 (V:V) in deionized H_2O.

Perturbation experiments

FoxC, FoxF, FoxL1, MyoD2, Tbx6, FGFR1 and Six1/2 antisense Morpholino Oligonucleotides (MO) were obtained from Gene Tools (Pilomath, OR, USA) and injected at different concentrations in the presence of 0.12 M KCl. Various MO concentrations were tested and the lowest that enabled the observation of a phenotype was used for the experiments (500 μM for the FGFR1 MO and 200 μM for FoxC, FoxF,

FoxL1, MyoD2, Tbx6 and Six1/2). Second MOs for FoxC, FoxF, FoxL1, Six1/2 and Tbx6 were used to confirm the morphant phenotypes (Figure 4-figure supplement 1 and 4). As a control experiment, a Standard Morpholino Control oligo end modified with 3'-Carboxyfluorescein (control-fluo MO, Gene Tools) was injected in parallel at the same concentration as the corresponding experiments. Embryos injected with FoxY, FoxC, FoxF, FoxL1, MyoD2, Tbx6 and Six1/2 MOs displayed a normal gross morphology, similar to uninjected or fluo-control MO injected embryos up to the pluteus stage, except for the effects on the coelomic pouches and, in the case of FoxC MO, the apical organ, thus suggesting confined effects of these MOs in the expression domains of the corresponding targeted genes. To test FGFR1 MO efficacy, embryos were injected with mRNA containing the MO target sequence fused to the 5' of the *gfp*-coding sequence (500 ng/µl) with or without the MO (Figure 2-figure supplement 2). FoxY MO was kindly provided by Stefan Materna (Caltech, USA). MO sequences used in this study are listed in Supplementary File 1.

Chemical treatments

SU5402 was dissolved in DMSO and added to a final concentration of 20 µM at 26 h, 28 h, 30 h or 36 h up to the collection time. Higher concentrations than this were lethal to the embryos soon after the addition of the drug and addition of the drug after 30 h did not show any effect. U0126 was dissolved in DMSO and added to a final concentration of 10 µM at 24 h as reported previously (51). A corresponding volume of DMSO was added as controls. A table summarizing the drug treatments and the observed phenotypes is seen in Figure 2-figure supplement 1.

PCR cloning and construction of expression plasmids

583 The primers used to amplify FGFA from embryonic cDNA were designed based on the
584 gene models. 5' and 3' sequences were extended by the FirstChoice RLM-RACE Kit
585 (Ambion). The complete mRNA sequences were deposited into GenBank (Sp-FGFA,
586 HQ107979). FGFR1 was amplified based on the published sequence (U17164). Both
587 primer sets are in Supplementary File 1. To construct the dominant-negative form of
588 FGFR1, a DNA fragment containing the signal peptide, the extracellular and
589 transmembrane domains was amplified by PCR (forward primer: 5'-
590 CGGGATCCATGAGTCTGCCGCGTTGTCC-3', reverse primer: 5'-
591 CCATCGATTGTCTCGAGGGAACTCCCAC-3') and cloned into the pCS2+MT vector.
592
593 *Whole Mount In Situ Hybridization (WMISH)*
594 In situ RNA probe sequences for *FoxY, FoxC, FoxF, Ese, Nanos, Six1/2, SoxE, MHC* and
595 *Gcm* are as previously published (*FoxY*: (88); *FoxC, FoxF*: (41); *Ese*: (52); *Nanos:* (55);
596 *Gcm:* (88); *Six1/2, SoxE* and *MHC:* (31)). Labeled probes were transcribed from
597 linearized DNA using digoxigenin-11-UTP or fluorescein-12-UTP (Roche, Indianapolis,
598 IN, USA), or labeled with DNP (Mirus, Madison, WI, USA) following kit instructions.
599 For WMISH with single probe, we followed the protocol outlined in (89). Double
600 Fluorescent In Situ Hybridization (FISH) and immunohistochemistry coupled to FISH
601 were performed as described in (31). For triple FISH the third signal was developed using
602 a 488 fluorophore-conjugated tyramide (Invitrogen). Embryos were imaged with a Zeiss
603 Axio Imager M1. Fluorescent in situ hybridizations were imaged with a Zeiss 510Meta
604 confocal microscope.

605 *Whole mount immunohistochemistry*
606 Embryos were collected by gentle centrifugation and fixed in 2% paraformaldehyde in
607 PBS for 15 minutes, washed 3 times in PBST and incubated in 4% sheep serum and 1

608 mg/ml BSA in PBST for 30 min. Embryos were then incubated with a primary antibody
609 (anti-*Sp*MHC, rabbit polyclonal antibody, diluted 1:600, PRIMM, Italy or a commercially
610 available anti-P-Elk1 (Serine 383), mouse monoclonal antibody, dilution 1:100, Santa
611 Cruz Biotechnology, Santa Cruz, USA) overnight at 4°C, washed 4 times in PBST and
612 followed by another incubation in 4% sheep serum and 1 mg/ml BSA in PBST for 30
613 min. Similarly, embryos were then incubated with a secondary antibody (anti rabbit-
614 AlexaFluor 555, Invitrogen or anti mouse-HRP) diluted 1:1000 for 1 h in RT, washed 4
615 times in PBST and imaged with a Zeiss 510Meta confocal microscope.
616
617 *RNA extraction and Transcriptional profiling*
618 Total RNA was isolated from cultures of various embryonic stages, approx. 100 embryos
619 per replica. The RNA was extracted with RNAquous (Ambion). The samples were
620 treated with DNase I (Ambion) to remove DNA contamination as described by the
621 manufacturer. First-strand cDNA was synthesized from total RNA using the VILO kit
622 (Invitrogen) according to the manufacturer's protocol. Expression levels were quantified
623 using the NanoString nCounter (NanoString, UCL, London) with a custom designed
624 probe set of 40 genes (Supplementary File 1). Samples were processed according to
625 manufacturers' instructions and data processed as described previously (44). Thresholds
626 of 2- and 0.5-fold differences were chosen as significant changes (44). Some data points
627 were supplemented with results from quantitative real time PCR (qPCR) analyses. qPCR
628 was conducted as described (90), using the ViiA 7 REAL TIME PCR detection system
629 and SYBR green chemistry (Applied Biosystems, Foster City, CA, USA). The primer
630 sequences used are included in Supplementary File 1. ddCt values were calculated
631 between experiment and control embryos and converted to fold differences to be
632 comparable to the NanoString data. Fold changes were calculated using *poly-ubiquitin* as

a reference and a threshold of 2-fold difference was chosen as a significant change (91). Normalized perturbation data against control are reported in Figure 6-source data 1 and raw data are reported in Figure 6-source data 2.

Sequence and Phylogenetic analysis

The signal peptide cleavage site of FGFA was predicted by SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/). The core sequences of FGFs from different organisms were aligned using the Clustal W program, and the alignment was confirmed manually. After removing gaps, the verified alignments were used to construct phylogenetic trees with the MacVector software based on the Neighbor-joining method. Bootstrap support values were calculated by 1000 pseudoreplications. All phylogenetic trees were illustrated with the FigTree program (http://tree.bio.ed.ac.uk/software/figtree/). The phylogenetic tree for FGFRs was constructed in the same manner as the FGF tree but was based on the alignments of the tyrosine kinase domains.

ABBREVIATIONS

AB NSM, aboral non-skeletogenic mesoderm; BLAST, Basic Local Alignment Search Tool; CDS, coding sequence; DIG, digoxigenin; Dn, dominant negative mutation form; DNP, dinitrophenol; FISH, fluorescent *in situ* hybridization; h, hours post-fertilization; lv, lateral view; M, muscles; MO, morpholino antisense nucleotide; MOPS, 3-(N-morpholino) propanesulfonic acid; MHC, Myosin Heavy Chain; MRF, Myogenic Regulatory Factor; NCBI, National Center for Biotechnology Information; NSM, non-skeletogenic mesoderm; NSMR, Non-Skeletogenic Mesoderm Repressor; OR AN NSM, oral animal non-skeletogenic mesoderm; OR ECTO, oral ectoderm; OV, oral view; PBS, Phosphate Buffered Saline; PCR, Polymerase Chain Reaction; PMCs, Primary

Mesenchyme Cells; qPCR, Quantitatively PCR; SM, small micromere; VV, vegetal view; WMISH, whole mount *in situ* hybridization

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Andy Ransick, Stefan Materna and Paola Oliveri for kindly donating some of the reagents and clones used in this study, Paola Oliveri also for helping with the NanoString analysis and for useful discussions. Aaron Tolwson and Claudia Cuomo for their help with qPCR. Giovanna Benvenuto for support with confocal microscopy. Han-Ru Li for her help in morpholino injections and phalloidin staining.

REFERENCES

1. Wilm B, James RG, Schultheiss TM, Hogan BL. The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. Developmental biology. 2004;271(1):176-89.
2. Mahlapuu M, Ormestad M, Enerback S, Carlsson P. The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. Development. 2001;128(2):155-66.
3. Bentzinger CF, Wang YX, Rudnicki MA. Building muscle: molecular regulation of myogenesis. Cold Spring Harb Perspect Biol. 2012;4(2).
4. Bryson-Richardson RJ, Currie PD. The genetics of vertebrate myogenesis. Nature reviews Genetics. 2008;9(8):632-46.
5. Yokoyama S, Asahara H. The myogenic transcriptional network. Cellular and molecular life sciences : CMLS. 2011;68(11):1843-9.
6. Molkentin JD, Olson EN. Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(18):9366-73.
7. Delfini MC, De La Celle M, Gros J, Serralbo O, Marics I, Seux M, et al. The timing of emergence of muscle progenitors is controlled by an FGF/ERK/SNAIL1 pathway. Developmental biology. 2009;333(2):229-37.
8. Marcelle C, Stark MR, Bronner-Fraser M. Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. Development. 1997;124(20):3955-63.
9. Vasyutina E, Lenhard DC, Birchmeier C. Notch function in myogenesis. Cell cycle. 2007;6(12):1451-4.
10. Groves JA, Hammond CL, Hughes SM. Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish. Development. 2005;132(19):4211-22.

11. Marics I, Padilla F, Guillemot JF, Scaal M, Marcelle C. FGFR4 signaling is a necessary step in limb muscle differentiation. Development. 2002;129(19):4559-69.
12. Beiman M, Shilo BZ, Volk T. Heartless, a Drosophila FGF receptor homolog, is essential for cell migration and establishment of several mesodermal lineages. Genes & development. 1996;10(23):2993-3002.
13. Beh J, Shi W, Levine M, Davidson B, Christiaen L. FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian Ciona intestinalis. Development. 2007;134(18):3297-305.
14. Tokuoka M, Kumano G, Nishida H. FGF9/16/20 and Wnt-5alpha signals are involved in specification of secondary muscle fate in embryos of the ascidian, Halocynthia roretzi. Development genes and evolution. 2007;217(7):515-27.
15. Burdine RD, Branda CS, Stern MJ. EGL-17(FGF) expression coordinates the attraction of the migrating sex myoblasts with vulval induction in C. elegans. Development. 1998;125(6):1083-93.
16. Lo TW, Branda CS, Huang P, Sasson IE, Goodman SJ, Stern MJ. Different isoforms of the C. elegans FGF receptor are required for attraction and repulsion of the migrating sex myoblasts. Developmental biology. 2008;318(2):268-75.
17. Bertrand S, Camasses A, Somorjai I, Belgacem MR, Chabrol O, Escande ML, et al. Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(22):9160-5.
18. Green SA, Norris RP, Terasaki M, Lowe CJ. FGF signaling induces mesoderm in the hemichordate Saccoglossus kowalevskii. Development. 2013;140(5):1024-33.
19. Ciglar L, Furlong EE. Conservation and divergence in developmental networks: a view from Drosophila myogenesis. Curr Opin Cell Biol. 2009;21(6):754-60.
20. Burton PM. Insights from diploblasts; the evolution of mesoderm and muscle. Journal of experimental zoology Part B, Molecular and developmental evolution. 2008;310(1):5-14.
21. Seipel K, Schmid V. Evolution of striated muscle: jellyfish and the origin of triploblasty. Developmental biology. 2005;282(1):14-26.
22. Steinmetz PR, Kraus JE, Larroux C, Hammel JU, Amon-Hassenzahl A, Houliston E, et al. Independent evolution of striated muscles in cnidarians and bilaterians. Nature. 2012;487(7406):231-4.
23. Oliveri P, Davidson EH. Gene regulatory network controlling embryonic specification in the sea urchin. Current opinion in genetics & development. 2004;14(4):351-60.
24. Ben-Tabou de-Leon S, Davidson EH. Deciphering the underlying mechanism of specification and differentiation: the sea urchin gene regulatory network. Science's STKE : signal transduction knowledge environment. 2006;2006(361):pe47.
25. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, et al. A genomic regulatory network for development. Science. 2002;295(5560):1669-78.
26. Materna SC, Ransick A, Li E, Davidson EH. Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. Developmental biology. 2013;375(1):92-104.
27. Peter IS, Davidson EH. A gene regulatory network controlling the embryonic specification of endoderm. Nature. 2011;474(7353):635-9.
28. Burke RD. Structure of the Digestive Tract of the Pluteus Larva of Dendraster Excentricus Echinodermata, Echinoida. Zoomorphology. 1981;98: 209-25.
29. Burke RD, Alvarez CM. Development of the esophageal muscles in embryos of the sea urchin Strongylocentrotus purpuratus. Cell Tissue Res. 1988;252(2):411-7.

30. Ruffins SW, Ettensohn CA. A fate map of the vegetal plate of the sea urchin (Lytechinus variegatus) mesenchyme blastula. Development. 1996;122(1):253-63.
31. Andrikou C, Iovene E, Rizzo F, Oliveri P, Arnone MI. Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. EvoDevo. 2013;4(1):33.
32. Luo YJ, Su YH. Opposing nodal and BMP signals regulate left-right asymmetry in the sea urchin larva. PLoS biology. 2012;10(10):e1001402.
33. Materna SC, Swartz SZ, Smith J. Notch and Nodal control forkhead factor expression in the specification of multipotent progenitors in sea urchin. Development. 2013;140(8):1796-806.
34. Ransick A, Davidson EH. Cis-regulatory logic driving glial cells missing: self-sustaining circuitry in later embryogenesis. Developmental biology. 2012;364(2):259-67.
35. Solek CM, Oliveri P, Loza-Coll M, Schrankel CS, Ho EC, Wang G, et al. An ancient role for Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes. Developmental biology. 2013;382(1):280-92.
36. Wu SY, Yang YP, McClay DR. Twist is an essential regulator of the skeletogenic gene regulatory network in the sea urchin embryo. Developmental biology. 2008;319(2):406-15.
37. Sherwood DR, McClay DR. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. Development. 1999;126(8):1703-13.
38. Sweet HC, Gehring M, Ettensohn CA. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. Development. 2002;129(8):1945-55.
39. Walton KD, Warner J, Hertzler PH, McClay DR. Hedgehog signaling patterns mesoderm in the sea urchin. Developmental biology. 2009;331(1):26-37.
40. Warner JF, McCarthy AM, Morris RL, McClay DR. Hedgehog signaling requires motile cilia in the sea urchin. Molecular biology and evolution. 2014;31(1):18-22.
41. Tu Q, Brown CT, Davidson EH, Oliveri P. Sea urchin Forkhead gene family: phylogeny and embryonic expression. Developmental biology. 2006;300(1):49-62.
42. Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Developmental biology. 2006;300(1):90-107.
43. Poustka AJ, Kuhn A, Groth D, Weise V, Yaguchi S, Burke RD, et al. A global view of gene expression in lithium and zinc treated sea urchin embryos: new components of gene regulatory networks. Genome Biol. 2007;8(5):R85.
44. Materna SC, Davidson EH. A comprehensive analysis of Delta signaling in pre-gastrular sea urchin embryos. Developmental biology. 2012;364(1):77-87.
45. Lapraz F, Rottinger E, Duboc V, Range R, Duloquin L, Walton K, et al. RTK and TGF-beta signaling pathways genes in the sea urchin genome. Developmental biology. 2006;300(1):132-52.
46. McCoon PE, Angerer RC, Angerer LM. SpFGFR, a new member of the fibroblast growth factor receptor family, is developmentally regulated during early sea urchin development. The Journal of biological chemistry. 1996;271(33):20119-25.
47. Rottinger E, Saudemont A, Duboc V, Besnardeau L, McClay D, Lepage T. FGF signals guide migration of mesenchymal cells, control skeletal morphogenesis [corrected] and regulate gastrulation during sea urchin development. Development. 2008;135(2):353-65.
48. Mohammadi M, McMahon G, Sun L, Tang C, Hirth P, Yeh BK, et al. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. Science. 1997;276(5314):955-60.

796 49. Favata MF, Horiuchi KY, Manos EJ, Daulerio AJ, Stradley DA, Feeser WS, et al.
797 Identification of a novel inhibitor of mitogen-activated protein kinase kinase. The Journal
798 of biological chemistry. 1998;273(29):18623-32.
799 50. Annunziata R, Arnone MI. A dynamic regulatory network explains ParaHox gene
800 control of gut patterning in the sea urchin. Development. 2014.
801 51. Fernandez-Serra M, Consales C, Livigni A, Arnone MI. Role of the ERK-
802 mediated signaling pathway in mesenchyme formation and differentiation in the sea
803 urchin embryo. Developmental biology. 2004;268(2):384-402.
804 52. Rizzo F, Fernandez-Serra M, Squarzoni P, Archimandritis A, Arnone MI.
805 Identification and developmental expression of the ets gene family in the sea urchin
806 (Strongylocentrotus purpuratus). Developmental biology. 2006;300(1):35-48.
807 53. Song JL, Wessel GM. The forkhead transcription factor FoxY regulates Nanos.
808 Mol Reprod Dev. 2012;79(10):680-8.
809 54. Gentsch GE, Owens ND, Martin SR, Piccinelli P, Faial T, Trotter MW, et al. In
810 vivo T-box transcription factor profiling reveals joint regulation of embryonic
811 neuromesodermal bipotency. Cell reports. 2013;4(6):1185-96.
812 55. Juliano CE, Voronina E, Stack C, Aldrich M, Cameron AR, Wessel GM. Germ
813 line determinants are not localized early in sea urchin development, but do accumulate in
814 the small micromere lineage. Developmental biology. 2006;300(1):406-15.
815 56. Hibino T, Nishino A, Amemiya S. Phylogenetic correspondence of the body axes
816 in bilaterians is revealed by the right-sided expression of Pitx genes in echinoderm larvae.
817 Development, growth & differentiation. 2006;48(9):587-95.
818 57. Hibino T, Ishii Y, Levin M, Nishino A. Ion flow regulates left-right asymmetry in
819 sea urchin development. Development genes and evolution. 2006;216(5):265-76.
820 58. Materna SC, Nam J, Davidson EH. High accuracy, high-resolution prevalence
821 measurement for the majority of locally expressed regulatory genes in early sea urchin
822 development. Gene expression patterns : GEP. 2010;10(4-5):177-84.
823 59. Bertrand S, Iwema T, Escriva H. FGF signaling emerged concomitantly with the
824 origin of Eumetazoans. Molecular biology and evolution. 2014;31(2):310-8.
825 60. Itoh N, Ornitz DM. Fibroblast growth factors: from molecular evolution to roles
826 in development, metabolism and disease. Journal of biochemistry. 2011;149(2):121-30.
827 61. Dorey K, Amaya E. FGF signalling: diverse roles during early vertebrate
828 embryogenesis. Development. 2010;137(22):3731-42.
829 62. Ko BJ, Kim SM, Park KH, Park HS, Mantzoros CS. Levels of circulating
830 selenoprotein P, fibroblast growth factor (FGF) 21 and FGF23 in relation to the metabolic
831 syndrome in young children. International journal of obesity. 2014.
832 63. Naiche LA, Holder N, Lewandoski M. FGF4 and FGF8 comprise the wavefront
833 activity that controls somitogenesis. Proceedings of the National Academy of Sciences of
834 the United States of America. 2011;108(10):4018-23.
835 64. Neugebauer JM, Yost HJ. FGF signaling is required for brain left-right
836 asymmetry and brain midline formation. Developmental biology. 2014;386(1):123-34.
837 65. Oki S, Kitajima K, Meno C. Dissecting the role of Fgf signaling during
838 gastrulation and left-right axis formation in mouse embryos using chemical inhibitors.
839 Developmental dynamics : an official publication of the American Association of
840 Anatomists. 2010;239(6):1768-78.
841 66. Sandhu DS, Baichoo E, Roberts LR. Fibroblast growth factor signaling in liver
842 carcinogenesis. Hepatology. 2014;59(3):1166-73.
843 67. Muha V, Muller HA. Functions and Mechanisms of Fibroblast Growth Factor
844 (FGF) Signalling in Drosophila melanogaster. International journal of molecular
845 sciences. 2013;14(3):5920-37.

68. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010;468(7326):911-20.
69. Wang MM. Notch signaling and Notch signaling modifiers. The international journal of biochemistry & cell biology. 2011;43(11):1550-62.
70. Baldauf SL. A Search for the Origins of Animals and Fungi: Comparing and Combining Molecular Data. The American naturalist. 1999;154(S4):S178-S88.
71. Hannenhalli S, Kaestner KH. The evolution of Fox genes and their role in development and disease. Nature reviews Genetics. 2009;10(4):233-40.
72. Mazet F, Amemiya CT, Shimeld SM. An ancient Fox gene cluster in bilaterian animals. Current biology : CB. 2006;16(9):R314-6.
73. Shimeld SM, Boyle MJ, Brunet T, Luke GN, Seaver EC. Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. Developmental biology. 2010;340(2):234-48.
74. Shimeld SM, Degnan B, Luke GN. Evolutionary genomics of the Fox genes: origin of gene families and the ancestry of gene clusters. Genomics. 2010;95(5):256-60.
75. Amin NM, Shi H, Liu J. The FoxF/FoxC factor LET-381 directly regulates both cell fate specification and cell differentiation in C. elegans mesoderm development. Development. 2010;137(9):1451-60.
76. Fritzenwanker JH, Gerhart J, Freeman RM, Jr., Lowe CJ. The Fox/Forkhead transcription factor family of the hemichordate Saccoglossus kowalevskii. EvoDevo. 2014;5:17.
77. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature. 2009;462(7269):65-70.
78. Macneil LT, Walhout AJ. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome research. 2011;21(5):645-57.
79. Davidson EH. Network design principles from the sea urchin embryo. Current opinion in genetics & development. 2009;19(6):535-40.
80. Duboc V, Lapraz F, Saudemont A, Bessodes N, Mekpoh F, Haillot E, et al. Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. Development. 2010;137(2):223-35.
81. Ransick A, Davidson EH. cis-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. Developmental biology. 2006;297(2):587-602.
82. Lee PY, Nam J, Davidson EH. Exclusive developmental functions of gatae cis-regulatory modules in the Strongylocentrorus purpuratus embryo. Developmental biology. 2007;307(2):434-45.
83. Materna SC, Ransick A, Li E, Davidson EH. Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. Developmental biology. 2013;375(1):92-104.
84. Amin NM, Lim SE, Shi H, Chan TL, Liu J. A conserved Six-Eya cassette acts downstream of Wnt signaling to direct non-myogenic versus myogenic fates in the C. elegans postembryonic mesoderm. Developmental biology. 2009;331(2):350-60.
85. Balagopalan L, Keller CA, Abmayr SM. Loss-of-function mutations reveal that the Drosophila nautilus gene is not essential for embryonic myogenesis or viability. Developmental biology. 2001;231(2):374-82.
86. Olson EN, Klein WH. bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes & development. 1994;8(1):1-8.

87. Andrikou C, Arnone MI. Too many ways to make a muscle: evolution of GRNs governing myogenesis. Zoologischer Anzeiger - A Journal of Comparative Zoology. 2015;http://dx.doi.org/10.1016/j.jcz.2015.03.005.
88. Ransick A, Rast JP, Minokawa T, Calestani C, Davidson EH. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Developmental biology. 2002;246(1):132-47.
89. Minokawa T, Rast JP, Arenas-Mena C, Franco CB, Davidson EH. Expression patterns of four different regulatory genes that function during sea urchin development. Gene expression patterns : GEP. 2004;4(4):449-56.
90. Rast JP, Amore G, Calestani C, Livi CB, Ransick A, Davidson EH. Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Developmental biology. 2000;228(2):270-86.
91. Materna SC, Oliveri P. A protocol for unraveling gene regulatory networks. Nature protocols. 2008;3(12):1876-87.

FIGURE LEGENDS

Figure 1: Expression analysis of genes encoding sea urchin FGF signaling components and *FoxC* by double FISH.

FGFR1 and *FGF* were stained in green and *FoxC* in red at the very early gastrula stage (30-32 h). Nuclei were labeled blue with DAPI. Yellow circles indicated by yellow arrowheads show cells co-expressing the analyzed genes. Panels A and B are stacks of merged confocal Z sections of all three channels, while separate channels over DAPI are presented in the other panels. Insets in panels A-A'' show representative single confocal sections to confirm that the two genes are indeed expressed in the same cell. Embryos in A-A'' are seen in a lateral view along the animal-top/vegetal-down axis. Embryos in B-B" are displayed in a vegetal view. fv, frontal view; vv, vegetal view; o, oral, ab, aboral. The position of the putative unspecified myoblast precursors is indicated in figure supplement 1. Phylogenetic analyses of sea urchin FGF and FGFR protein sequences are reported in figure supplements 2 and 3, respectively. A coexpression analysis of *FGFR1* and *FoxC* at late gastrula stage (48 h) is reported in figure supplement 4.

Figure 2: Perturbation of the FGF pathway.

To analyze the phenotype of FGF perturbation, bright field images were taken with DIC. Effects on muscle formation were also tested by detection of *MHC* expression by FISH or of MHC protein localization by immunostaining on pluteus larvae (72 h). The ciliary band and gut internal cilia were stained with an anti-acetylated tubulin antibody (AcT). Panels (A-D) show the effect of SU5402 in the formation of the coelomic pouches (B) and *MHC* expression (D). Panels (E-H) show the effect of anti-FGFR1 translation MO in the formation of the coelomic pouches, MHC protein localization and gut morphology. Two representative phenotype embryos, both with impaired muscles while differing for gut sphincter formation, are reported in F (normal gut, 70% of cases) and H (reduced sphincters, 30% of cases). Panels (E, G, I and J) show the effect in MHC protein localization caused by injection of FGFR1 dominant negative RNA (DnRNA) (J). Panels (A, C) show control embryos treated with DMSO. Panel (G) shows a larva injected with a fluo-control MO and panels (E, I) show control un-injected larvae (for MO injection

controls see also Materials and methods and Figure S6). The inset in panel H is a magnified view of the cilia at the apical organ. Pictures in C, D and G-J are stacks of merged confocal Z sections. MHC was stained in red and acetylated tubulin in green. Nuclei were labeled blue with DAPI. Spicules are seen in DIC analysis as reflecting polarized light objects. All embryos are seen in frontal view except the ones in panels E, F and I that are seen in lateral view with the oral side on the right (fv, frontal view; lv, lateral view). White arrows indicate the position of cardiac sphincters, whilst yellow and red arrows show, where present, the pyloric and anal sphincters, respectively. Black lines indicate pigment cells (pc). White lines indicate muscle fibers (mf). The asterisks indicate the absence of coelomic pouches (cp). A summary of SU5402 and U0126 treatments as well as MHC protein expression analysis after MEK pathway perturbation, are reported in figure supplement 1. Control MO experiments are reported in figure supplement 2. Coexpression analysis of genes encoding putative MAPK effectors and *FoxC* as well as P-Elk protein detection are reported in figure supplement 3.

Figure 3: Spatial analysis of gene expression after FGF pathway perturbation by FISH.

FoxY (A-D) *FoxC* (E-H), *FoxF* (I-L), *MHC* (I, J), and *Ese* (A, B) transcript localization tested by FISH in control embryos (A, C, E, G, I, K) and in embryos treated with SU5402 (B, F, J) or injected with FGFR1 MO (D, H, L) (for MO injection controls see also Materials and methods and Figure 2-figure supplement 2). Panels A, B, I and J show double FISH. *FoxY* was stained in green, *FoxC* and *FoxF* in red, *MHC* in cyan and *Ese* in magenta. Nuclei were labeled blue with DAPI. Each picture is a stack of merged confocal Z sections. Yellow circles indicated by yellow arrowheads show cells co-expressing the analyzed genes. The orientation of the larvae is reported for each panel: fv, frontal view; av, animal view; lv, lateral view.

Figure 4: MHC protein detected by immunostaining after perturbation of putative myogenic regulators.

MHC protein localization was tested by immunostaining in fluo-control MO injected pluteus larvae (72 h) (A) and in embryos of the same age injected with MOs against *FoxY* (B), *FoxC* (C), *FoxF* (D), *FoxL1* (E), *MyoD2* (F), *Six1/2N* (G) and *Tbx6* (H) (for MO injection controls see also Materials and methods and Figure 2-figure supplement 2). The ciliary band and gut internal cilia were stained by immunohistochemistry with an anti-acetylated tubulin antibody. Each picture is a stack of merged confocal Z sections with MHC in red and acetylated tubulin in green. Nuclei were labeled blue with DAPI. All embryos are seen in lateral view with the oral side on the right. White arrows indicate the position of cardiac sphincters. White lines indicate muscle fibers (mf). Below each panel, statistics of muscle fiber phenotype observed are reported as normal (6-7 mf), mild (4-5 mf) or strong (0-2 mf). A coexpression analysis of *Six1/2* and *FoxC* is reported in figure supplement 1. Analysis of the temporal expression profile of two distinct Six1/2 isoforms and visualization of pigmentation after perturbing Six1/2N isoform are reported in figure supplement 2.

Figure 5: Spatial analysis of gene expression after MO perturbation of selected putative myogenic regulators by double FISH).

FoxC, FoxY, *FoxF, MHC* and *nanos* transcripts were detected by FISH in fluo-control MO injected embryos (A, C, E, G, I) and in embryos injected with MOs against *FoxY* (B, D, F), *FoxC* (H) and *FoxF* (J). All images are obtained as stacks of merged confocal Z sections. Panels G-H show double FISH. In panel G, single channels over DAPI are shown as insets. *FoxY* was stained in green, *FoxC* and *FoxF* in red, *MHC* in cyan and *Nanos* in magenta. Nuclei were labeled blue with DAPI. Yellow circles indicated by yellow arrowheads show cells co-expressing the analyzed genes. The orientation of the embryos is indicated in each panel: fv, frontal view; lv, lateral view.

Figure 6: Effects of FoxY, FoxC, FoxF, FoxL1, MyoD2, Six1/2 and Tbx6 perturbations on transcript levels of selected mesodermal genes at 44 h and 48 h.

Each diamond represents a single measurement of three independent biological experiments. Fold differences were calculated between experiments and control counts using the quantitative data obtained from the NanoString nCounter. 1 fold change represents no change; ≥ 2 indicates increased expression level significantly (blue labels); ≤ 0.5 indicates decreased expression level significantly (red labels). Asterisks indicate perturbation effects as measured in independent biological experiments by qPCR. NanoString and qPCR perturbation data normalized against controls are provided in source data 1 and raw NanoString data are provided in source data 2.

Figure 7: Schematic representation and view from all nuclei of the NSM regulatory interactions in early, mid and late sea urchin gastrulae.

On the left side, three developmental stages of the sea urchin embryo are schematized: (A) early-, (B) mid- and (C) late-gastrula stage. On the right side, the genetic interactions found within this study are summarized. Different colors are used for each domain showing exclusive regulatory states: Oral Animal NSM (OR AN NSM), salmon pink; NSM, blue; Aboral NSM (AB NSM), light blue; Small Micromere derivatives (SM), light green; Myogenic domain (M), light red; Endoderm (ENDO), yellow; Oral Ectoderm (OR ECTO), light gray. Genes are presented as horizontal thick lines and their names are reported below the thick lines. The wiring among the genes is shown with solid lines, although none of them has been demonstrated to be direct. Arrows represent positive regulation, bars represent repression and white bullets, together with the dashed lines, indicate signaling events. Genes that are expressed in more than one domain, for which the putative inputs were revealed by NanoString but not validated by spatial expression analysis, are shown on a shaded background. A coexpression analysis of several genes included in the GRN diagrams is reported in figure supplement 1. Numbers associated to inputs indicate the evidence for all interactions reported and are listed in source data 1.

FIGURE SUPPLEMENTS, SOURCE DATA AND SUPPLEMENTARY FILES

Figure 1-figure supplement 1: Three-color FISH of *Gcm, Ese* and *FoxA*.

Co-staining of *Gcm* as a marker for aboral pigment cell precursors, *Ese* for oral blastocoelar cell precursors and *FoxA* for endoderm at 26 h. *Gcm* transcript was stained in light green, *Ese* in red and *FoxA* in cyan. Nuclei were labeled blue with DAPI. Yellow circles indicate the small micromeres and white circles indicated by white arrows show single cells that do not express any of the analyzed genes. In 85% of the embryos analyzed, approximately two triple-negative cells were observed that might represent myoblast precursors. All embryos are in a vegetal view. Panels A, B and C are merged confocal stacks while panels A'-A''' depict separate channels over DAPI. Panel D is a schematic representation of the vegetal surface of a sea urchin mesenchyme blastula orientated along the oral right/aboral left (O/Ab) axis, seen from the vegetal pole. The different NSM domains identified by distinct regulatory signatures at the vegetal plate are shown in different colors as indicated in the legend. Primary mesenchyme cells are not visible as they have already ingressed into the blastocoel at this stage.

Figure 1-figure supplement 2: Phylogenetic analysis of the sea urchin FGFA protein sequence.

(A) The protein domain structure of the SpFGFA contains a signal peptide and a FGF core domain. (B) Phylogenetic analysis of FGF proteins. The tree was constructed by the neighbor joining method based on the multiple alignments of the FGF core domains from various organisms. Bootstrap values over 50 percent are shown at the branch points. Sea urchin FGFs (Sp-FgfA and Pl-FgfA) do not group with the seven vertebrate FGF families from A to G. The scale indicates the % amino acid difference with Poisson correction. GenBank accession numbers for FGFs are: acorn worm SkFgf20 like, NM_001171225; amphioxus AmphiFgf1/2, EU606032; AmphiFgf8/17/18, EU606035; AmphiFgf9/16/20, EU606036; ascidian CiFgf8/17/18, NM_001032476; CiFgf11/12/13/14, NM_001032561; Caenorhabditis elegans CeEGL17, NM_075706; Drosophila melanogaster DmPYR, AY55396; DmTHS, NM_136857; human HsFgf19, NM_005117; mouse MmFgf1, NM_010197; MmFgf2, NM_008006; MmFgf3, Y00848; MmFgf4, NM_010202; MmFgf5, NM_010203; MmFgf6, NM_010204; MmFgf7, NM_008008; MmFgf8, NM_010205; MmFgf9, U33535; MmFgf10, NM_008002; MmFgf11, NM_010198; MmFgf12, NM_183064; MmFgf13, NM_010200; MmFgf14, NM_010201; MmFgf15, NM_008003; MmFgf16, AB049219; MmFgf17, NM_008004; MmFgf18, NM_008005; MmFgf20, AB049218; MmFgf21, NM_020013; MmFgf22, NM_023304; MmFgf23, NM_022657; sea urchin Sp-FgfA, HQ107979; Pl-FgfA, EF157978.

Figure 1-figure supplement 3: Phylogenetic analysis of the sea urchin FGFR protein sequences.

(A) The protein domain structures of SpFGFR1 and SpFGFR2. The Fibronectin III domain (FN3), three Ig domains (Ig1-3), Acid Box (AB), transmembrane region (TM) and Tyrosine kinase domain are indicated. (B) For the phylogenetic analysis, the tyrosine kinase domains of FGFRs were aligned, and the tree was built in the same manner as the FGF tree. SpFGFR1 and SpFGFR2 are not orthologous to the human FGFR1 and FGFR2, respectively. GenBank accession numbers for FGFRs are: human HsFgfR1, AB208919; HsFgfR2, NM_000141; HsFgfR3, NM_000142; HsFgfR4, AY892920;

ascidian CiFgfR, NM_001044355; C. elegans CeEGL-15, NM_077441; D. melanogaster DmBreathless, NM_168577; DmHeartless, NM_169784; sea urchin SpFgfR1, NM_214537; SpFgfR2, JF499690.

Figure 1-figure supplement 4: Coexpression analysis of *FGFR1* and *FoxC* by double FISH.

Relative spatial expression domains of *FGFR1* and *FoxC* at late gastrula stage (48 h). Image is a stack of merged confocal Z sections in all channels. Inset shows representative single confocal section of the tip of the archenteron. Color code of channel association to each gene is shown in each panel. Nuclei are stained blue with DAPI. Embryo is seen in a frontal view.

Figure 2-figure supplement 1: Summary of SU5402 and U0126 treatments and MHC protein detection by immunostaining after MEK pathway perturbation.

(A) Scheme summarizing the drug treatments performed and the morphological phenotypes observed. (B, C) MHC protein localization was tested by immunostaining in (B) control pluteus larva and (C) pluteus larva treated with the MEK inhibitor U0126 as indicated in (A) by the green line. The ciliary band and gut internal cilia were stained with an anti-acetylated tubulin antibody. MHC was stained in red and acetylated tubulin in green. Nuclei were labeled blue with DAPI. All embryos are seen in lateral view with the oral side on the right.

Figure 2-figure supplement 2: Control experiments for MOs.

Panels A- D show control uninjected embryos at very early gastrula (A), late gastrula (B) and pluteus larva (C-D). Panels (E-G) show embryos injected with fluo-control MO at very early gastrula (E), late gastrula (F, H) and pluteus larva (G, I). Panels J and K show the effect in the translation of FGFR1-GFP fusion protein expression after FGFR1 MO injection (K-K') compared to controls injected with FGFR1-GFP mRNA only (J-J'). A-C, E-G, J and K are bright field images taken with DIC, E'-G', J' and K' are fluorescent images, whilst D, H and I are stacks of merged confocal Z sections. In panels D, H and I, MHC was immunostained in red, acetylated tubulin in green and nuclei were labeled blue with DAPI. Embryos in panels D and I are the same shown in Figure 2I and Figure 4A, respectively. Embryos in panels A-E, G and I-K are seen in lateral view while those in panels B, F and H are in frontal view. fv, frontal view; lv, lateral view; cp, coelomic pouches; pc, pigment cells; sp, spicules; mo, mouth; st, stomach; in, intestine.

Figure 2-figure supplement 3: Immunostaining of P-Elk and expression analysis of genes encoding putative MAPK effectors and *FoxC* by double FISH.

(A, B) Localization of P-Elk protein by immunostaining. (C-F) Spatial expression domains of (C, D) *Erg* and (E, F) *Ets1/2* with respect to *FoxC* by double FISH. Panels A-B are DIC images and panels C-F are stacks of merged confocal Z sections. Insets in panels E-F show representative single confocal sections. FoxC is red, Erg is green and Ets1/2 is cyan. Nuclei were stained blue with DAPI. Embryos in A and C panels are viewed from the vegetal pole while all the others are seen in frontal view.

Figure 4-figure supplement 1: Control experiments for MOs.
Circumesophagael muscles were tested by phalloidin staining in fluo-control MO injected pluteus larvae (72 h) (A-A') and in embryos of the same age injected with MOs against *FoxC* (B-B'), *FoxF* (C-C') and *FoxL1* (D-D') at different concentrations. Each picture is a stack of merged confocal Z sections. Phalloidin is seen in green and nuclei are labeled blue with Hoechst. All embryos are seen in lateral view with the oral side on the left. Below each panel, statistics of muscle fiber phenotype observed are reported.

Figure 4-figure supplement 2: Coexpression analysis of *Six1/2* and *FoxC* by double FISH.
Relative spatial expression domains of *Six1/2* and *FoxC* at the mid gastrula stage (42 h). Image is a stack of merged confocal Z sections in all channels. Inset shows representative single confocal section of the tip of the archenteron. Color code of channel association to each gene is shown in each panel. Nuclei are stained blue with DAPI. Embryo is seen in a frontal view.

Figure 4-figure supplement 3: The two *Six1/2* isoforms.
(A) Upstream sequence of the of the *Six1/2* gene. Two ATGs are shown in red. The upstream one, highlighted in bold, corresponds to the first ATG in the long isoform (*Six1/2N*) that is probably generated by an alternative transcription start. The downstream ATG indeed corresponds to the first ATG of the short isoform which transcription starts a few nucleotides upstream of it (Andrew Ransick, personal communication). Highlighted in different colors show the regions where the different set of qPCR primers were designed: the ones used to amplify the upstream sequence belonging to *Six1/2N* isoform only, are in yellow, while the ones used to amplify part of the homeobox domain (in bold), common to both isoforms, are highlighted in olive green. The target sequence used to design the MO against the long isoform *Six1/2N* is highlighted in cyan. (B) Temporal expression profiles of *Six1/2* distinct isoforms during sea urchin embryogenesis. Graph shows the number of transcripts per embryo during embryogenesis revealed by qPCR. Six1/2HD represents the sum of the number of transcripts of the two isoforms while Six1/2N shows only the number of transcripts for *Six1/2N*. The columns represent average of various measurements and the error bars are standard deviations. (C-D) Bright field images were taken with DIC of (C) control un-injected larva and (D) *Six1/2N* morphant (72 h) for visualizing the effect on pigmentation.

Figure 4-figure supplement 4: Control experiments for MOs.
Circumesophagael muscles were tested by phalloidin staining in fluo-control MO injected pluteus larvae (72 h) (A-A') and in embryos of the same age injected with MOs against *Six1/2* (B-B') and *Tbx6* (C-C') at a concentration of 100 µM. Each picture is a stack of merged confocal Z sections. Phalloidin is seen in green and nuclei are labeled blue with Hoechst. All embryos are seen in lateral view with the oral side on the left. Below each panel, statistics of muscle fiber phenotype observed are reported.

Figure 6-source data 1: Perturbation data derived from NanoString and qPCR analysis showing fold differences of gene expression in MO-injected embryos after normalization against controls.

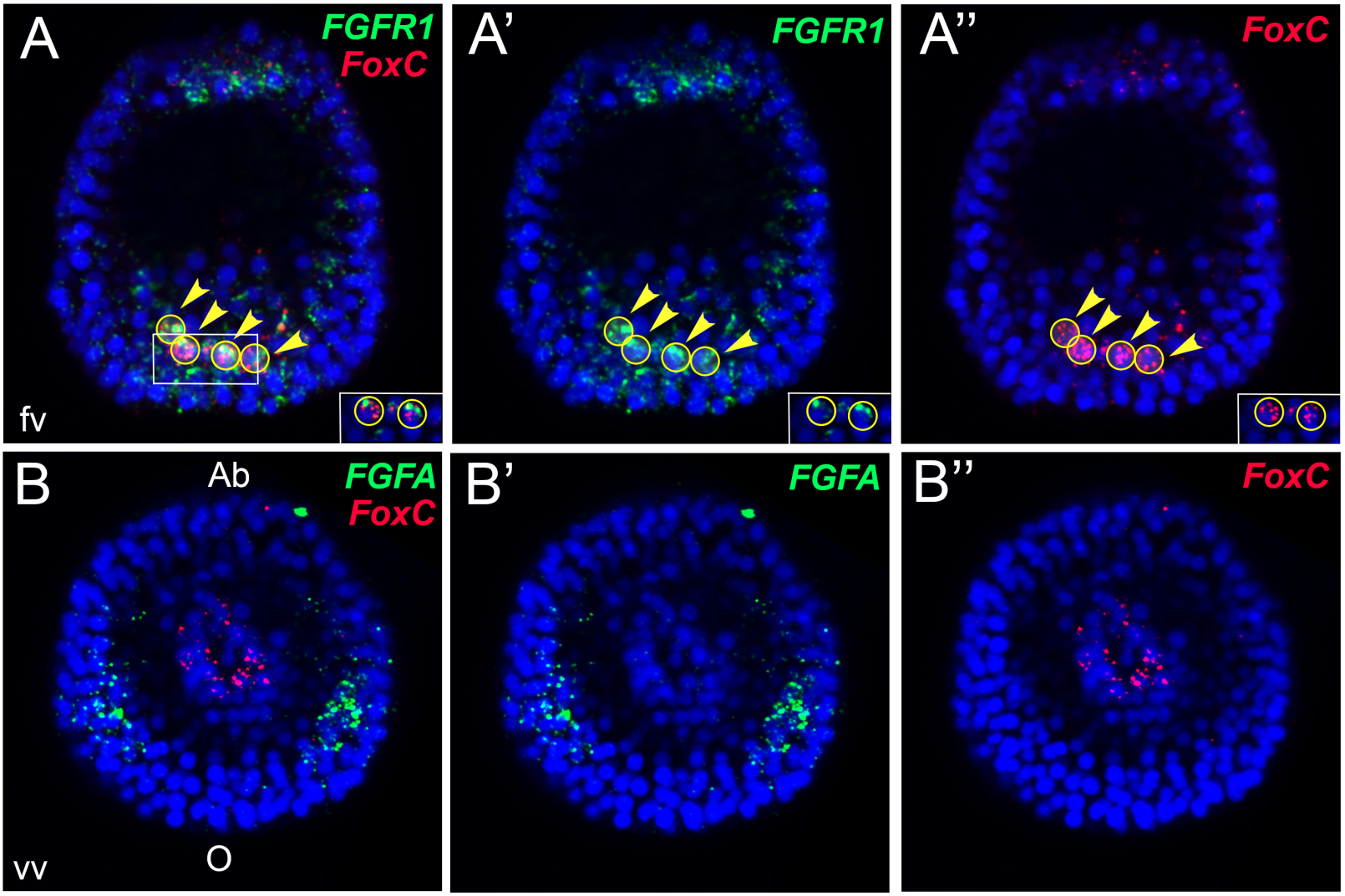
Figure 6-source data 2: Raw data derived from NanoString analysis.

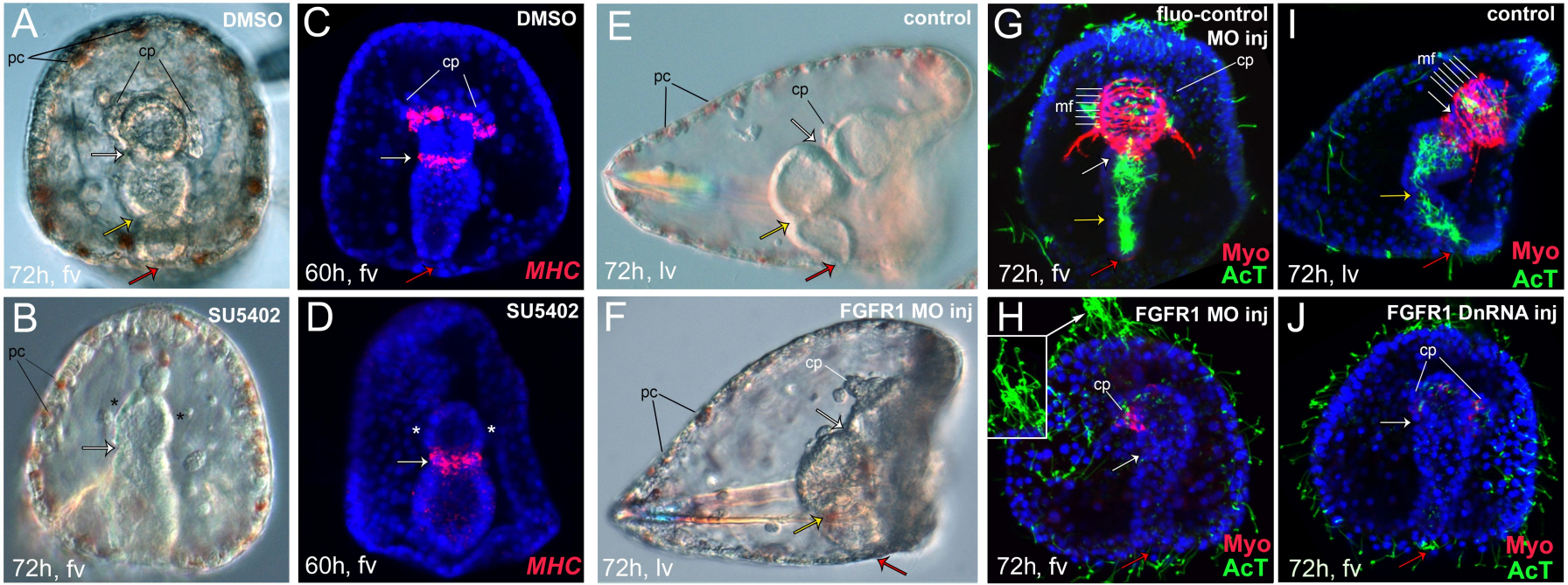
Figure 7-figure supplement 1: Coexpression analysis of genes encoding mesodermal factors by double FISH.

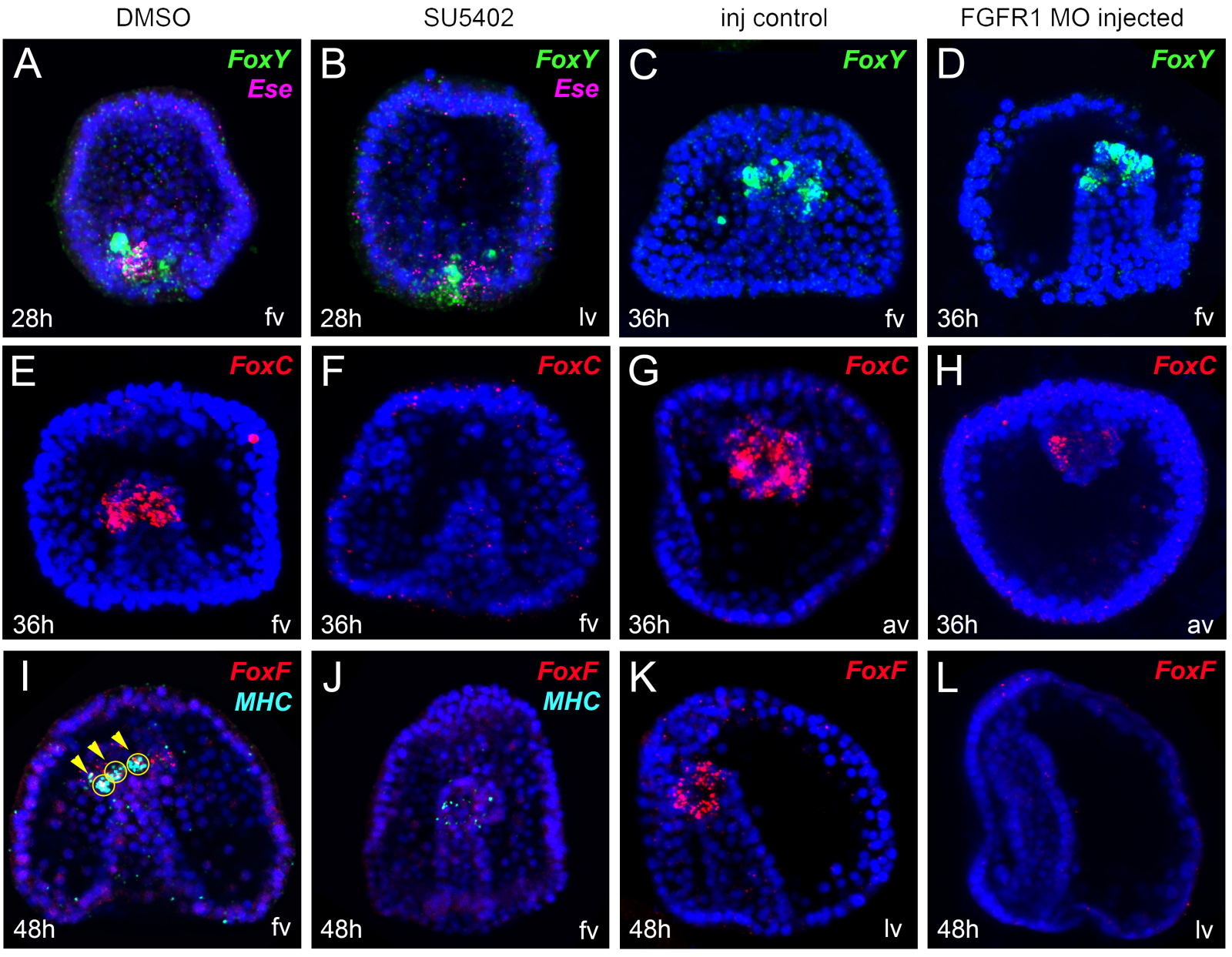
Relative spatial expression domains of (A) *SoxE* and *FoxC* at the mid gastrula stage (40 h), (B) *Not* and *FoxC* at the mid gastrula stage (44 h) and (C) *Pitx2* and *FoxC* at the late gastrula stage (48 h). Each picture is a stack of merged confocal *Z* sections in all channels. Inset in panel A shows a representative single confocal section of the tip of the archenteron. Color code of channel association to each gene is shown in each panel. Nuclei are stained blue with DAPI. All embryos are in a frontal view.

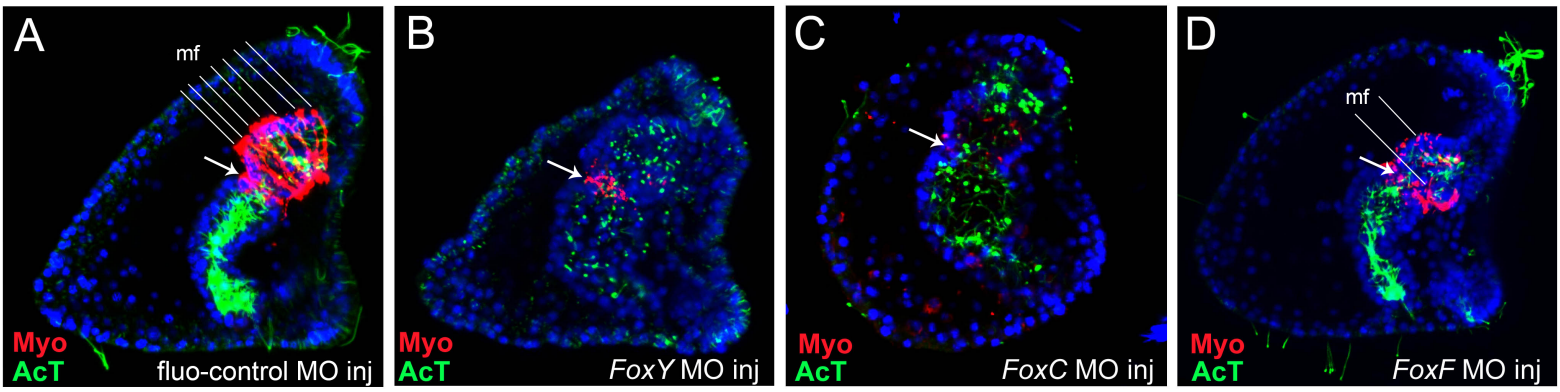
Figure 7-source data 1: Evidence for all inputs reported in Figure 7.

Supplementary File 1: Primers, MOs and nanoprobe sequences.

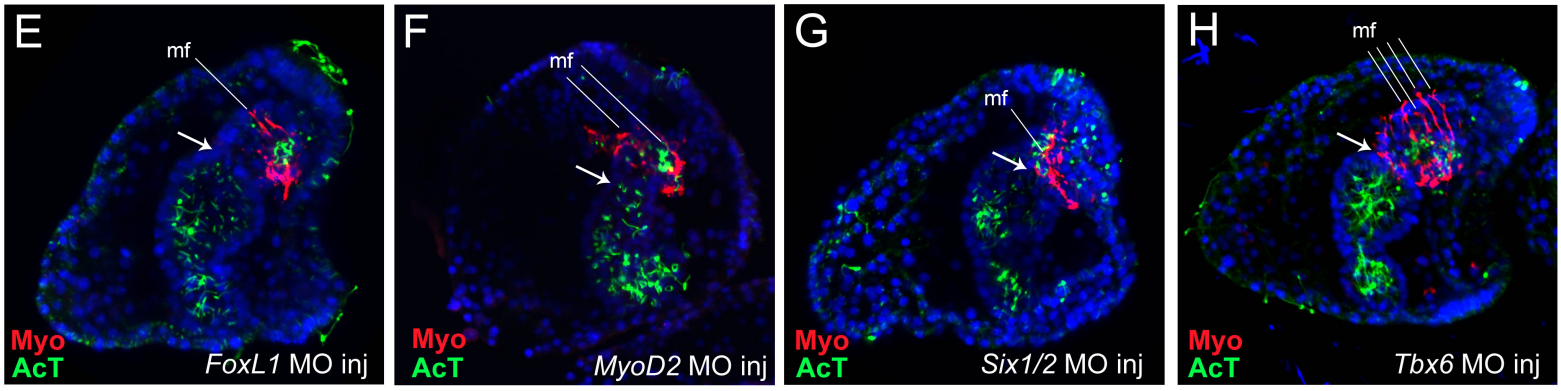

A
FGFR1
FoxC
fv
A'
FGFR1
A''
FoxC
B
Ab
FGFA
FoxC
vv
O
B'
FGFA
B''
FoxC


A
DMSO
pc
cp
72h, fv
B
SU5402
pc
72h, fv
C
DMSO
cp
60h, fv
MHC
D
SU5402
60h, fv
MHC
E
control
pc
cp
72h, lv
F
FGFR1 MO inj
cp
pc
72h, lv
G
fluo-control
MO inj
cp
mf
72h, fv
Myo
AcT
H
FGFR1 MO inj
cp
72h, fv
Myo
AcT
I
control
mf
72h, lv
Myo
AcT
J
FGFR1 DnRNA inj
cp
72h, fv
Myo
AcT

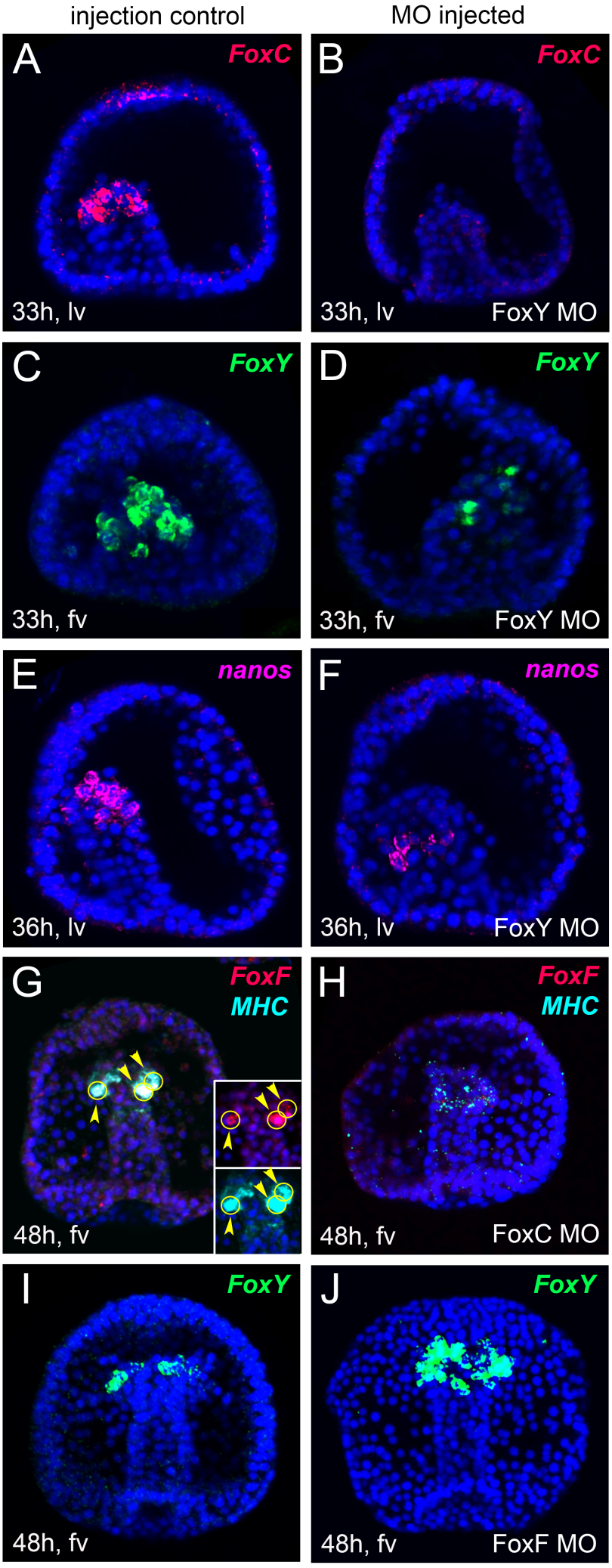

DMSO
SU5402
inj control
FGFR1 MO injected
A
FoxY
Ese
28h
fv
B
FoxY
Ese
28h
lv
C
FoxY
36h
fv
D
FoxY
36h
fv
E
FoxC
36h
fv
F
FoxC
36h
fv
G
FoxC
36h
av
H
FoxC
36h
av
I
FoxF
MHC
48h
fv
J
FoxF
MHC
48h
fv
K
FoxF
48h
lv
L
FoxF
48h
lv

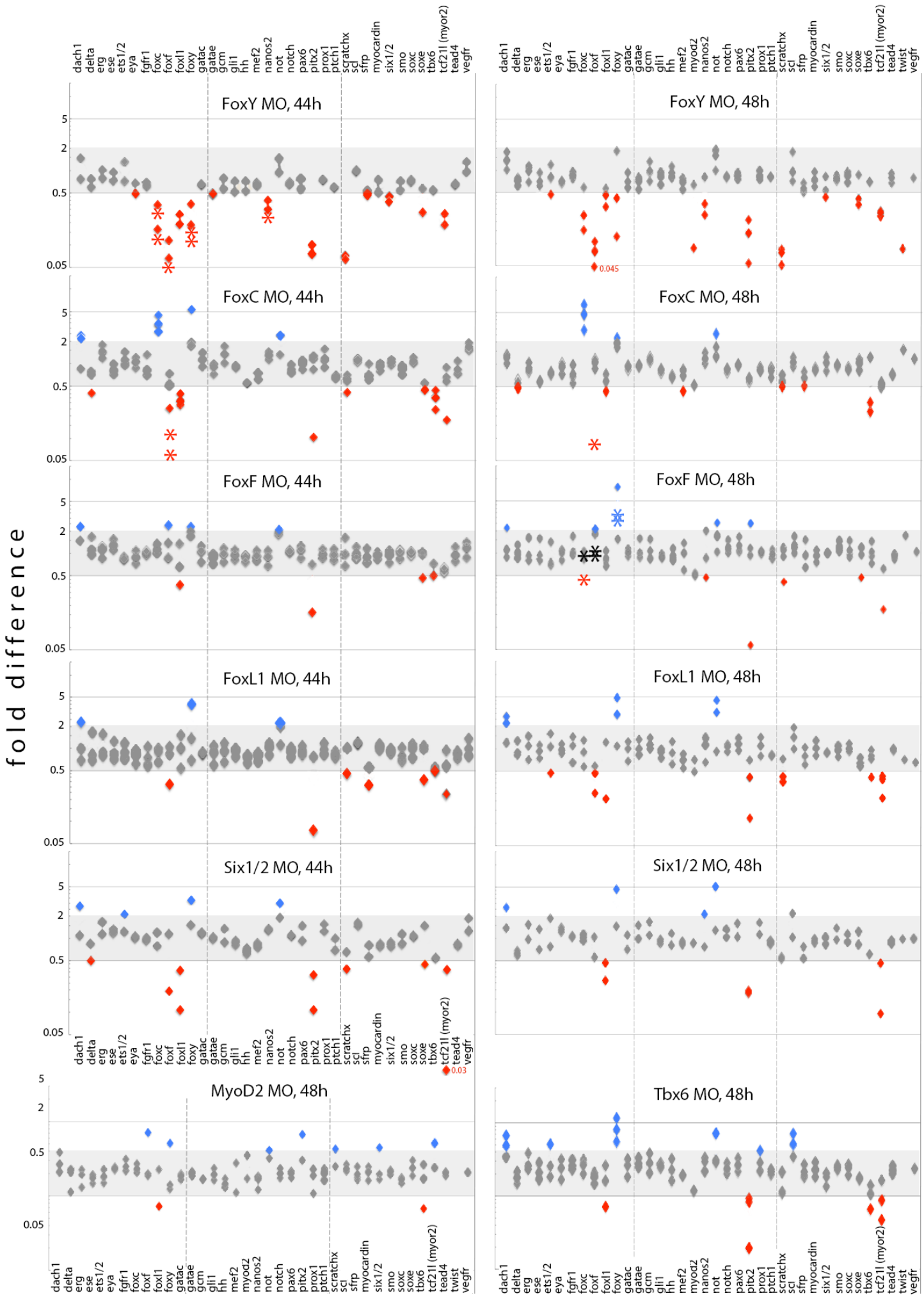


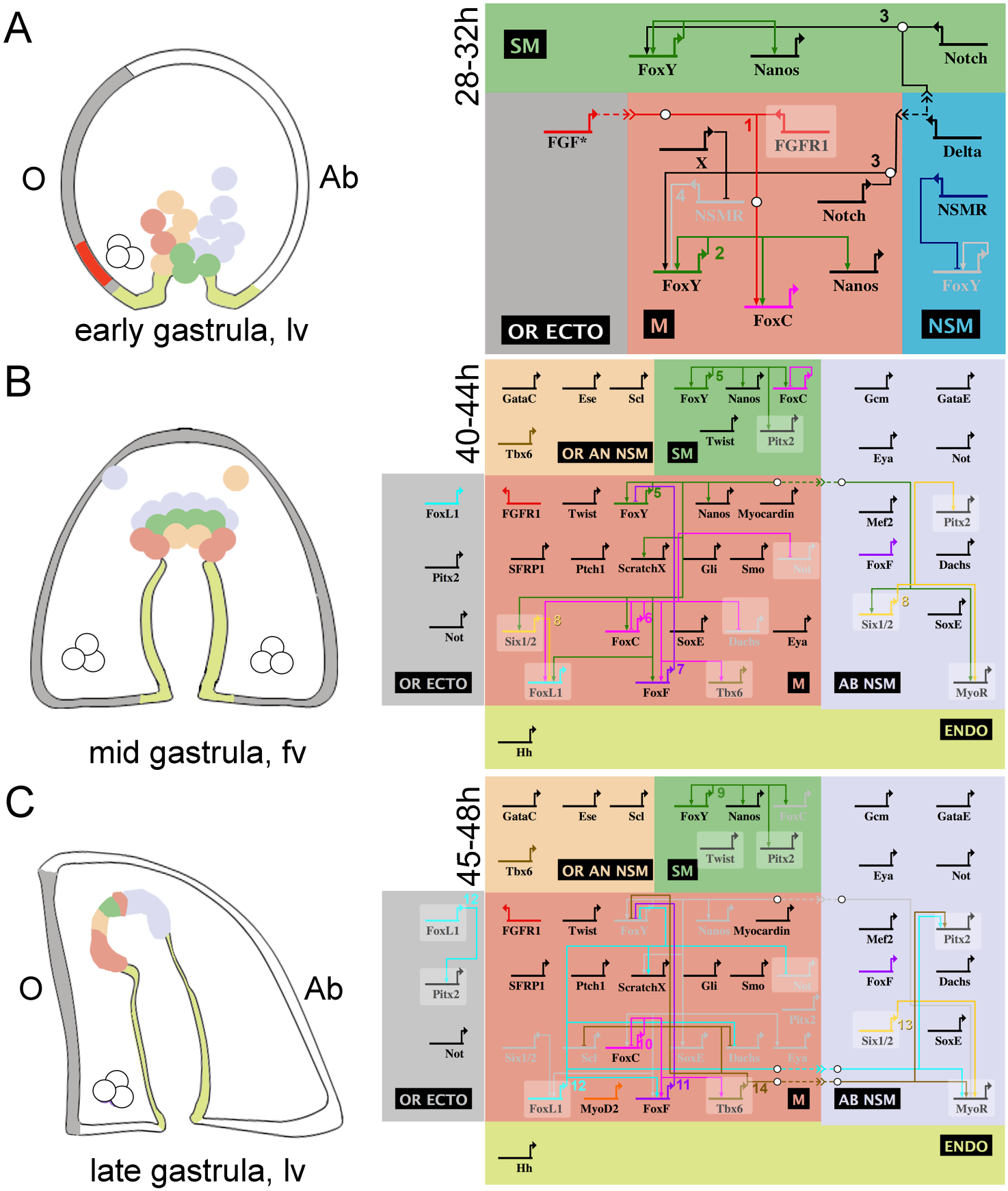
100% normal, n=198 | 10% mild, 90% strong, n=195 | 20% mild, 80% strong, n=170 | 10% mild, 90% strong, n=152



5% mild, 95% strong, n=160 | 15% mild, 85% strong, n=145 | 20% mild, 80% strong, n=157 | 5% normal, 95% mild, n=183


injection control
MO injected
A
FoxC
33h, lv
B
FoxC
33h, lv
FoxY MO
C
FoxY
33h, fv
D
FoxY
33h, fv
FoxY MO
E
nanos
36h, lv
F
nanos
36h, lv
FoxY MO
G
FoxF
MHC
48h, fv
H
FoxF
MHC
48h, fv
FoxC MO
I
FoxY
48h, fv
J
FoxY
48h, fv
FoxF MO


fold difference
FoxY MO, 44h
FoxY MO, 48h
FoxC MO, 44h
FoxC MO, 48h
FoxF MO, 44h
FoxF MO, 48h
FoxL1 MO, 44h
FoxL1 MO, 48h
Six1/2 MO, 44h
Six1/2 MO, 48h
MyoD2 MO, 48h
Tbx6 MO, 48h
5
2
0.5
0.05
0.045
0.03
dach1 delta erg ese ets1/2 eya fgfr1 foxc foxf foxl1 foxy gatac gatae gcm gli1 hh mef2 myod2 nanos2 not notch pax6 pitx2 prox1 ptch1 scratchx scl sfrp myocardin six1/2 smo soxc soxe tbx6 tcf21l (myor2) tead4 twist vegfr


A
O
Ab
early gastrula, lv
28-32h
SM
FoxY
Nanos
Notch
FGF*
X
FGFR1
Delta
NSMR
FoxY
Nanos
FoxC
OR ECTO
M
NSM
B
mid gastrula, fv
40-44h
GataC
Ese
Scl
Tbx6
OR AN NSM
FoxY
Nanos
FoxC
Twist
Pitx2
SM
Gcm
GataE
Eya
Not
FoxL1
Pitx2
Not
FGFR1
Twist
FoxY
Nanos
Myocardin
SFRP1
Ptch1
ScratchX
Gli
Smo
Six1/2
FoxC
SoxE
Dachs
Eya
FoxL1
FoxF
Tbx6
M
Mef2
Pitx2
FoxF
Dachs
Six1/2
SoxE
MyoR
AB NSM
OR ECTO
Hh
ENDO
C
late gastrula, lv
45-48h
GataC
Ese
Scl
Tbx6
OR AN NSM
FoxY
Nanos
FoxC
Twist
Pitx2
SM
Gcm
GataE
Eya
Not
FoxL1
Pitx2
Not
FGFR1
Twist
FoxY
Nanos
Myocardin
SFRP1
Ptch1
ScratchX
Gli
Smo
Six1/2
Scl
FoxC
SoxE
Dachs
Eya
Pitx2
FoxL1
MyoD2
FoxF
Tbx6
M
Mef2
Pitx2
FoxF
Dachs
Six1/2
SoxE
MyoR
AB NSM
OR ECTO
Hh
ENDO