© 2024. Published by The Company of Biologists Ltd | Development (2024) 151, dev202278. doi:10.1242/dev.202278

RESEARCH ARTICLE

Integrative multi-omics increase resolution of the sea urchin posterior gut gene regulatory network at single-cell level

Danila Voronov[1], Periklis Paganos[1], Marta S. Magri[2], Claudia Cuomo[1], Ignacio Maeso[3], Jose Luis Gómez-Skarmeta[2] and Maria Ina Arnone[1,*]

ABSTRACT

Drafting gene regulatory networks (GRNs) requires embryological knowledge pertaining to the cell type families, information on the regulatory genes, causal data from gene knockdown experiments and validations of the identified interactions by cis-regulatory analysis. We use multi-omics involving next-generation sequencing to obtain the necessary information for drafting the *Strongylocentrotus purpuratus* (Sp) posterior gut GRN. Here, we present an update to the GRN using: (1) a single-cell RNA-sequencing-derived cell atlas highlighting the 2 day-post-fertilization (dpf) sea urchin gastrula cell type families, as well as the genes expressed at the single-cell level; (2) a set of putative cis-regulatory modules and transcription factor-binding sites obtained from chromatin accessibility ATAC-seq data; and (3) interactions directionality obtained from differential bulk RNA sequencing following knockdown of the transcription factor Sp-Pdx1, a key regulator of gut patterning in sea urchins. Combining these datasets, we draft the GRN for the hindgut *Sp-Pdx1*-positive cells in the 2 dpf gastrula embryo. Overall, our data suggest the complex connectivity of the posterior gut GRN and increase the resolution of gene regulatory cascades operating within it.

KEY WORDS: Gene regulatory networks, Sea urchin, Embryo, Gut, scRNA-seq, ATAC-seq, RNA-seq, *Pdx1*

INTRODUCTION

Gene regulatory networks (GRNs) are used to describe the molecular underpinnings of developmental processes that lead to establishment of various cell and tissue types in a developing embryo. GRNs consist of two main components: genes, which are the nodes of the network, and their interactions, which are the edges of the network. Such networks allow visualization of inputs and outputs of transcription factors (TFs), signaling molecules and terminal differentiation genes in a time- and location-specific manner.

Echinoderms, and in particular, the purple sea urchin *Strongylocentrotus purpuratus* (Sp), have long been excellent experimental systems for evolutionary developmental studies (Arnone et al., 2015; McClay, 2011). Together with hemichordates, echinoderms form the Ambulacraria group, which is a phylogenetic sister group to chordates (Röttinger and Lowe, 2012). GRNs have been used to study and describe the molecular underpinnings of embryonic development. The sea urchin endomesoderm GRN, in particular, is one of the best studied (Cary et al., 2020; Ettensohn, 2020; Massri et al., 2023; Peter and Davidson, 2010). Additionally, the sea urchin embryonic gut GRN, which operates downstream of the endomesoderm GRN, from mid-gastrula up until 3 days post fertilization (dpf) pluteus larva, has also been a subject of multiple studies (reviewed by Annunziata et al., 2019), showing the crucial role of two ParaHox genes, *Sp-Pdx1* and *Sp-Cdx*, in the development of the sea urchin hindgut and pyloric sphincter.

The traditional protocol for drafting GRNs requires embryological knowledge pertaining to the cell and tissue types present in the embryo, knowledge of the regulatory genes describing the regulatory state of the tissue or cell type, causal information from perturbation experiments and validations of the GRNs through the cis-regulatory analysis, identifying the cis-regulatory modules (CRMs) of the regulatory genes (Materna and Oliveri, 2008). The process of identification of the cell types and expressed genes usually involves gene expression visualization techniques such as fluorescence *in situ* hybridization (FISH), whereas identification of CRMs is usually achieved via sequence alignment with evolutionarily closely related species (Lee et al., 2007; Livi and Davidson, 2007) to identify conserved regions that could play a role as CRMs. Such methods allow identification of only a subset of the expressed genes, usually those that are highly expressed and for which *in situ* probes were made, as well as those with a limited number of CRMs around them. However, each cell expresses multiple genes that are controlled by various CRMs (Cui et al., 2017; de-Leon and Davidson, 2010; Paganos et al., 2021); thus, high-throughput approaches are essential to build complete GRNs.

With the advent of technologies involving next-generation sequencing, such as assay for transposase-accessible chromatin with next-generation sequencing (ATAC-seq) (Buenrostro et al., 2015) and bulk and single-cell RNA sequencing (RNA-seq and scRNA-seq, respectively), high-throughput identification of CRMs and TFs bound to them (Shashikant et al., 2018), gene expression profiles (Davidson et al., 2022; Massri et al., 2021; Paganos et al., 2021, 2022c), and causal dynamics (Lowe et al., 2016; Rafiq et al., 2014) became possible. For example, Shashikant et al. (2018) used a combination of ATAC-seq and DNase I-hypersensitive site sequencing (DNase-seq) to assess chromatin accessibility in sea urchin skeletogenic cells and were able to predict a set of primary mesenchyme cell CRMs. Paganos et al. (2021) have used scRNA-seq data to identify the cell type families present in the 3 dpf sea urchin pluteus larva and to obtain the expression profiles of genes within every identified cell population at an unprecedented resolution.

Here, we present a cell type family atlas of the 2 dpf *S. purpuratus* gastrula obtained using scRNA-seq data and the set of putative gut

[1]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy. [2]Centro Andaluz de Biología del Desarrollo, CSIC/Universidad Pablo de Olavide, 41013 Sevilla, Spain. [3]Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain.

*Author for correspondence (miarnone@szn.it)

D.V., 0000-0002-2972-6484; P.P., 0000-0001-9525-4625; M.S.M., 0000-0001-5711-7304; C.C., 0009-0003-6995-0151; I.M., 0000-0002-6440-8457; J.L.G.-S., 0000-0001-5125-4332; M.I.A., 0000-0002-9012-7624

Handling Editor: Samantha Morris
Received 24 August 2023; Accepted 9 July 2024

Development (2024) 151, dev202278. doi:10.1242/dev.202278

CRMs at 2 dpf obtained with embryonic gut and whole-embryo ATAC-seq, which we combined with differential expression analyses to increase resolution of the hindgut GRN draft around the *Sp-Pdx1* gene in *Sp-Pdx1*-positive cells. We used ATAC-seq data for predicting putative CRMs and performing TF footprinting, along with scRNA-seq data to narrow genome-wide predictions to specific cell type families. Furthermore, we used the available bulk differential RNA-seq data to give causal information to the interactions within a GRN of the cells of the hindgut expressing the key gut patterning ParaHox gene *Sp-Pdx1* (Annunziata and Arnone, 2014; Cole et al., 2009). Thus, we used multi-omics datasets to obtain the information necessary for GRN drafting as per the established logic for drafting GRNs (Materna and Oliveri, 2008), and provide insight to whether the GRN interactions are predicted to be direct or not. Finally, using a combination of reporter gene-mediated cis-regulatory analysis and TF knockdown in trans, we provide *in vivo* validation of an interaction predicted by our approach between the Hox gene *Sp-Hox11/13b* and the ParaHox gene *Sp-Pdx1*.

RESULTS

Chromatin accessibility predicts putative CRMs in sea urchin gastrula nuclei

In order to identify the locations of putative CRMs (pCRMs) in the sea urchin gastrula and predict which TFs are bound to them, we generated ATAC-seq libraries for the sea urchin gastrula whole embryo and the corresponding isolated gut tissue in two replicates each (Fig. S1A,B). The consensus set of peaks was obtained by merging all the independent peaks from gut and whole-embryo ATAC-seq data, resulting in 30,866 open chromatin regions (OCRs) (Table S1). The majority of these OCRs are in the intergenic regions (40.44%) and in the introns (28.27%); promoter regions have 12.95% of the open chromatin peak set (Fig. 1A). This distribution is similar to that seen in other ATAC-seq studies (Marlétaz et al.,

Fig. 1. Open chromatin regions in the sea urchin gastrula. (A) Pie chart of the proportion of putative cis-regulatory modules (pCRMs) relative to the genome annotation features. (B) Pie chart of the proportion of transcription factor footprints in pCRMs associated with genome annotation features. (C) Pie chart of the proportion of the presence of published known hindgut CRMs in the ATAC-seq-predicted pCRMs at 2 days post fertilization (dpf). (D) Genome browser tracks of *Sp-Blimp1* and *Sp-Hox11/13b* pCRMs overlapping with published CRMs for these genes with scaled ATAC-seq coverage tracks for gut and whole-embryo datasets. (E) Genome browser tracks of *Sp-Pdx1* and *Sp-Cdx* pCRMs with scaled ATAC-seq coverage tracks for gut and whole-embryo datasets. No CRMs were previously published for these genes. TF, transcription factor; TSS, transcription start site; TTS, transcription termination site.

Development (2024) 151, dev202278. doi:10.1242/dev.202278

2018; Shashikant et al., 2018) and other methods of predicting pCRMs (Khor et al., 2019; Khor et al., 2021). We also performed TF footprinting analysis to find TFs that could be bound at these OCRs. This resulted in 279,197 footprints genome wide (Table S2), with 36.06% of them in the intergenic OCRs, 27.10% in the intronic OCRs and 17.10% in the promoter-transcription start site-associated OCRs (Fig. 1B). In order to assess whether the OCRs correspond to pCRMs, we looked for the presence of published confirmed cis-regulatory regions in our consensus peak set. To do this, we looked for overlap between the confirmed published hindgut CRMs (Arnone et al., 1998; Cui et al., 2017; de-Leon and Davidson, 2010; Lee et al., 2007; Livi and Davidson, 2007; McCarty and Coffman, 2013; Smith et al., 2008) and our 2 dpf-stage consensus peaks. We found that out of 22 confirmed published CRMs, 18 are present in the consensus peak set (Fig. 1C). For instance, our data recovered all the three known CRMs for *Sp-Blimp1* – regions CR3, CR5 and 43 (Smith et al., 2008) (Fig. 1D) – and could potentially improve the resolution of CRM prediction, as the OCRs are in many cases shorter than the known published CRMs, and CRM function could be confined to these shorter regions (Fig. 1D) (Cui et al., 2017). Thus, the consensus peak set could help identify pCRMs for the 2 dpf-stage *S. purpuratus* embryos. Some of the known CRMs were not found, such as CRM D for *Sp-Hox11/13b* (Cui et al., 2017) (Fig. 1D), which could be due to these locations being open at different stages of development; therefore, these CRMs were not further analyzed in this study. Our dataset also suggested additional regions of chromatin that could serve as previously unidentified pCRMs as exemplified by pCRMs around the key posterior gut patterning drivers *Sp-Pdx1* and *Sp-Cdx* (Annunziata and Arnone, 2014; Cole et al., 2009). Specifically, there are five pCRMs for *Sp-Pdx1* and four for *Sp-Cdx* at 2 dpf (Fig. 1E). *Sp-Pdx1* pCRMs are further explored in later sections of this study.

Diversity of cell type families in the sea urchin gastrula

In order to draft cell type-specific GRNs in the 2 dpf *S. purpuratus* late gastrula stage, we performed scRNA-seq on five samples originating from three independent biological replicates (Fig. S1C). Embryos were dissociated into single cells using an enzyme-free dissociation protocol, previously developed by our group (Paganos et al., 2021), which allows the sequencing of live and healthy cells. Isolated single cells were processed using the 10× Chromium scRNA-seq capturing system (Fig. 2A). In total, transcriptomes from 15,341 cells were included in the final analysis. Computational analysis, including data integration and Louvain graph clustering, resulted in the identification of 20 distinct cell clusters (Fig. 2A,E), corresponding to individual cell types or a set of closely related cell types.

Next, we explored the identity of the 20 identified cell clusters. Cell type family identities were assigned to each cluster based on the expression of previously described cell type markers, exploration of the total amount of genes expressed within them (Table S3), as well as taking advantage of the transcriptional signatures previously identified in the 3 dpf pluteus larva through scRNA-seq (Paganos et al., 2021) (Fig. 2B; Fig. S2). For instance, the following genes were used as specific cell type markers allowing us to recognize distinct cell type families: *Sp-Hnf6* and *Sp-Fbsl_2* (ciliary band) (Paganos et al., 2021; Poustka et al., 2004); *Sp-Frizz5/8* (anterior neuroectoderm) (Cui et al., 2014); *Sp-Spec1* and *Sp-E2_D1* (aboral ectoderm) (Amore et al., 2003; Paganos et al., 2021); *Sp-Bra* and *Sp-FoxABL* (oral ectoderm) (Paganos et al., 2021; Wei et al., 2012); *Sp-Synb* (neurons) (Burke et al., 2006); *Sp-FoxC* and *Sp-Mlckb* (myoblasts) (Andrikou et al., 2013; Paganos et al., 2021); *Sp-Nan2* and *Sp-Vasa* (small micromere descendants) (Juliano et al., 2010); *Sp-MacpfA2* (globular cells) (Ho et al., 2017); *Sp-Pks1* and *Sp-Hypp_1249* (immune cells) (Paganos et al., 2021; Perillo et al., 2020); *Sp-FcolI/II/III* (blastocoelar cells) (Paganos et al., 2021); *Sp-Msp130* and *Sp-Hypp_2386* (skeleton) (Harkey et al., 1992; Paganos et al., 2021); *Sp-Hox11/13b*, *Sp-Cdx* and *Sp-FoxI* (hindgut) (Annunziata and Arnone, 2014; Tu et al., 2006); *Sp-Chp* and *Sp-ManrC1a* (midgut) (Annunziata and Arnone, 2014; Annunziata et al., 2019); *Sp-Ptf1a* (exocrine pancreas-like precursors) (Paganos et al., 2022b; Perillo et al., 2016); and *Sp-Brn1/2/4* (foregut) (Cole and Arnone, 2009). As the ultimate goal of this study was the reconstruction of the posterior gut GRN draft and especially the one that is orchestrated by the TF Sp-Pdx1, we further explored our data by plotting for *Sp-Pdx1* as well as TFs known to pattern distinct domains of the posterior archenteron (Annunziata and Arnone, 2014; Paganos et al., 2021) (Fig. 2C). Our analysis revealed that the majority of these TFs are expressed in the two clusters that we recognized as 'Hindgut', supporting our cell type annotation. Interestingly, transcripts for *Sp-Pdx1* and, in total, eleven out of the thirteen investigated TFs were found present in the 'Hindgut (1)' cluster. In more detail, the molecular fingerprint of this cell type family, as revealed by scRNA-seq, includes the TFs Sp-FoxA, Sp-SoxC, Sp-Blimp1, Sp-Gatae, Sp-Pdx1, Sp-Rhox3, Sp-FoxI, Sp-FoxD, Sp-Cdx, Sp-Bra and Sp-Hox11/13b. Furthermore, the presence of the TFs Sp-Bra and Sp-Hox11/13b in this cluster, known to be expressed in the most posterior part of the developing archenteron, confirms that this cell type family corresponds to the most posterior hindgut domain, even though *Sp-Pdx1* and *Sp-Bra* are not expressed in the same cells of this domain, as revealed by FISH (Fig. S3). This suggests heterogeneity of the 'Hindgut (1)' cluster containing cells corresponding to more than one cell type.

Finally, to confirm the identities assigned to the cell clusters, we performed FISH on a selected set of genes with known and unknown expression patterns at this stage (Fig. 2D). The FISH results were in line with the initial predictions and our cell type annotation. Similar to our previous findings in the case of the 3 dpf *S. purpuratus* pluteus larva (Paganos et al., 2021), there was one cluster with a poorly defined molecular signature that likely represents not fully differentiated cells, which we refer to as 'Undefined'. The persistence of this cluster at both time points is indicative of this cluster representing an actual cell type rather than it being a technical artifact.

RNA-seq data uncovers the targets of Sp-Pdx1 in the developing gut

Causal information about the function of a gene is crucial to drafting GRNs. Taking advantage of the already available bulk RNA-seq data after Pdx1 knockdown at 2 dpf gastrula obtained from Annunziata and Arnone (2014), after re-analysis using an up-to-date differential expression analysis pipeline, we were able to get the list of differentially expressed genes (DEGs) between the untreated sea urchin embryos and those that were injected with *Sp-Pdx1* morpholino oligonucleotides (MOs) (Cole et al., 2009). The re-analysis of this dataset allowed us to provide additional information regarding these libraries. Principal component analysis performed on the untreated and injected groups of embryos showed that 97% of variance between the samples could be explained by the injection of the *Sp-Pdx1* MO (Fig. S1D), which indicates that the difference in gene expression is due to the *Sp-Pdx1* perturbation and, thus, the dataset can be used to infer causal information downstream of Sp-Pdx1. In line with previous findings, our differential expression analysis identified a large number of deregulated genes totaling to 2680 DEGs with an adjusted *P*-value of less than 0.05 (Table S4). Of these, 1661 were upregulated after *Sp-Pdx1* perturbation and 1019

Development (2024) 151, dev202278. doi:10.1242/dev.202278

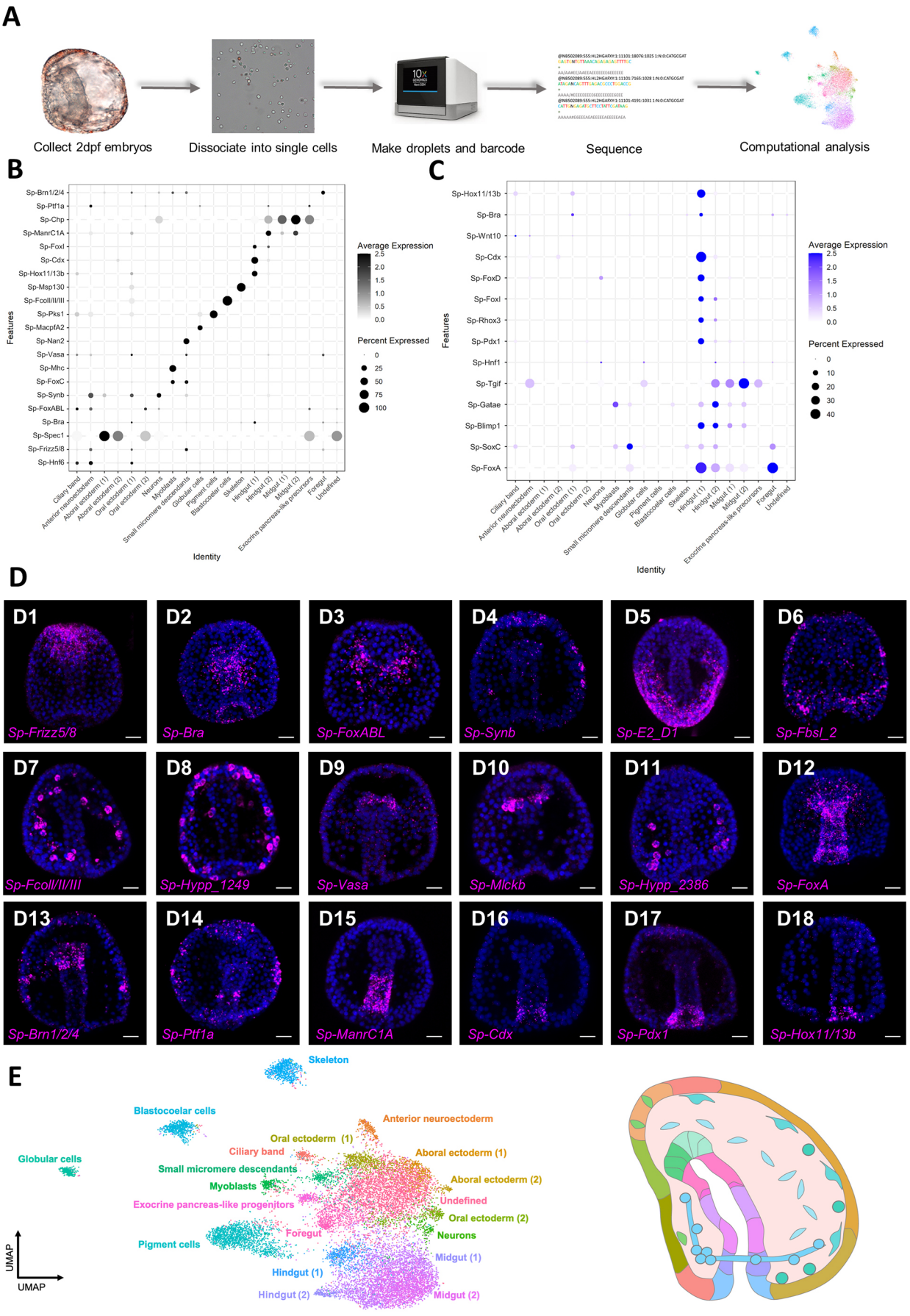


Fig. 2. See next page for legend.

Development (2024) 151, dev202278. doi:10.1242/dev.202278

Fig. 2. Cell type family atlas of the 2 dpf *Strongylocentrotus purpuratus* gastrula stage. (A) General scheme of scRNA-seq data generation and analysis. (B) Dot plot of marker genes used for determining cluster identities. (C) Dot plot of known hindgut genes, highlighting expression in the 'Hindgut (1)' cluster. In B,C, 'percent expressed' is the percentage of cells in the cluster expressing the gene and 'average expression' is the z-score-scaled average expression measured in transcripts per 10,000. (D) Fluorescence *in situ* hybridization of *S. purpuratus* 2 dpf gastrula embryos using antisense probes for *Sp-Frizz5/8* (D1), *Sp-Bra* (D2), *Sp-FoxABL* (D3), *Sp-Synb* (D4), *Sp-E2_D1* (D5), *Sp-Fbsl_2* (D6), *Sp-FcolIII/III* (D7), *Sp-Hypp_1249* (D8), *Sp-Vasa* (D9), *Sp-Mlckb* (D10), *Sp-Hypp_2386* (D11), *Sp-FoxA* (D12), *Sp-Brn1/2/4* (D13), *Sp-Ptf1a* (D14), *Sp-ManrC1A* (D15), *Sp-Cdx* (D16), *Sp-Pdx1* (D17) and *Sp-Hox11/13b* (D18). All embryos are oriented in the oral view, except the embryos shown in panels D7 and D14, which are oriented in the lateral view, and the embryo in D5, which is placed in the dorsal view. Nuclei are depicted in blue (DAPI). Images are representative of three biological replicates with at least 150 embryos per replicate. Scale bar: 20 µm. (E) Uniform Manifold Approximation and Projection (UMAP) plot of identified cell clusters in 2 dpf gastrula, with the clusters highlighted in a schematic representation of 2 dpf sea urchin gastrula using color coding.

were downregulated (Fig. 3A). According to scRNA-seq data, *Sp-Pdx1*-positive cells are found in the hindgut [clusters 'Hindgut (1)' and 'Hindgut (2)'] and in the midgut [clusters 'Midgut (1)' and 'Midgut (2)'], and a low number of cells are found in the clusters corresponding to 'Oral ectoderm (2)' and 'Ciliary band' (Fig. 3D). Notably, the average scaled expression of this gene in clusters other than 'Hindgut (1)' was less than 0.5 (Fig. 2C; Table S3). The cell type families of DEGs could be inferred using the scRNA-seq data (Fig. 3C). The downregulated genes, i.e. those that are activated by Sp-Pdx1 in the untreated embryos, belonged to gut cell families, especially to the hindgut cell clusters, whereas the upregulated genes belonged mostly to the anterior neuroectoderm and ciliary band cell clusters, although there were some genes in the anterior neuroectoderm that were downregulated after *Sp-Pdx1* MO injection (Fig. 3B,C). Consequently, the RNA-seq data shed light on what genes are downstream of Sp-Pdx1 and whether they are upregulated or downregulated by this TF. This information can be used to iteratively predict causal information of the interactions suggested by the combination of ATAC-seq and scRNA-seq datasets.

Gastrula hindgut GRN

Our omics data allow drafting GRNs for an individual cell type in the 2 dpf *S. purpuratus* gastrula. For instance, the complete GRN for the 2 dpf gastrula hindgut can be drafted, in particular, the interactions among various TFs involved in building the embryonic hindgut, with 512 interactions involving 91 individual TFs (Fig. S4A; Table S5). This GRN can be narrowed down, keeping only the nodes that are co-expressed with *Sp-Pdx1* in the same cells (Fig. S4B; Table S6).

Previously, the core of the hindgut GRN was reconstructed by Annunziata and Arnone (2014) and a global genome view of the GRN operating within the nuclei of these cells was reviewed by Annunziata et al. (2019). Our combinatorial data allow us to update and refine the published GRN for the hindgut region and reconstruct the GRN for *Sp-Pdx1*-positive cells at higher resolution. The GRN presented in Annunziata et al. (2019) (Fig. 4A), which represented the most complete version of the posterior gut at that time, shows possible interactions among the different genes comprising the GRN. However, with the available information at that point, it was unclear whether these interactions could be direct or indirect. Using our datasets towards answering the very same question, we found that most of the interactions appear to be indirect and each gene is wired through one or more intermediate nodes. Below, we report several cases for which we present a refined and/or alternative gene regulatory connectivity.

Sp-Pdx1

Sp-Pdx1 was shown to control *Sp-Cdx* (Fig. 4A); however, the actual link between the two remained an open question. Our data suggest that this interaction is likely indirect as there are no Sp-Pdx1 TF-binding sites above the threshold near *Sp-Cdx*. This interaction may instead be routed through either Sp-FoxD or Sp-Osr (Fig. 4B) as revealed by this study.

Sp-Hox11/13b

Sp-Hox11/13b was previously shown to have a self-repressive loop and an input on *Sp-Cdx* (Fig. 4A). The combinatorial data, again, favor the indirect nature of the self-control loop. This loop can instead be wired through Sp-Pdx1, which represses *Sp-Hox11/13b* expression. Additionally, the effect Sp-Hox11/13b on *Sp-Cdx* can be explained through Sp-Hox11/13b directly activating *Sp-Osr*, which then in turn affects *Sp-Cdx* directly. We also identified a direct input of Sp-Hox11/13b on *Sp-Pdx1*, which was not reported before (Fig. 4C).

Sp-Cdx

Similarly to Sp-Hox11/13b, Sp-Cdx also has a self-regulatory loop (Fig. 4A). In this case, our combinatorial analyses also suggest multiple equivalent indirect paths from Sp-Cdx back to *Sp-Cdx*, which is in line with the *Sp-Cdx* self-regulation previously reported, such as through Sp-Runt1 to Sp-Blimp1 onto *Sp-FoxD* and then to *Sp-Cdx* (Fig. 4D).

Sp-Gatae

Sp-Gatae was shown to have an activatory effect on *Sp-Hox11/13b* and on itself (Fig. 4A). Our data indicate that these effects are also indirect and that the path to *Sp-Hox11/13b* can go through upregulating *Sp-Blimp1*, which itself upregulates *Sp-Hox11/13b*. The effect on *Sp-Hox11/13b* can also be explained through the same intermediates that are likely to account for the self-feedback loop of Sp-Gatae such as Sp-Irf and Sp-SoxC (Fig. 4E).

Sp-Blimp1

Sp-Blimp1 was previously shown to affect *Sp-Pdx1* (Fig. 4A), and it is likely that this interaction occurs through a single intermediate, either Sp-Tcf or Sp-Hox11/13b or possibly both. Thus, this interaction between Sp-Blimp1 and *Sp-Pdx1* is also not direct (Fig. 4F).

Sp-FoxA

In the Annunziata et al. (2019) GRN, *Sp-FoxA* has many outputs: Sp-FoxA, Sp-Cdx, Sp-Hox11/13b and Sp-Bra (Fig. 4A). The input on *Sp-Cdx* could be a direct one, as supported by TF *in silico* footprinting. The data for the rest of the interactions suggest that they are indirect and happen through intermediates. Sp-FoxA is likely to affect itself through a loop via Sp-Hb9, whereas the input on *Sp-Hox11/13b* can be explained via Sp-Osr or Sp-Runt1 to Sp-Tcf activations. Sp-Tcf, eventually, has a direct input on *Sp-Hox11/13b* (Fig. 4G).

Sp-Bra

Sp-Bra is not expressed in the hindgut cells expressing *Sp-Pdx1* at the 2 dpf gastrula stage (Fig. S3), but if all the cells of the 'Hindgut (1)' cluster are included then the wiring of *Sp-Bra* can also be recovered (Fig. S5). Our data confirm the direct input of Sp-Bra on *Sp-FoxA* with an Sp-Bra footprint proximal to the *Sp-FoxA* gene

Development (2024) 151, dev202278. doi:10.1242/dev.202278

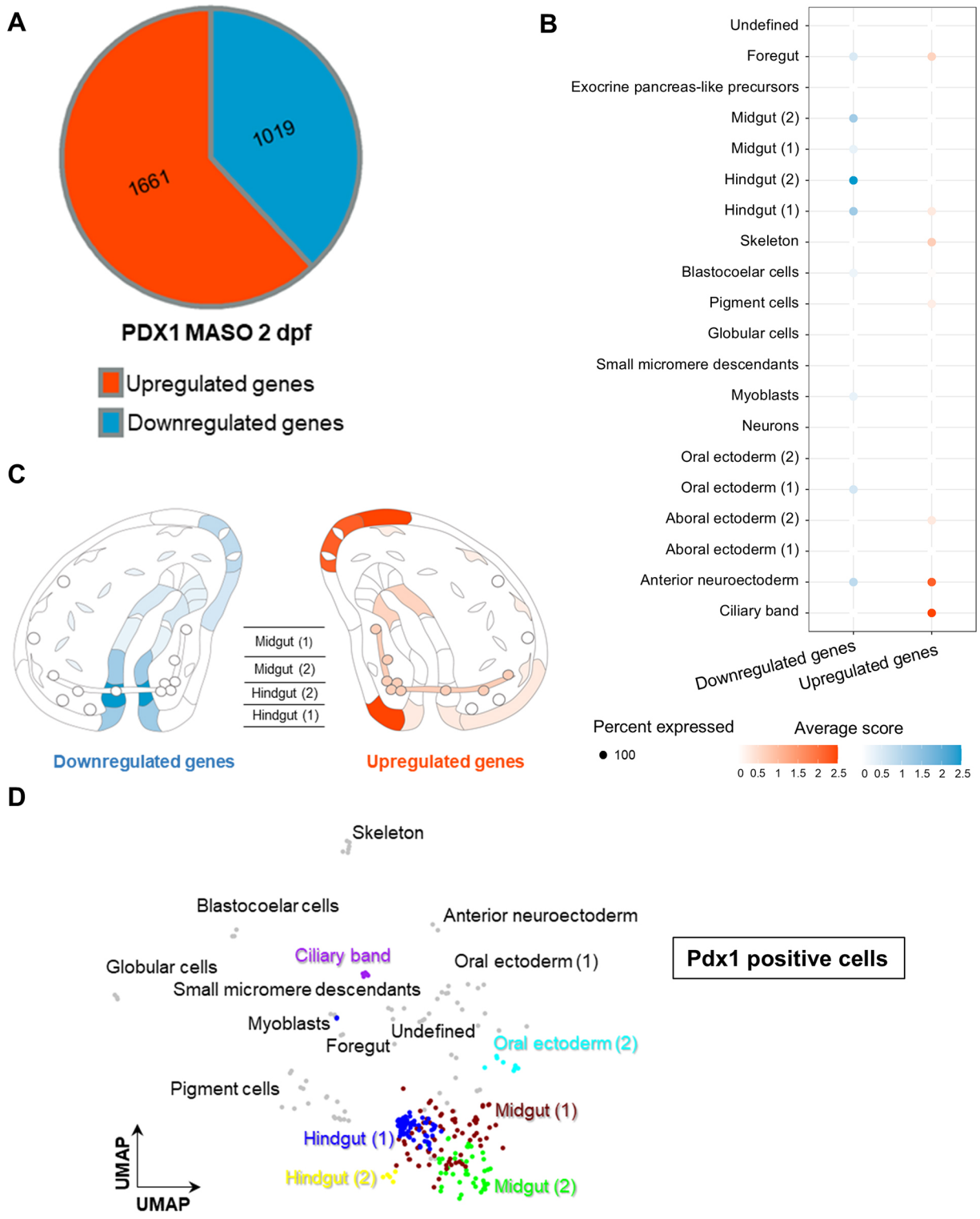


Fig. 3. Differential RNA-seq of *Sp-Pdx1* knockdown embryos. (A) Pie chart of the proportion of genes significantly upregulated and downregulated after *Sp-Pdx1* knockdown by morpholino antisense oligonucleotide (MASO). (B) Dot plot of cell clusters to which upregulated and downregulated genes belong. 'Percent expressed' is the percentage of cells in the cluster expressing the gene and 'average expression' is the z-score-scaled average expression measured in transcripts per 10,000. (C) Diagrams of sea urchin 2 dpf gastrulae highlighting the localization of upregulated and downregulated genes. (D) UMAP plot of cells within the identified cell clusters at 2 dpf that express *Sp-Pdx1*.

(de-Leon and Davidson, 2010; Schwaiger et al., 2022), whereas the Sp-Bra effect on *Sp-Cdx* is likely indirect and could go either through Sp-FoxA or Sp-Osr as intermediates. In addition, our data at 2 dpf suggest that the inputs of Sp-FoxA on *Sp-Bra* and on itself are indirect, the effect on *Sp-Bra* could be wired through Sp-Tcf and Sp-Cdx, Sp-Runt1 and Sp-Osr intermediates, whereas self-feedback loop of Sp-FoxA can be explained by a loop through Sp-Tcf and Sp-Bra (Fig. S5).

Development (2024) 151, dev202278. doi:10.1242/dev.202278

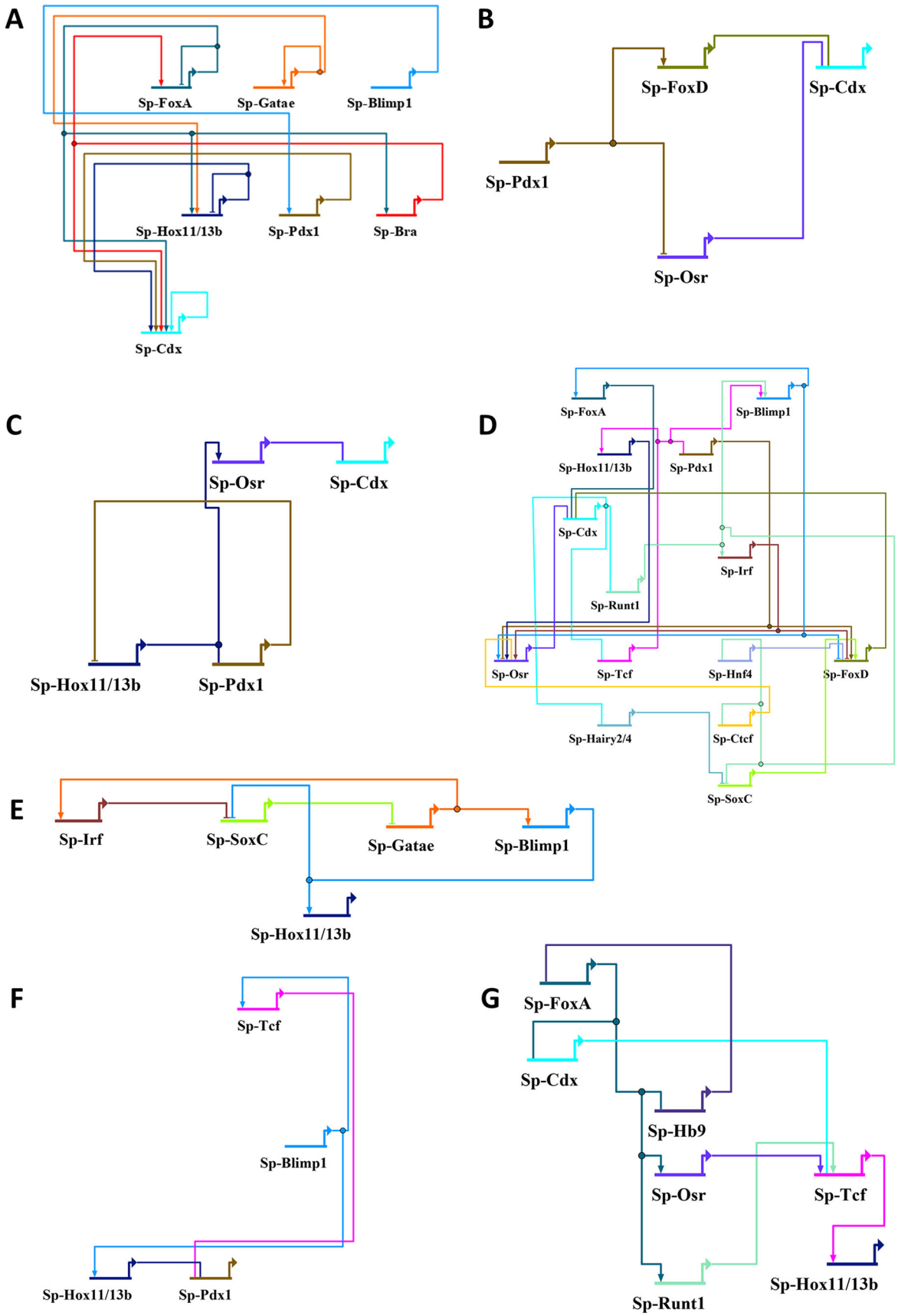


Fig. 4. 2 dpf *S. purpuratus* gastrula stage hindgut gene regulatory networks. (A) Global genome view of the core transcription factors operating in the 2 dpf hindgut gene regulatory network (GRN) as reviewed by Annunziata et al. (2019). (B-G) *In silico*-drafted GRNs connecting Sp-Pdx1 (B), Sp-Hox11/13b (C), Sp-Cdx (D), Sp-Gatae (E), Sp-Blimp1 (F) and Sp-FoxA (G) with downstream genes of the core GRN.

Development (2024) 151, dev202278. doi:10.1242/dev.202278

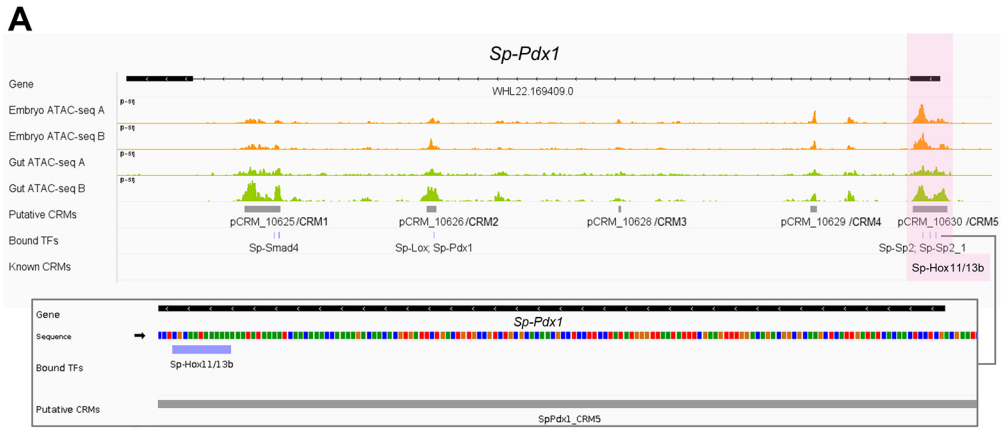
Altogether, the presented omics datasets not only allow reconstruction of a GRN draft without prior knowledge, but also empower better resolution of the interactions identified previously by supplementing additional information on whether these interactions are direct or indirect and by identifying the intermediate nodes within cells of interest through combinatorial analyses.

CRM5 allows control of *Sp-Pdx1* by Sp-Hox11/13b

The identified direct input of Sp-Hox11/13b on *Sp-Pdx1* was not reported previously. Consequently, we focused our attention on exploring this input. Our CRM predictions and the TF footprinting indicate a binding site for Sp-Hox11/13b in pCRM5, which is in the 5′ region of the *Sp-Pdx1* gene and overlaps with its first exon (Fig. 5A). To determine whether this pCRM is indeed an active cis-regulatory element, a GFP reporter construct with this element was microinjected into *S. purpuratus* zygotes. The microinjected embryos were allowed to develop until the 2 dpf gastrula stage and then GFP expression driven by this pCRM was visualized. At 2 dpf, the CRM5-GFP-tag construct exhibited expression consistent with known Pdx1-positive domains including the lateral neurons (Fig. 5B,G) and the hindgut regions (Fig. 5C,G), as well as other regions of the gut such as the future stomach (Fig. 5D,G). At the 3 dpf pluteus stage, this CRM is capable of driving GFP expression in the predominant Pdx1-expression domain in the larval digestive tract, specifically, in the pyloric sphincter (Fig. 5E). To determine whether Sp-Hox11/13b could affect *Sp-Pdx1* expression through pCRM5, *Sp-Hox11/13b* MO, previously shown to strongly downregulate Sp-Hox11/13b function (Arenas-Mena et al., 2006), was co-injected with the CRM5-GFP reporter construct. At 2 dpf, the presence of the *Sp-Hox11/13b* MO decreased the overall reporter construct expression, in addition to greatly inhibiting the GFP expression in the endodermal regions, such as the hindgut (Fig. 5F,G), with a more than tenfold decrease in the percentage of endodermally expressing embryos (Fig. 5G). This suggests that Sp-Hox11/13b controls *Sp-Pdx1* expression in the hindgut through *Sp-Pdx1* CRM5.

Multi-omics increases hindgut GRN draft resolution

Combinatorial omics and *in vivo* validations allow drafting an updated version of the GRN published in Annunziata and Arnone (2014) within the cells that express the Sp-Pdx1 TF (Fig. 6). Chromatin accessibility information allowed us to identify pCRMs and TF-binding sites within them, whereas single-cell data allowed us to filter these to contain only genes from cells that express the *Sp-Pdx1* gene. Such filtering to cells expressing a particular TF allowed us to assess its role in controlling downstream genes in the same cells, when adding the bulk RNA-seq data after *Sp-Pdx1* knockdown. This gives a detailed and novel GRN draft for a particular subset of hindgut cells expressing *Sp-Pdx1* cells. Notably, these cells express all the TFs present in the GRN reviewed in Annunziata et al. (2019), except *Sp-Bra*. The reason for this is that, at 2 dpf, *Sp-Bra* is expressed in a different subset of 'Hindgut (1)' cells and is not co-expressed with *Sp-Pdx1*; thus, it was excluded from the updated GRN (Fig. 6; Fig. S3). *Sp-Hox11/13b* MO-injected embryos showed a decrease in hindgut expression of a putative *Sp-Pdx1* CRM5-GFP construct that normally shows an *Sp-Pdx1*-like expression pattern. This allowed us to validate the predicted possible cis-regulatory direct effect of Sp-Hox11/13b on *Sp-Pdx1* via *Sp-Pdx1* CRM5 (Fig. 6, green diamond). The cis-regulatory effect of Sp-Tcf on *Sp-Hox11/13b* and *Sp-Blimp1* has been



G	Total embryos	Total fluorescent	Ectoderm only	Endoderm only	Ectoderm & Endoderm
SpPdx1_CRM5	175	67	26	27	14
Percentage of total embryos	100%	38.3%	14.9%	15.4%	8%
Sp-Pdx1_CRM5 + Sp-Hox11/13b MO	212	16	13	3	0
Percentage of total embryos	100%	7.5%	6.1%	1.4%	0%

Fig. 5. *Sp-Pdx1* CRM5 control of *Sp-Pdx1*. (A) Genome browser tracks of *Sp-Pdx1* pCRMs, with scaled ATAC-seq coverage tracks for gut and whole-embryo datasets, indicating the position and sequence of the Sp-Hox11/13b-binding site within CRM5. The zoom-in shows the Sp-Hox11/13b-binding site sequence using color coding of nucleotides: A is green, T is red, C is blue and G is orange. (B-F) *Sp-Pdx1* CRM5-driven GFP expression in the lateral neurons of the 2 dpf gastrula stage (B), in the hindgut of the 2 dpf gastrula stage (C), in the midgut of the 2 dpf gastrula stage (D), in the midgut of the 3 dpf pluteus (E) and in the 2 dpf gastrula stage, indicating absence of endodermal expression after Sp-Hox11/13b knockdown by morpholino oligonucleotide (MO) (F). Dotted regions in B,C,F indicate the outline of the embryo and the embryonic gut. Scale bars: 20 µm. (G) Embryo scoring table showing the effect of *Sp-Hox11/13b* knockdown on the *Sp-Pdx1* CRM5-driven GFP expression in different germ layers.

Development (2024) 151, dev202278. doi:10.1242/dev.202278

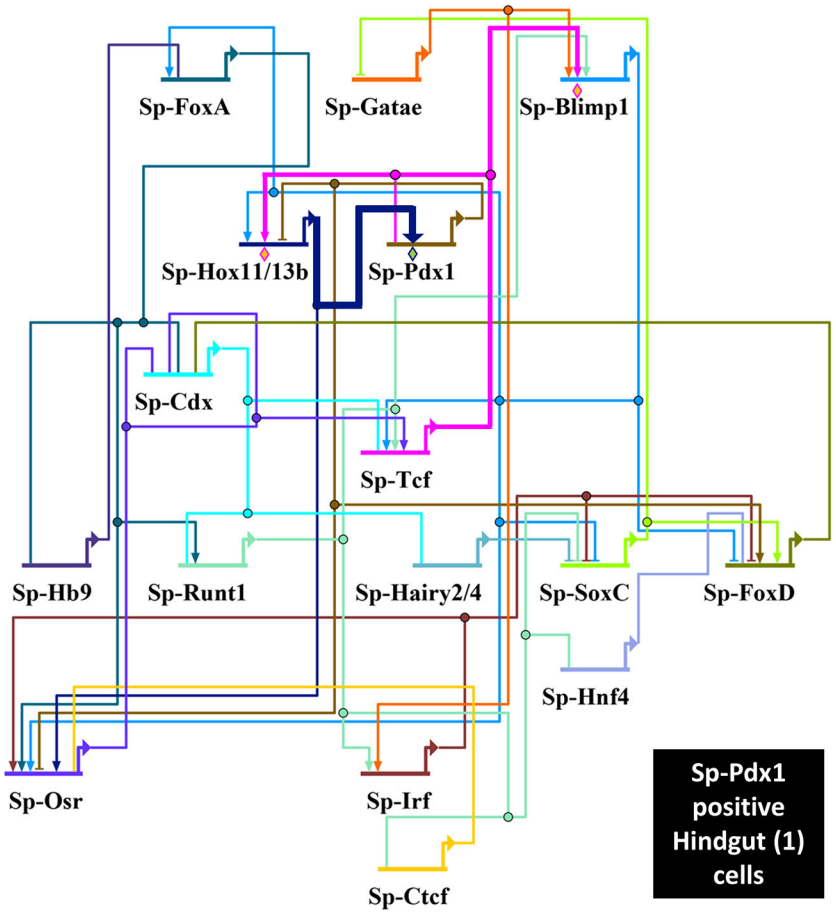


Fig. 6. GRN operating in *Sp-Pdx1*-positive hindgut cells. Updated GRN of previously published nodes within *Sp-Pdx1*-positive hindgut cells drafted using the presented multi-omics approach. Validation of interactions by cis-regulatory analysis are indicated by orange diamonds (Cui et al., 2017; Smith et al., 2008).

shown previously at blastula 18-24 h post fertilization (Cui et al., 2017; Smith et al., 2008), and our data suggest that these effects persist in the 2 dpf gastrula as well (Fig. 6, orange diamonds). In addition, we have also shown that many of the previously identified interactions are indirect and happen via intermediates, suggesting high complexity and depth of the 2 dpf sea urchin hindgut GRN, even for nodes co-expressed with *Sp-Pdx1* (Fig. 6).

DISCUSSION

The predicted GRN recapitulates and refines the existing GRN

The sea urchin endomesoderm GRN is one of the most studied GRNs. However, the GRNs were traditionally drafted for morphologically identifiable tissues that may or may not express all the nodes at a particular time point. In addition, historically, GRNs had limited information, mostly due to lack of high-throughput technologies at the time, on whether a particular interaction is direct or indirect, unless cis-regulatory analysis was performed. Lack of information on the directness allows only the overall final effect of perturbation on a gene to be identified. Integrative multi-omics allows us to address these issues. The GRN draft presented in this study is the first for a cell population expressing a particular TF, in this case, *Sp-Pdx1*, and all the nodes are shown to be co-expressed with *Sp-Pdx1* in the same cells by scRNA-seq data. ATAC-seq data, in contrast, points to the interactions of the GRN that are likely to be direct and those that are not, and the intermediates through which they could go. These intermediates allow deduction of the most parsimonious effect of perturbation of a gene on the downstream targets, both direct or indirect. Thus, the available omics data allow us to tackle sea urchin GRNs at unprecedented resolution, which refines the existing GRNs and highlights the complexity and depth of the regulatory wiring responsible for the development of cell types, tissues and other embryonic structures. Our data suggest that the hindgut GRN and, likely, the whole endomesoderm GRN are complex and diverse with many TF genes serving as intermediates in the interactions of the GRN nodes.

The published interactions of the hindgut GRN were recapitulated through our approach, albeit showing that most of them are likely indirect, increasing the resolution of the interactions of the nodes. This approach stays within the logic of Materna and Oliveri (2008) for drafting GRNs, as the embryological knowledge is obtained via scRNA-seq data, whereas the information on the regulatory genes such as TFs as well as CRMs that establish the regulatory state are predicted via ATAC-seq and scRNA-seq datasets. Finally, the causal information can be added through differential bulk RNA-seq after gene perturbation. Thus, all the necessary components for GRN drafting as per Materna and Oliveri (2008) can be obtained through multi-omics and their integration.

Hints to potential similarities with vertebrate gut GRNs

The possibility of using vertebrate position weight matrices (PWMs) for our TF footprinting also highlights the remarkable, albeit perhaps expected, conservation of the gut GRN nodes between sea urchin and vertebrates, which was previously noted (Annunziata et al., 2014; Annunziata et al., 2019; Grainger et al., 2010).

Sp-Pdx1 and Sp-Cdx are homologs of vertebrate PDX1 and CDX2, respectively, which have well-known functions in the vertebrate gut, with both genes encoding these proteins being

Development (2024) 151, dev202278. doi:10.1242/dev.202278

crucial for gastrointestinal (GI) tract development (Gao et al., 2009). CDX2 controls intestinal development, whereas PDX1 has a prominent role in the development of the vertebrate pancreas and duodenum, with PDX1 controlling *CDX2* in the vertebrate duodenum (Teo et al., 2015). Data also suggest interplays between PDX1, *GATA6* (potential homolog of Sp-Gatae), *HES1* (potential homolog of *Sp-Hairy2/4*), *FOXA2* (*Sp-FoxA*), SOX4 and SOX12 (potential homologs of Sp-SoxC), *OSR1* or *OSR2* (*Sp-Osr*) and *MNX1* (*Sp-Hb9*) in human pancreatic progenitors, adult pancreatic cells and other GI cells (Douchi et al., 2022; Gracz and Magness, 2011; Han et al., 2020; Ito, 2011; Teo et al., 2015; Wang et al., 2018b; Zhao et al., 2022). FOXA2 in turn can directly control PDX1 through its gene enhancers, suggesting a regulatory loop between the genes encoding these two proteins (Gao et al., 2008). *HNF4A* (*Sp-Hnf4*) is also controlled by PDX1 in the human adult pancreas (Thomas et al., 2001). RUNX3 (potential homolog of Sp-Runt1) regulates *CDX2* and *PDX1* along with β-catenin and T cell factors (TCFs) (Douchi et al., 2022; Ito, 2011), which are potential homologs of Sp-Tcf. Gene-to-gene relationships among vertebrate *PRDM1* (homolog of *Sp-Blimp1*), *IRF1* (homolog of *Sp-Irf*), *HOXD13* (*Sp-Hox11/13b*), *FOXD1*, *FOXD2* or *FOXD3* (*Sp-FoxD*), and *PDX1* are unclear, which could be due to differences in protein localization as well as simply lack of research linking them on the molecular level. These genes are involved in the development of various parts of the vertebrate GI tract but may not be spatially co-expressed with *PDX1* (Kim et al., 2023; Mould et al., 2015; Muraro et al., 2016; Wang et al., 2018a; Yahagi et al., 2004; Zhou et al., 2022) (Fig. 6).

To summarize, it appears that, for many vertebrate homologs of the genes present in the updated hindgut GRN, *PDX1* plays a regulatory role (see above) in the pancreas and the posterior regions of the gut. Despite the observed conservation of topology of expression of TFs and signaling molecules, the exploration of the gene-regulatory interactions connecting such factors so far mostly highlighted divergence in the architecture of the vertebrate and echinoderm gut GRNs (Arnone et al., 2016).

Our updated gut GRN draft reveals complex interactions in the *S. purpuratus* 2 dpf gastrula posterior gut GRN, further highlighting a degree of conservation between sea urchin and vertebrates of players previously observed (Pdx1, Cdx, FoxA, Blimp1, etc.) and, in addition, suggests the involvement of TFs such as Osr, Runt, Hnf4, Irf, etc., which were absent in the previous sea urchin gut GRN reconstruction. Sp-Pdx1 has an effect on the sea urchin vertebrate homologs of posterior gut genes; however, further research into the comparison between vertebrate GI GRNs with the sea urchin gut GRNs is necessary to properly assess the extent of the rewirings that have occurred in the posterior gut GRN around Pdx1 between vertebrates and echinoderms.

Limitations and strengths of the approach

The multi-omics approach used in this work has its limitations. The majority of TF PWMs are available for human or mouse TFs (Fornes et al., 2020), with only very few sea urchin PWMs published. Thus, human homologs are used for TF footprinting, which could lead to false negatives or positives if the sea urchin homologs have different binding sequences. In addition, when using vertebrate PWMs, certain TFs have similar PWMs such as forkhead genes (FOXD, FOXI, FOXA, etc.) or SRY genes (e.g. SOXC), which complicates identification of TFs binding to a particular locus in a CRM and could also lead to incorrect identification of the homolog affecting a particular gene in other organisms such as vertebrates, as discussed in the previous section. Our approach does not currently allow us to resolve such ambiguities, so all TFs potentially bound to the CRM are indicated in the GRN draft (Fig. S4B, Table S6). The other side of this issue is that there could also be false negatives in our footprinting identification of TFs bound to OCRs. Vertebrate PWMs do not represent all the possible binding motifs for *S. purpuratus* genes; so, potentially some interactions could not be identified. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) data would help address these ambiguities.

The tools available are designed for cell cultures but the ATAC-seq and bulk RNA-seq data used in this study comes from whole embryos and tissues. These data are likely to involve a lot of noise and could lead to certain cell-specific OCRs and TF footprints being missed by our analyses, for example, some gut target gene-CRM-TF interactions could potentially be hidden by reads coming from mesodermal cells still attached at the tip of the archenteron (presumptive coelomic pouches or muscles). These cells have a different repertoire of TFs bound to their DNA, but these cells are likely to be left over during the gut tissue preparation. This can be alleviated by using single-cell ATAC-seq data, which would allow identification of OCRs and TF footprints with much higher specificity. In addition, adding ChIP-seq data to the pipeline would allow identification at actual locations of TF binding, rather than those predicted from footprinting, even though ChIP-seq approaches are not high throughput and focus on a single TF. Linking the pCRM to its target gene is also tricky, as a single CRM may control multiple genes other than the closest one, which, in some cases, could itself be unaffected by a given pCRM. Possible distal interactions between pCRMs and target genes via chromatin 3D conformation as well as TF-CRM interactions via cofactors are also missed by this approach, with more datasets necessary to alleviate this issue. Additionally, information in some cases may be hidden due to lateral inhibition, which could lead to a TF activating certain genes in a given cell type, but inhibiting genes in another cell type due to cell-cell signaling, hindering the assignment of a directionality dynamic (Cary et al., 2020; Castro et al., 2005). All this could lead to false positives and false negatives. Thus, the indicative nature of our approach highlights the need for *in vivo* validations. With published CRM analysis data lacking information for 2 dpf, more experimental *in vivo* validations are necessary to validate the results of our approach at this stage of embryo development.

It is also important to note that our approach does not take into account cases in which TFs operate within a cell type when they are not being actively expressed by its cells. Such are the cases of protein remnants from a previous developmental time point that could still be active, due to a longer protein half-life. For instance, *Sp-Pdx1* could affect more cells than those actively expressing it at the 2 dpf gastrula stage as the Sp-Pdx1 protein could persist in the cells and continue to function after active transcription and translation of this TF have ended. Thus, more cells of the hindgut could have genes under Sp-Pdx1 control if these cells were expressing *Sp-Pdx1* earlier in development. This could also explain the widespread, in terms of the number of downregulated genes, effect of Sp-Pdx1 knockdown observed in cells of the ‘Hindgut (2)’ cluster, as shown in Fig. 3C,D.

Thus, due to limitations discussed, for GRNs drafted using *in silico* approaches, similar to the approaches described in this study, *in vivo* validations are necessary to confirm the predicted interactions. Sea urchins, with the vast molecular toolkit available, are a great experimental system to perform such validations, for example, through *in vivo* reporter construct experiments as exemplified in this study by cis-regulatory analysis of Sp-Pdx1-pCRM5 combined with *Sp-Hox11/13b* perturbation.

Development (2024) 151, dev202278. doi:10.1242/dev.202278

Despite the highlighted limitations, the importance of multimodal approaches to improve GRN inference has been recently highlighted (van der Sande et al., 2023). For vertebrate experimental systems, approaches using scRNA-seq or a combination of scRNA-seq and single-cell ATAC-seq of the same cells have emerged (Yang et al., 2023; Zhang et al., 2023), which, similarly to our approach, allow drafting of single-cell GRNs. These approaches are robust in identifying nodes and interactions of the GRNs, although most of them lack information on whether a given identified interaction is activatory or inhibitory. In our approach, we attempted to obtain the most parsimonious prediction to the directness of interactions predicted by ATAC-seq and scRNA-seq datasets using RNA-seq data. Although the method presented in the current study lacks single-cell ATAC-seq data, the permissive rather than deterministic nature of chromatin accessibility allows us to use tissue-level ATAC-seq data for GRN drafting, despite the potentially higher noise. Applying the omics data and integrating them as we present in the current study gives the necessary information for GRN drafting at the cell type level for the 2 dpf sea urchin embryo, which can be followed with targeted *in vivo* validations.

Conclusions

Overall, our multi-omics approach drafts GRNs that allow us to increase the resolution and highlight the high interconnectivity of the sea urchin hindgut GRN nodes around *Sp-Pdx1*. Thus, within echinoderm research, such an approach has the potential of giving a more holistic view of gene regulation in a specific cell type in a developing embryo. Moreover, this approach is applicable to any experimental system, including emerging model organisms, as multi-omics datasets become more and more available owing to the recent advances in next-generation sequencing technologies.

MATERIALS AND METHODS

Animal husbandry and culture of embryos

Adult *Strongylocentrotus purpuratus* individuals were obtained from Patrick Leahy (Kerckhoff Marine Laboratory, California Institute of Technology, Pasadena, CA, USA) and maintained in circulating seawater aquaria at Stazione Zoologica Anton Dohrn in Naples, Italy. Gametes were obtained by vigorously shaking the animals. Embryos were cultured at 15°C in filtered Mediterranean Sea water diluted 9:1 with deionized water, and collected at 2 dpf.

Gut tissue separation

Sea urchin gut tissue was obtained by adapting existing protocols (Juliano et al., 2014; McClay, 2004). Echinoderm embryos were grown until the 2 dpf gastrula stage. The embryo suspensions were collected into 1.5 ml tubes using a 40 µm Nitex mesh filter, concentrated by centrifuging at 500 ***g*** for 5 min at 4°C, and then washed once in 1 ml of Ca^{2+}- and Mg^{2+}-free sea water. Then, the embryos were concentrated again by centrifuging at 500 ***g*** for 5 min at 4°C and treated with 1 M glycine and 0.02 M EDTA in Ca^{2+}- and Mg^{2+}-free sea water for 10 min on ice. Embryos were continuously pipetted up and down carefully using a P1000 micropipette to mechanically separate other tissue cells from the gut tissue until gut separation was achieved. This was performed in 1% agarose plates under the dissecting microscope to visually control the dissociation process. The guts were then individually collected via pipetting into 1.5 ml Eppendorf tubes with 100 µl of artificial sea water (28.3 g NaCl, 0.77 g KCl, 5.41 g $MgCl_2{\cdot}6H_2O$, 3.42 g $MgSO_4$ or 7.13 g $MgSO_4{\cdot}7H_2O$, 0.2 g $NaHCO_3$, 1.56 g $CaCl_2{\cdot}2H_2O$ per 1 l of deionized water) on ice.

ATAC-seq library synthesis

ATAC-seq libraries were generated as described by Magri et al. (2021). Around 270 gastrula stage embryos or 400 embryonic guts were collected per biological replicate and washed on the 40 µm Nitex mesh filter with artificial sea water and transferred into 1.5 ml tubes. The samples were then centrifuged at 500 ***g*** for 5 min at 4°C, washed twice with 200 µl artificial sea water in the 1.5 ml tube and then resuspended in 50 µl of lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.2% IGEPAL CA-630), followed by lysis by pipetting up and down for 3-5 min. Half of the lysate was used for counting the released nuclei under a Zeiss AxioImager M1 microscope using a hemocytometer with DAPI dye (1 µl of 1:100 diluted DAPI in the 25 µl of the released nuclei). From the other half, around 75,000 nuclei were used for the tagmentation reaction by splitting the sample, if necessary, to contain the required number of nuclei, centrifuging the sample at 500 ***g***, removing lysis buffer and then incubating for 30 min at 37°C with 25 µl of 2× tagmentation buffer [20 mM Tris-HCl, 10 mM $MgCl_2$, 20% (v/v) dimethylformamide], 23.75 µl of nuclease-free water and 1.25 µl of Tn5 enzyme (Illumina; provided by J.L.G.-S.).

The tagmented DNA was purified using MinElute Kit (Qiagen) following the manufacturer's instructions and eluted in 10 µl of elution buffer. The eluted DNA was then amplified to obtain the library for sequencing [10 µl of eluted tagmented DNA, 10 µl of nuclease-free water, 2.5 µl 10 µM Nextera Primer 1, 2.5 µl 10 µM Nextera Primer 2.X (where X is the unique Nextera barcode used for sequencing) and 25 µl of NEBNext High-Fidelity 2× PCR Master Mix (New England BioLabs)] using the following thermocycler program: 72°C for 5 min; 98°C for 30 s; then 15 cycles of 98°C for 10 s, 63°C for 30 s and 72°C for 1 min; followed by a hold step at 4°C. The amplified library was then purified with the MinElute Kit following the manufacturer's instructions and eluted in 20 µl of elution buffer. The concentration of the resulting library was checked using the Qubit dsDNA BR Assay Kit (Molecular Probes) and its quality was assessed by running 70 µg of the library on a 2% agarose 1× TAE gel. Libraries exhibiting nucleosomal size patterns on the gel were sent for sequencing. Sequencing was performed by BGI Tech (Hong Kong), resulting in 49 bp paired-end reads (Illumina HiSeq 4000) with the mean of 57 million reads.

ATAC-seq data analysis

The raw ATAC-seq reads were trimmed from adapter sequences and bad-quality bases using Trimmomatic 0.38 (Bolger et al., 2014) using the following parameters: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:6 CROP:40 SLIDINGWINDOW:3:25 MINLEN:25 for paired-end data and -phred33 quality scores. The good-quality reads from each replicate were then aligned to the *S. purpuratus* genome v3.1 linear scaffolds assembly file (Sea Urchin Genome Sequencing Consortium et al., 2006) using Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012). The mapped reads in BAM format were then converted into BED files using BEDtools v2.27.1 (Quinlan and Hall, 2010), the BAM file was filtered using SAMtools v1.7.1 (Li et al., 2009) to keep aligned reads of quality greater than 30, fix mates and remove duplicates with default parameters. In addition, only fragments of less than 130 bp in size were kept. The resulting files were then used in MACS2 v2.1.2 software (Zhang et al., 2008) to call peaks using BED as the input file format as well as setting –extsize to 100, –shift to 50 and using the –nomodel setting with genome size of 815,936,258 for the *S. purpuratus* genome v3.1; all the other MACS2 settings were kept at default. The BEDtools v2.27.1 intersect tool (Quinlan and Hall, 2010) was used with default settings to combine replicates and obtain a consensus set of peaks, which were then all merged using BEDtools merge to the pCRM list.

The resulting gut ATAC-seq BAM files were merged using BAMtools and used as input in TOBIAS v0.11.1-dev (Bentsen et al., 2020) along with the genome and CRM list to perform TF footprinting analysis. The TOBIAS commands ATACorrect and ScoreBigwig were run with default parameters, BINDetect was run using JASPAR2020 vertebrate motif database PWMs (Fornes et al., 2020) using –motif-pvalue of 1e−5. Motifs for each TF footprint with a score of at least 0.5 were kept, to obtain a table with each pCRM and TFs bound to it. HOMER v4.10.3 (Heinz et al., 2010) annotatePeaks.pl was used to assign pCRMs to *S. purpuratus* genes. BLAST v2.6.0+ (Altschul et al., 1990) with max_target_seqs of 1 and max_hsps of 1 was used to identify the single best sea urchin homolog for each of vertebrate TFs. OCRs, TF footprint locations, gene models and open chromatin coverage were visualized using Integrative Genomics Viewer v2.16.0 (Robinson et al., 2011).

Development (2024) 151, dev202278. doi:10.1242/dev.202278

Single-cell embryo dissociation

Dissociation of the 2 dpf *S. purpuratus* gastrulae into single cells was performed as described by Paganos et al. (2021). Embryos were collected, concentrated using a 40 µm Nitex mesh filter and centrifuged at 500 ***g*** for 10 min. Sea water was removed and embryos were resuspended in Ca^{2+}- and Mg^{2+}-free artificial sea water. Embryos were concentrated at 500 g for 10 min at 4°C and resuspended in a dissociation buffer containing 1 M glycine and 0.02 M EDTA in Ca^{2+}- and Mg^{2+}-free artificial sea water. Embryos were incubated on ice for 10 min and mixed gently approximately every 2 min, monitoring the progress of the dissociation. Dissociated cells were concentrated at the bottom of the tube by centrifugation at 700 ***g*** for 5 min and washed several times with Ca^{2+}- and Mg^{2+}-free artificial sea water. Cell viability was assessed using propidium iodide and fluorescein diacetate; samples with cell viability ≥90% were further processed. Single cells were counted using a hemocytometer and diluted according to the manufacturer's protocol (10× Genomics).

scRNA-seq and data analysis

scRNA-seq and the analysis of the 2 dpf *S. purpuratus* gastrula stage data were performed as described by Paganos et al. (2021). scRNA-seq was performed using the 10× Genomics single-cell-capturing system. Cells from three biological replicates, ranging from 6000 to 20,000 cells, were loaded on the 10× Genomics Chromium Controller. Single-cell cDNA libraries were prepared using the Chromium Single Cell 3′ Reagent Kit (v3 and v3.1). Libraries were sequenced by GeneCore (European Molecular Biology Laboratory, Heidelberg, Germany) for 75 bp paired-end reads (Illumina NextSeq 500). Cell Ranger Software Suite v3.0.2 (10× Genomics) was used for the alignment of the scRNA-seq output reads and generation of features, barcodes and matrices. The genomic index was made in Cell Ranger using the *S. purpuratus* genome v3.1 and its associated annotation (Kudtarkar and Cameron, 2017; Sea Urchin Genome Sequencing Consortium et al., 2006). Cell Ranger output matrices for three biological replicates (with two of these having two technical replicates) were used for further analysis in Seurat v3.0.2 R package (Stuart et al., 2019). Genes that were transcribed in less than three cells and cells that had less than a minimum of 200 transcribed genes were excluded from the analysis. The cutoff number of transcribed genes was determined based on feature scatter plots and varied depending on the replicate. In total, 15,341 cells passed the quality checks and were further analyzed. Datasets were normalized and variable genes were found using the variance-stabilizing transformation (VST) method with a maximum of 2000 variable features. Data integration was performed via identification of anchors among the five different objects. Next, the datasets were scaled and principal component analysis was performed. The shared nearest-neighbor (SNN) graph was computed with 20 dimensions (resolution 1.0) to identify the clusters. Genes co-expressed with *Sp-Pdx1* were identified by first identifying *Sp-Pdx1*-expressing cells by adding cell scores for *Sp-Pdx1* expression using AddModuleScore (Stuart et al., 2019), and then subsetting to contain only cells with genes co-expressed with *Sp-Pdx1* in the 'Hindgut (1)' cluster. We use the added *Sp-Pdx1* module score cutoff of 1.5 and the cutoff for average expression of genes of 0.5. Transcripts of all genes per cell type of interest were identified by converting a Seurat DotPlot for these cells into a table with all these transcripts as features (ggplot2 v3.2.0 R package; https://ggplot2.tidyverse.org). All resulting tables containing the genes transcribed within different cell type families were further annotated by adding PFAM terms for associated proteins (Finn et al., 2014; Trapnell et al., 2010), Gene Ontology terms and descriptions from Echinobase (https://www.echinobase.org/) (Kudtarkar and Cameron, 2017; Telmer et al., 2024).

MO injections

Microinjections were performed as follows. The sea urchin eggs were dejellied in acidic seawater (pH 4.5) for <1 min, then washed in filtered sea water and rowed onto 4% protamine sulfate plates filled with 50 mg of *p*-aminobenzoic acid in 100 ml of filtered sea water. The microinjecting needle was pulled from borosilicate glass with capillary using a P-97 micropipette puller (Sutter Instrument, Novato, CA, USA); then, samples were loaded with the injection solution using a Microloader pipette tip (Eppendorf). The loaded needle tip was broken off using a scratch in the middle of the protamine plate with eggs. The eggs were fertilized with a few drops of diluted sea urchin sperm and injected with approximately 2 pl of microinjection solution. Injected eggs were washed twice with filtered sea water and incubated at 15°C overnight; then, the hatched embryos were transferred to four-well plates (Thermo Fisher Scientific) with filtered sea water to grow until the 2 dpf gastrula stage.

The MO microinjection solutions containing 100 µM of final MO concentration were heated up to 75°C for 5 min and then passed through a 0.22 µm PVDF micro-filter (Millipore) placed in a 500 µl tube by centrifugation at 2500 ***g*** for 2 min. After filtration, the filtrate was centrifuged for 15 min at 16,000 ***g*** prior to microinjection. The translation MO sequence for *Sp-Hox11/13b* is the same as that previously described (Arenas-Mena et al., 2006).

RNA-seq data analysis

Raw reads for 2 dpf *Sp-Pdx1* MO-injected and untreated embryos were published by Annunziata and Arnone (2014) and the bulk RNA-seq data can be found at https://osf.io/cbsxr/files/. The reads were re-analyzed for this study, adhering to more up-to-date pipelines. FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess the quality of sequencing data. Bad-quality sequences were trimmed from the reads using Trimmomatic v0.38 (Bolger et al., 2014) with the following options: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:6 CROP:90 HEADCROP:10 SLIDINGWINDOW:3:25 MINLEN:25 using -phred33 base-quality encoding for paired-end reads. Only the paired output of Trimmomatic was used as input in Salmon v0.11.3 (Patro et al., 2017) for transcript quantification with automatic library detection and all settings left as default. Salmon index was made using transcriptome sequences in FASTA format corresponding with *S. purpuratus* genome v3.1 from Echinobase (https://www.echinobase.org/) (Cary et al., 2018; Tu et al., 2012) with a k-value of 25. The Salmon output quantification files were used in the DESeq2 (Love et al., 2014) R package for differential gene expression analysis; the tximport (Soneson et al., 2015) package was used to link every transcript to a WHL identifier denoting genes in the genome annotation. The analysis was performed using local fit and condition (untreated or MO injected) as the factor. Differentially expressed transcripts with an adjusted *P*-value of less than 0.05 using independent hypothesis weighting were considered significant. The resulting transcripts were annotated by adding *S. purpuratus* gene names.

In silico GRN drafting

The general flowchart of the method is described in Fig. S6. The table with each putative sea urchin CRM and TFs bound to it, along with their target genes (described in ATAC-seq data analysis), was filtered to have only TFs and target genes present in a single cell type family. For this, the tables containing the genes transcribed within different cell type families, obtained in scRNA-seq data analysis as described above, were cut to have only genes expressed in the hindgut cluster. This list was used for filtering the ATAC-seq analysis results. The resulting table of TFs and target genes was further filtered to have only TFs (Lambert et al., 2018) both as target and effector. In order to build a network downstream of Sp-Pdx, the data were further narrowed down to contain only genes expressed in *Sp-Pdx1*-positive cells. For this, AddModuleScore from Seurat (Stuart et al., 2019) was used to assess expression of *Sp-Pdx1* in the 2 dpf cells; cells with a module score higher than the cutoff were analyzed further. Genes expressed in these cells with scaled average expression greater than 0.5 were used to build the GRN downstream of *Sp-Pdx1*-positive cells. To predict causal information of the interactions within the Sp-Pdx1 network, the results of the bulk RNA-seq differential expression analyses between untreated embryos and *Sp-Pdx1* MO-injected embryos were used. With the help of a custom R v4.1.2 (https://www.r-project.org/) script, direct targets of Sp-Pdx1 (based on ATAC-seq data) within *Sp-Pdx1*-positive cells were assigned causal dynamics information; then, using the signs of the direct targets, the targets of these direct targets were assigned dynamics information. The same was then performed for the targets of indirect targets using the signs of direct targets and indirect targets. This process was performed iteratively, getting more removed from Sp-Pdx1, to give the most parsimonious causal information to the genes downstream of Sp-Pdx1 in the GRN through iterative assignment

of interaction signs. For instance, the *Sp-Pdx1* gene in this case is the effector (Efct) gene that has direct (D-tgt) and indirect (I-tgt) targets (Fig. S7). If a direct target of *Sp-Pdx1* (D-tgt1) is expressed less when *Sp-Pdx1* is knocked down and its direct target, i.e. an indirect *Sp-Pdx1* target (I-tgt1), is also expressed less when *Sp-Pdx1* is knocked down, then the Sp-Pdx1 TF activates D-tgt1 and D-tgt1 in turn activates I-tgt1 (Fig. S7). However, if an indirect *Sp-Pdx1* target (I-tgt2), which is a direct target of a direct *Sp-Pdx1* target (D-tgt2), is downregulated by *Sp-Pdx1* MO, thus upregulated by the Sp-Pdx1 TF, and the direct *Sp-Pdx1* target (D-tgt2) is downregulated by the Sp-Pdx1 TF, then we can deduce that D-tgt2 downregulates the I-tgt2 (Fig. S7). The same approach can be used with the combination of upregulating and downregulating interactions, such as activation of I-tgt3 by Sp-Pdx1 (Fig. S7). Again, from the effect of Efct on D-tgt3 and overall effect of Efct on I-tgt3, the effect of D-tgt3 on I-tgt3 can be deduced. This process can be performed iteratively for other indirect targets that are more removed from *Sp-Pdx1*, i.e. those that have more intermediates between them and *Sp-Pdx1*, to assign signs to interactions downstream of *Sp-Pdx1*. After assigning signs to interactions, the identifiers used throughout the pipeline are then converted to gene names for visualization in BioTapestry software (https://biotapestry.systemsbiology.net/).

CRM-GFP reporter construct synthesis

Tagging of pCRMs was performed according to a protocol described by Nam et al. (2010). The plasmids with DNA tags were provided by Dr Jongmin Nam (Rutgers University–Camden, Camden, NJ, USA). 50 bp flanks from the genome were added to the pCRMs to design primers using Primer3web v4.1.0 (Untergasser et al., 2012) falling within the added 50 bp to ensure that the whole CRM is amplified. 18 bp of the reverse complement of the beginning of the DNA tag sequences were added to the 5′ of the reverse primer so that the CRM could be combined with the DNA tag in an equal amount by overlap PCR (Xiong et al., 2006) using Expand High-Fidelity PLUS PCR (Sigma-Aldrich). The resulting fragment was run on a 2% agarose 1× TAE gel, the corresponding band was cut out, and the gel was purified using the GenElute Gel Extraction Kit (Sigma-Aldrich) according to the manufacturer's guidelines and eluted in 50 µl of kit elution buffer. The eluted DNA was then purified again using QIAquick PCR Purification Kit (Qiagen) and eluted in 30 µl of kit elution buffer. DNA yield and concentration were assessed using NanoDrop ND-1000 (Thermo Fisher Scientific).

CRM-GFP reporter construct microinjections

The tagged CRM was used to make microinjection solutions according to Nam et al. (2010): 0.5 µl of tagged CRM solution, 1.2 µl of 1 mM KCl, 0.275 µl carrier DNA {genomic DNA sheared with HindIII enzyme [2 U for 1 µg of DNA in SuRE/Cut Buffer B (Roche) for 3 h at 37°C], purified using QIAquick PCR Purification Kit (Qiagen) and diluted to 500 ng/µl} and water up to 10 µl (Arnone et al., 2004; Nam et al., 2010). The prepared solutions were then centrifuged at 16,000 ***g*** for 15 min prior to microinjections. The microinjection procedure was the same as that described in the 'MO injections' section above.

FISH

FISH was performed as described by Paganos et al. (2022a). In brief, embryos were collected at 2 dpf and fixed in 4% paraformaldehyde (PFA) in MOPS buffer overnight at 4°C. The following day, specimens were washed with MOPS buffer several times and then stored in 70% ethanol at −20°C. Antisense mRNA probes were generated as described by Perillo et al. (2021). The primer sequences used for cDNA isolation were the same as previously described: *Sp-Synb* (Burke et al., 2006); *Sp-FoxA*, *Sp-Pdx1*, *Sp-Brn1/2/4*, *Sp-Cdx*, *Sp-ManrC1a*, *Sp-Hox11/13b* (Annunziata and Arnone, 2014); *Sp-FoxC* (Andrikou et al., 2015); *Sp-Ptf1a* (Perillo et al., 2016); *Sp-Pks1* (Perillo et al., 2020); *Sp-E2_D1*, *Sp-Fbsl_2*, *Sp-Bra*, *Sp-FoxABL*, *Sp-Frizz5/8*, *Sp-FcolI/II/III*, *Sp-Hypp_1249*, *Sp-Mlckb* and *Sp-Hypp_2386* (Paganos et al., 2021). Probes were synthesized from linearized DNA and labeled during synthesis with digoxigenin-11-UTP (Roche) nucleotides. The fluorescence signal was developed using the fluorophore-conjugated tyramide technology using TSA Plus Cyanine 3 and 5 kits (Akoya Biosciences). Specimens were imaged using a Zeiss LSM 700 confocal microscope.

Acknowledgements

We thank Elijah Kareem Lowe for setting the foundation for designing the approach pipeline and Lorena Buono for help with scripts. We thank Rossella Annunziata for invaluable feedback on the manuscript and Davide Caramiello for taking care of the animals. We also thank Detlev Arendt, Jacob M. Musser and the European Molecular Biology Laboratory GeneCore for single-cell transcriptomics.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: D.V., M.I.A.; Methodology: D.V., P.P., M.S.M., C.C., M.I.A.; Software: D.V.; Validation: D.V., P.P.; Formal analysis: D.V., P.P., M.S.M.; Investigation: D.V., P.P., M.S.M., C.C.; Resources: I.M., J.L.G.-S., M.I.A.; Data curation: D.V., P.P., M.S.M.; Writing - original draft: D.V., P.P.; Writing - review & editing: M.S.M., C.C., I.M., M.I.A.; Visualization: D.V., P.P.; Supervision: I.M., J.L.G.-S., M.I.A.; Project administration: I.M., J.L.G.-S., M.I.A.; Funding acquisition: J.L.G.-S., M.I.A.

Funding

D.V. and C.C. were supported by the Stazione Zoologica Anton Dohrn PhD fellowships. This work was supported by the H2020 Marie Skłodowska-Curie Actions Innovative Training Network EvoCELL (grant number 766053 to M.I.A. and fellowship to P.P.). J.L.G.-S. was supported by the Spanish government (Ministerio de Economía y Competitividad; grant BFU2016-74961-P), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 740041) and the institutional grant Unidad de Excelencia María de Maeztu (MDM-2016-0687 to the Department of Gene Regulation and Morphogenesis, Centro Andaluz de Biología del Desarrollo). M.S.M. has been granted a fellowship of the Programme for the Training of Researchers by the Ministry of Economy, Industry and Competitiveness of Spain (Ministerio de Economía y Competitividad; BES-2014-068494). I.M. acknowledges support from the Spanish Ministry of Science and Innovation (Agencia Estatal de Investigación) and the European Union (grants RYC-2016-20089, PGC2018-099392-A-I00 and PID2021-128728NB-I00).

Data availability

The raw ATAC-seq and scRNA-seq data generated for this work and described in this study can be found at NCBI Gene Expression Omnibus (accession number GSE262916). Scripts associated with the data analysis pipeline can be found at https://github.com/Danvor/spur_2dpf_hindgut_pdx1_downstream_grn. The table with various gene name synonyms can be found at the same GitHub link.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410. doi:10.1016/S0022-2836(05)80360-2

Amore, G., Yavrouian, R. G., Peterson, K. J., Ransick, A., Mcclay, D. R. and Davidson, E. H. (2003). Spdeadringer, a sea urchin embryo gene required separately in skeletogenic and oral ectoderm gene regulatory networks. *Dev. Biol.* **261**, 55-81. doi:10.1016/S0012-1606(03)00278-1

Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P. and Arnone, M. I. (2013). Myogenesis in the sea urchin embryo : the molecular fingerprint of the myoblast precursors Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. *Evodevo* **4**, 33. doi:10.1186/2041-9139-4-33

Andrikou, C., Pai, C.-Y., Su, Y.-H. and Arnone, M. I. (2015). Logics and properties of a genetic regulatory program that drives embryonic muscle development in an echinoderm. *Elife* **4**, e07343. doi:10.7554/eLife.07343

Annunziata, R. and Arnone, M. I. (2014). A dynamic regulatory network explains ParaHox gene control of gut patterning in the sea urchin. *Development* **141**, 2462-2472. doi:10.1242/dev.105775

Annunziata, R., Perillo, M., Andrikou, C., Cole, A. G., Martinez, P. and Arnone, M. I. (2014). Pattern and process during sea urchin gut morphogenesis: the regulatory landscape. *Genesis* **52**, 251-268. doi:10.1002/dvg.22738

Annunziata, R., Andrikou, C., Perillo, M., Cuomo, C. and Arnone, M. I. (2019). Development and evolution of gut structures: from molecules to function. *Cell Tissue Res.* **377**, 445-458. doi:10.1007/s00441-019-03093-9

Arenas-Mena, C., Cameron, R. A. and Davidson, E. H. (2006). Hindgut specification and cell-adhesion functions of Sphox11/13b in the endoderm of the sea urchin embryo. *Dev. Growth Differ.* **48**, 463-472. doi:10.1111/j.1440-169X.2006.00883.x

Arnone, M. I., Martin, E. L. and Davidson, E. H. (1998). Cis-regulation downstream of cell type specification: a single compact element controls the complex expression of the CyIIa gene in sea urchin embryos. *Development* **125**, 1381-1395. doi:10.1242/dev.125.8.1381

Arnone, M. I., Dmochowski, I. J. and Gache, C. (2004). Using reporter genes to study cis-regulatory elements. *Methods Cell Biol.* **74**, 621-652. doi:10.1016/S0091-679X(04)74025-X

Arnone, M. I., Byrne, M. and Martinez, P. (2015). Echinodermata. In: *Evolutionary Developmental Biology of Invertebrates*, vol. 6 (ed. A. Wanninger), pp. 1-58. Vienna: Springer. doi:10.1007/978-3-7091-1856-6_1

Arnone, M. I., Andrikou, C. and Annunziata, R. (2016). Echinoderm systems for gene regulatory studies in evolution and development. *Curr. Opin. Genet. Dev.* **39**, 129-137. doi:10.1016/j.gde.2016.05.027

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T. et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267. doi:10.1038/s41467-020-18035-1

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120. doi:10.1093/bioinformatics/btu170

Buenrostro, J. D., Wu, B., Chang, H. Y. and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109

Burke, R. D., Osborne, L., Wang, D., Murabe, N., Yaguchi, S. and Nakajima, Y. (2006). Neuron-specific expression of a synaptotagmin gene in the sea urchin Strongylocentrotus purpuratus. *J. Comp. Neurol.* **496**, 244-251. doi:10.1002/cne.20939

Cary, G. A., Cameron, R. A. and Hinman, V. F. (2018). EchinoBase: tools for echinoderm genome analyses. *Methods Mol. Biol.* **1757**, 349-369. doi:10.1007/978-1-4939-7737-6_12

Cary, G. A., Mccauley, B. S., Zueva, O., Pattinato, J., Longabaugh, W. and Hinman, V. F. (2020). Systematic comparison of sea urchin and sea star developmental gene regulatory networks explains how novelty is incorporated in early development. *Nat. Commun.* **11**, 6235. doi:10.1038/s41467-020-20023-4

Castro, B., Barolo, S., Bailey, A. M. and Posakony, J. W. (2005). Lateral inhibition in proneural clusters: cis-regulatory logic and default repression by Suppressor of Hairless. *Development* **132**, 3333-3344. doi:10.1242/dev.01920

Cole, A. G. and Arnone, M. I. (2009). Fluorescent in situ hybridization reveals multiple expression domains for SpBrn1/2/4 and identifies a unique ectodermal cell type that co-expresses the ParaHox gene SpLox. *Gene Expr. Patterns* **9**, 324-328. doi:10.1016/j.gep.2009.02.005

Cole, A. G., Rizzo, F., Martinez, P., Fernandez-Serra, M. and Arnone, M. I. (2009). Two ParaHox genes, SpLox and SpCdx, interact to partition the posterior endoderm in the formation of a functional gut. *Development* **136**, 541-549. doi:10.1242/dev.029959

Cui, M., Siriwon, N., Li, E., Davidson, E. H. and Peter, I. S. (2014). Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo. *Proc. Natl. Acad. Sci. USA* **111**, E5029-E5038. doi:10.1073/pnas.1419141111

Cui, M., Vielmas, E., Davidson, E. H. and Peter, I. S. (2017). Sequential response to multiple developmental network circuits encoded in an intronic cis- regulatory module of sea urchin hox11/13b. *Cell Rep.* **19**, 364-374. doi:10.1016/j.celrep.2017.03.039

Davidson, P. L., Guo, H., Swart, J. S., Massri, A. J., Edgar, A., Wang, L., Berrio, A., Devens, H. R., Koop, D., Cisternas, P. et al. (2022). Recent reconfiguration of an ancient developmental gene regulatory network in Heliocidaris sea urchins. *Nat. Ecol. Evol.* **6**, 1907-1920. doi:10.1038/s41559-022-01906-9

de-Leon, S. B.-T. and Davidson, E. H. (2010). Information processing at the foxa node of the sea urchin endomesoderm specification network. *Proc. Natl. Acad. Sci. USA* **107**, 10103-10108. doi:10.1073/pnas.1004824107

Douchi, D., Yamamura, A., Matsuo, J., Lee, J.-W., Nuttonmanit, N., Melissa Lim, Y. H., Suda, K., Shimura, M., Chen, S., Pang, S. et al. (2022). A point mutation R122C in RUNX3 promotes the expansion of isthmus stem cells and inhibits their differentiation in the stomach. *Cell Mol. Gastroenterol. Hepatol.* **13**, 1317-1345. doi:10.1016/j.jcmgh.2022.01.010

Ettensohn, C. A. (2020). The gene regulatory control of sea urchin gastrulation. *Mech. Dev.* **162**, 103599. doi:10.1016/j.mod.2020.103599

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J. et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-D230. doi:10.1093/nar/gkt1223

Fornes, O., Castro-Mondragon, J. A., Khan, A., Van Der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M. and Baranašić, D. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87-D92. doi:10.1093/nar/gkaa516

Gao, N., Lelay, J., Vatamaniuk, M. Z., Rieck, S., Friedman, J. R. and Kaestner, K. H. (2008). Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* **22**, 3435-3448. doi:10.1101/gad.1752608

Gao, N., White, P. and Kaestner, K. H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev. Cell* **16**, 588-599. doi:10.1016/j.devcel.2009.02.010

Gracz, A. D. and Magness, S. T. (2011). Sry-box (Sox) transcription factors in gastrointestinal physiology and disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* **300**, G503-G515. doi:10.1152/ajpgi.00489.2010

Grainger, S., Savory, J. G. A. and Lohnes, D. (2010). Cdx2 regulates patterning of the intestinal epithelium. *Dev. Biol.* **339**, 155-165. doi:10.1016/j.ydbio.2009.12.025

Han, L., Chaturvedi, P., Kishimoto, K., Koike, H., Nasr, T., Iwasawa, K., Giesbrecht, K., Witcher, P. C., Eicher, A., Haines, L. et al. (2020). Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis. *Nat. Commun.* **11**, 4158. doi:10.1038/s41467-020-17968-x

Harkey, M. A., Whiteley, H. R. and Whiteley, A. H. (1992). Differential expression of the msp130 gene among skeletal lineage cells in the sea urchin embryo: a three dimensional in situ hybridization analysis. *Mech. Dev.* **37**, 173-184. doi:10.1016/0925-4773(92)90079-Y

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589. doi:10.1016/j.molcel.2010.05.004

Ho, E. C., Buckley, K. M., Schrankel, C. S., Schuh, N. W., Hibino, T., Solek, C. M., Bae, K., Wang, G. and Rast, J. P. (2017). Perturbation of gut bacteria induces a coordinated cellular immune response in the purple sea urchin larva. *Immunol. Cell Biol.* **95**, 647. doi:10.1038/icb.2017.40

Ito, K. (2011). RUNX3 in oncogenic and anti-oncogenic signaling in gastrointestinal cancers. *J. Cell. Biochem.* **112**, 1243-1249. doi:10.1002/jcb.23047

Juliano, C. E., Yajima, M. and Wessel, G. M. (2010). Nanos functions to maintain the fate of the small micromere lineage in the sea urchin embryo. *Dev. Biol.* **337**, 220-232. doi:10.1016/j.ydbio.2009.10.030

Juliano, C., Swartz, S. Z. and Wessel, G. (2014). Isolating specific embryonic cells of the sea urchin by FACS. *Methods Mol. Biol.* **1128**, 187-196. doi:10.1007/978-1-62703-974-1_12

Khor, J. M., Guerrero-Santoro, J. and Ettensohn, C. A. (2019). Genome-wide identification of binding sites and gene targets of Alx1, a pivotal regulator of echinoderm skeletogenesis. *Development* **146**, dev180653. doi:10.1242/dev.180653

Khor, J. M., Guerrero-Santoro, J., Douglas, W. and Ettensohn, C. A. (2021). Global patterns of enhancer activity during sea urchin embryogenesis assessed by eRNA profiling. *Genome Res.* **31**, 1680-1692. doi:10.1101/gr.275684.121

Kim, H.-M., Kang, B., Park, S., Park, H., Kim, C. J., Lee, H., Yoo, M., Kweon, M.-N., Im, S.-H., Kim, T. I. et al. (2023). Forkhead box protein D2 suppresses colorectal cancer by reprogramming enhancer interactions. *Nucleic Acids Res.* **51**, 6143-6155. doi:10.1093/nar/gkad361

Kudtarkar, P. and Cameron, R. A. (2017). Echinobase: an expanding resource for echinoderm genomic information. *Database* **2017**, bax074. doi:10.1093/database/bax074

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R. and Weirauch, M. T. (2018). The human transcription factors. *Cell* **175**, 598-599. doi:10.1016/j.cell.2018.09.045

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923

Lee, P. Y., Nam, J. and Davidson, E. H. (2007). Exclusive developmental functions of gatae cis-regulatory modules in the Strongylocentrorus purpuratus embryo. *Dev. Biol.* **307**, 434-445. doi:10.1016/j.ydbio.2007.05.005

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. and **1000 Genome Project Data Processing Subgroup**. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079. doi:10.1093/bioinformatics/btp352

Livi, C. B. and Davidson, E. H. (2007). Regulation of spblimp1/krox1a, an alternatively transcribed isoform expressed in midgut and hindgut of the sea urchin gastrula. *Gene Expr. Patterns* **7**, 1-7. doi:10.1016/j.modgep.2006.04.009

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. doi:10.1186/s13059-014-0550-8

Lowe, E. K., Cuomo, C. and Arnone, M. I. (2016). A Differential Transcriptomic Approach to Compare Target Genes of Homologous Transcription Factors in Echinoderm Species. In *Dynamics of Mathematical Models in Biology* (ed. A. Rogato, V. Zazzu and M. Guarracino), pp. 55-63. Cham: Springer International Publishing. doi:10.1007/978-3-319-45723-9_5

Magri, M. S., Voronov, D., Ranđelović, J., Cuomo, C., Gómez-Skarmeta, J. L. and Arnone, M. I. (2021). ATAC-seq for assaying chromatin accessibility protocol using echinoderm embryos. *Methods Mol. Biol.* **2219**, 253-265. doi:10.1007/978-1-0716-0974-3_16

Marlétaz, F., Firbas, P. N., Maeso, I., Tena, J. J., Bogdanovic, O., Perry, M., Wyatt, C. D. R., De La Calle-Mustienes, E., Bertrand, S., Burguera, D. et al. (2018). Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64-70. doi:10.1038/s41586-018-0734-6

Massri, A. J., Greenstreet, L., Afanassiev, A., Berrio, A., Wray, G. A., Schiebinger, G. and Mcclay, D. R. (2021). Developmental single-cell transcriptomics in the Lytechinus variegatus sea urchin embryo. *Development* **148**, dev198614. doi:10.1242/dev.198614

Massri, A. J., Mcdonald, B., Wray, G. A. and Mcclay, D. R. (2023). Feedback circuits are numerous in embryonic gene regulatory networks and offer a

stabilizing influence on evolution of those networks. *Evodevo* **14**, 1-13. doi:10.1186/s13227-023-00214-y
Materna, S. C. and Oliveri, P. (2008). A protocol for unraveling gene regulatory networks. *Nat. Protoc.* **3**, 1876-1887. doi:10.1038/nprot.2008.187
Mccarty, C. M. and Coffman, J. A. (2013). Developmental cis-regulatory analysis of the cyclin D gene in the sea urchin Strongylocentrotus purpuratus. *Biochem. Biophys. Res. Commun.* **440**, 413-418. doi:10.1016/j.bbrc.2013.09.094
Mcclay, D. R. (2004). Methods for embryo dissociation and analysis of cell adhesion. *Methods Cell Biol.* **74**, 311-329. doi:10.1016/S0091-679X(04)74014-5
Mcclay, D. R. (2011). Evolutionary crossroads in developmental biology: sea urchins. *Development* **138**, 2639-2648. doi:10.1242/dev.048967
Mould, A. W., Morgan, M. A. J., Nelson, A. C., Bikoff, E. K. and Robertson, E. J. (2015). Blimp1/Prdm1 functions in opposition to Irf1 to maintain neonatal tolerance during postnatal intestinal maturation. *PLoS Genet.* **11**, e1005375. doi:10.1371/journal.pgen.1005375
Muraro, M. J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., Van Gurp, L., Engelse, M. A., Carlotti, F., De Koning, E. J. P. et al. (2016). A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385-394.e3. doi:10.1016/j.cels.2016.09.002
Nam, J., Dong, P., Tarpine, R., Istrail, S. and Davidson, E. H. (2010). Functional cis-regulatory genomics for systems biology. *Proc. Natl. Acad. Sci. USA* **107**, 3930-3935. doi:10.1073/pnas.1000147107
Paganos, P., Voronov, D., Musser, J. M., Arendt, D. and Arnone, M. I. (2021). Single-cell RNA sequencing of the larva reveals the blueprint of major cell types and nervous system of a non-chordate deuterostome. *Elife* **10**, e70416. doi:10.7554/eLife.70416
Paganos, P., Caccavale, F., La Vecchia, C., D'aniello, E., D'aniello, S. and Arnone, M. I. (2022a). FISH for all: a fast and efficient fluorescent in situ hybridization (FISH) protocol for marine embryos and larvae. *Front. Physiol.* **13**, 878062. doi:10.3389/fphys.2022.878062
Paganos, P., Ronchi, P., Carl, J., Mizzon, G., Martinez, P., Benvenuto, G. and Arnone, M. I. (2022b). Integrating single cell transcriptomics and volume electron microscopy confirms the presence of pancreatic acinar-like cells in sea urchins. *Front. Cell Dev. Biol.* **10**, 991664. doi:10.3389/fcell.2022.991664
Paganos, P., Ullrich-Lüter, E., Caccavale, F., Zakrzewski, A., Voronov, D., Fournon-Berodia, I., Cocurullo, M., Lüter, C. and Arnone, M. I. (2022c). A new model organism to investigate extraocular photoreception: opsin and retinal gene expression in the sea urchin. *Cells* **11**, 2636. doi:10.3390/cells11172636
Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417-419. doi:10.1038/nmeth.4197
Perillo, M., Wang, Y. J., Leach, S. D. and Arnone, M. I. (2016). A pancreatic exocrine-like cell regulatory circuit operating in the upper stomach of the sea urchin Strongylocentrotus purpuratus larva. *BMC Evol. Biol.* **16**, 117. doi:10.1186/s12862-016-0686-0
Perillo, M., Oulhen, N., Foster, S., Spurrell, M., Calestani, C. and Wessel, G. (2020). Regulation of dynamic pigment cell states at single-cell resolution. *Elife* **9**, e60388. doi:10.7554/eLife.60388
Perillo, M., Paganos, P., Spurrell, M., Arnone, M. I. and Wessel, G. M. (2021). Methodology for whole mount and fluorescent RNA in situ hybridization in echinoderms: single, double, and beyond. *Methods Mol. Biol.* **2219**, 195-216. doi:10.1007/978-1-0716-0974-3_12
Peter, I. S. and Davidson, E. H. (2010). The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.* **340**, 188-199. doi:10.1016/j.ydbio.2009.10.037
Poustka, A. J., Kühn, A., Radosavljevic, V., Wellenreuther, R., Lehrach, H. and Panopoulou, G. (2004). On the origin of the chordate central nervous system: expression of onecut in the sea urchin embryo. *Evol. Dev.* **6**, 227-236. doi:10.1111/j.1525-142X.2004.04028.x
Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842. doi:10.1093/bioinformatics/btq033
Rafiq, K., Shashikant, T., Mcmanus, C. J. and Ettensohn, C. A. (2014). Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins. *Development* **141**, 950-961. doi:10.1242/dev.105585
Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26. doi:10.1038/nbt.1754
Röttinger, E. and Lowe, C. J. (2012). Evolutionary crossroads in developmental biology: hemichordates. *Development* **139**, 2463-2475. doi:10.1242/dev.066712
Schwaiger, M., Andrikou, C., Dnyansagar, R., Murguia, P. F., Paganos, P., Voronov, D., Zimmermann, B., Lebedeva, T., Schmidt, H. A., Genikhovich, G. et al. (2022). An ancestral Wnt-Brachyury feedback loop in axial patterning and recruitment of mesoderm-determining target genes. *Nat. Ecol. Evol.* **6**, 1921-1939. doi:10.1038/s41559-022-01905-w
Sea Urchin Genome Sequencing Consortium, **Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D. et al.** (2006). The genome of the sea urchin Strongylocentrotus purpuratus. *Science* **314**, 941-952. doi:10.1126/science.1133609
Shashikant, T., Khor, J. M. and Ettensohn, C. A. (2018). Global analysis of primary mesenchyme cell cis-regulatory modules by chromatin accessibility profiling. *BMC Genomics* **19**, 206. doi:10.1186/s12864-018-4542-z
Smith, J., Kraemer, E., Liu, H., Theodoris, C. and Davidson, E. (2008). A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Dev. Biol.* **313**, 863-875. doi:10.1016/j.ydbio.2007.10.042
Soneson, C., Love, M. I. and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521. doi:10.12688/f1000research.7563.1
Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21. doi:10.1016/j.cell.2019.05.031
Telmer, C. A., Karimi, K., Chess, M. M., Agalakov, S., Arshinoff, B. I., Lotay, V., Wang, D. Z., Chu, S., Pells, T. J., Vize, P. D. et al. (2024). Echinobase: a resource to support the echinoderm research community. *Genetics* **227**, iyae002. doi:10.1093/genetics/iyae002
Teo, A. K. K., Tsuneyoshi, N., Hoon, S., Tan, E. K., Stanton, L. W., Wright, C. V. E. and Dunn, N. R. (2015). PDX1 binds and represses hepatic genes to ensure robust pancreatic commitment in differentiating human embryonic stem cells. *Stem Cell Rep.* **4**, 578-590. doi:10.1016/j.stemcr.2015.02.015
Thomas, H., Jaschkowitz, K., Bulman, M., Frayling, T. M., Mitchell, S. M., Roosen, S., Lingott-Frieg, A., Tack, C. J., Ellard, S. and Ryffel, G. U. (2001). A distant upstream promoter of the HNF-4alpha gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum. Mol. Genet.* **10**, 2089-2097. doi:10.1093/hmg/10.19.2089
Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J. and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511-515. doi:10.1038/nbt.1621
Tu, Q., Brown, C. T., Davidson, E. H. and Oliveri, P. (2006). Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev. Biol.* **300**, 49-62. doi:10.1016/j.ydbio.2006.09.031
Tu, Q., Cameron, R. A., Worley, K. C., Gibbs, R. A. and Davidson, E. H. (2012). Gene structure in the sea urchin Strongylocentrotus purpuratus based on transcriptome analysis. *Genome Res.* **22**, 2079-2087. doi:10.1101/gr.139170.112
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115. doi:10.1093/nar/gks596
Van Der Sande, M., Frölich, S. and Van Heeringen, S. J. (2023). Computational approaches to understand transcription regulation in development. *Biochem. Soc. Trans.* **51**, 1-12. doi:10.1042/BST20210145
Wang, L.-J., Wang, W.-L., Gao, H., Bai, Y.-Z. and Zhang, S.-C. (2018a). FOXD3/FOXD4 is required for the development of hindgut in the rat model of anorectal malformation. *Exp. Biol. Med.* **243**, 327-333. doi:10.1177/1535370217751073
Wang, X., Sterr, M., Burtscher, I., Chen, S., Hieronimus, A., Machicao, F., Staiger, H., Häring, H.-U., Lederer, G., Meitinger, T. et al. (2018b). Genome-wide analysis of PDX1 target genes in human pancreatic progenitors. *Mol. Metab.* **9**, 57-68. doi:10.1016/j.molmet.2018.01.011
Wei, Z., Range, R., Angerer, R. and Angerer, L. (2012). Axial patterning interactions in the sea urchin embryo: suppression of nodal by Wnt1 signaling. *Development* **139**, 1662-1669. doi:10.1242/dev.075051
Xiong, A.-S., Yao, Q.-H., Peng, R.-H., Duan, H., Li, X., Fan, H.-Q., Cheng, Z.-M. and Li, Y. (2006). PCR-based accurate synthesis of long DNA sequences. *Nat. Protoc.* **1**, 791-797. doi:10.1038/nprot.2006.103
Yahagi, N., Kosaki, R., Ito, T., Mitsuhashi, T., Shimada, H., Tomita, M., Takahashi, T. and Kosaki, K. (2004). Position-specific expression of Hox genes along the gastrointestinal tract. *Congenit. Anom.* **44**, 18-26. doi:10.1111/j.1741-4520.2003.00004.x
Yang, Y., Li, G., Zhong, Y., Xu, Q., Chen, B.-J., Lin, Y.-T., Chapkin, R. S. and Cai, J. J. (2023). Gene knockout inference with variational graph autoencoder learning single-cell gene regulatory networks. *Nucleic Acids Res.* **51**, 6578-6592. doi:10.1093/nar/gkad450
Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. doi:10.1186/gb-2008-9-9-r137
Zhang, S., Pyne, S., Pietrzak, S., Halberg, S., Mccalla, S. G., Siahpirani, A. F., Sridharan, R. and Roy, S. (2023). Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nat. Commun.* **14**, 3064. doi:10.1038/s41467-023-38637-9
Zhao, L., Song, W. and Chen, Y.-G. (2022). Mesenchymal-epithelial interaction regulates gastrointestinal tract development in mouse embryos. *Cell Rep.* **40**, 111053. doi:10.1016/j.celrep.2022.111053
Zhou, B., Zhang, J., Zhu, H. and Wu, S. (2022). A potential prognostic marker PRDM1 in pancreatic adenocarcinoma. *J. Oncol.* **2022**, 1934381. doi:10.1155/2022/1934381