Article

Genome-Wide Identification, Comparison, and Expression Analysis of Transcription Factors in Ascidian *Styela clava*

Jin Zhang [1,†], **Jiankai Wei** [1,2,3,†], **Haiyan Yu** [1,*] **and Bo Dong** [1,2,3,*]

1   Sars-Fang Centre, MoE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China; zhangjincoolcool@163.com (J.Z.); weijiankai@ouc.edu.cn (J.W.)
2   Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China
3   Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao 266003, China
*   Correspondence: haiyanyu@ouc.edu.cn (H.Y.); bodong@ouc.edu.cn (B.D.); Tel.: +86-0532-82032732 (B.D.)
†   These authors contributed equally to this work.

Abstract: Tunicates include diverse species, as they are model animals for evolutionary developmental biology study. The embryonic development of tunicates is known to be extensively regulated by transcription factors (TFs). *Styela clava*, the globally distributed invasive tunicate, exhibits a strong capacity for environmental adaptation. However, the TFs were not systematically identified and analyzed. In this study, we reported 553 TFs categorized into 60 families from *S. clava*, based on the whole genome data. Comparison of TFs analysis among the tunicate species revealed that the gene number in the zinc finger superfamily displayed the most significant discrepancy, indicating this family was under the highly evolutionary selection and might be related to species differentiation and environmental adaptation. The greatest number of TFs was discovered in the Cys2His2-type zinc finger protein (zf-C2H2) family in *S. clava*. From the point of temporal view, more than half the TFs were expressed at the early embryonic stage. The expression correlation analysis revealed the existence of a transition for TFs expression from early embryogenesis to the later larval development in *S. clava*. Eight *Hox* genes were identified to be located on one chromosome, exhibiting different arrangement and expression patterns, compared to *Ciona robusta* (*C. intestinalis* type A). In addition, a total of 23 *forkhead box* (*fox*) genes were identified in *S. clava*, and their expression profiles referred to their potential roles in neurodevelopment and sensory organ development. Our data, thus, provides crucial clues to the potential functions of TFs in development and environmental adaptation in the leathery sea squirt.

Keywords: *Styela clava*; transcription factor; zinc-finger protein; *hox*; *forkhead box*

1. Introduction

Transcription factors (TFs) specifically recognize the DNA sequence through DNA-binding domains (DBDs) on itself, controlling transcription, and performing the first step in decoding the DNA sequence [1]. All TFs have at least one DBD, through which they attach to a specific sequence of DNA fragment adjacent to the genes, and then the transcriptions of these genes are either activated or repressed [2]. DBD is a crucial standard for the identification and classification of TFs. The DBD, regulatory regions, and biological functions of TFs are largely conserved across metazoans, although the number and diversity of TFs in different organism are variable [2–4]. TFs play extensively and essentially regulatory functions in diversely biological processes, such as cell differentiation [5], organ development [6], inflammatory response [7], and body-axis building [8]. In metazoans, TFs were identified and classified into 78 TF families in the REGULATOR database [9]. Genome-wide TFs databases are also available, such as the Animal Transcription Factor

DataBase (Animal TFDB) [10]. In these databases, TFs of the representative species, such as *Homo sapiens* and *Drosophila melanogaster* are identified and classified.

Tunicates occupy a crucially phylogenetic status between invertebrate and vertebrate, which are closely related to the vertebrates in evolution. The characteristic development process makes them a great study model for developmental biology. In *Ciona robusta* (*C. intestinalis* type A), TFs were identified at the genome level [11–16]. Most *Ciona* TFs are expressed maternally and were demonstrated to regulate early embryonic development. For example, *brachyury*, belonging to the T-box TF family, regulates the notochord cell specification and morphogenesis, through the downstream genes mediated by cis-regulatory modules [17–20], such as the *nuclear factor of activated T cells 5* (*nfat5*), *T-box2/3* (*Tbx2/3*), and *signal transducer and activator of transcription 5* (*stat5*) [21,22]. Whereas, *FoxD* activates the expression of the *zinc finger of the cerebellum-like* (*ZicL*) part, which binds and promotes the expression of *brachyury* [23–25].

The various TFs in different animals were evolutionarily related to their regulatory ways. For example, *hox* genes were conserved throughout vertebrates [26,27], but in tunicates, the number and arrangement of *hox* genes were quite different, indicating the different roles of *hox* genes in tunicates [26,27].

Styela clava is a globally distributed ascidian species that shows a strong environmental adaptation [28–30]. In this study, we performed genome-wide identification and analysis for TFs in *S. clava*. The expanded and contracted TF families were also identified through comparison with other species. Furthermore, the expression profiles of TFs in *S. clava* were also analyzed. These results provided insights into the understanding of the regulatory mechanisms of TFs in embryogenesis and environmental adaptation in tunicates, and the evolution of the TF families.

2. Results

2.1. Identification of TFs in the S. clava Genome

A total of 17,428 protein-coding genes were identified, 15,734 of which were annotated in the *S. clava* genome [28]. Based on our previous study, we further identified 553 TF genes, which were distributed in 60 TF families in the *S. clava* genome, according to the domain annotation of proteins and the types of DBD. In these 60 TF families, the Cys2His2-type zinc finger protein (zf-C2H2) family had the largest number of TFs, with 154 genes (27.85% of total TFs) (Table 1). The Homeodomain family and the basic helix–loop–helix (bHLH) family were the second and third largest TF family, with 73 (13.20%) and 41 (7.41%) TFs, respectively (Table 1). There were 16 orphan TFs, belonging to the TF family with only one member including the ALL1-fused gene from chromosome 4 (AF-4), the CCAAT-binding transcription factor subunit B_nuclear transcription factor Y subunit alpha (CBFB_NFYA), and interferon regulatory factors-3 (IRF-3), etc. (Table 1).

Table 1. Number of TFs and TF families in *S. clava*, *C. robusta*, *M. oculata*, *O. dioica*, *B. leachii*, *B. schlosseri*, *H. sapiens*, *B. floridae*, and *C. elegans*.

TF Super-family	TF Family	*S. clava*	*C. robusta*	*M. oculata*	*O. dioica*	*B. leachii*	*B. schlosseri*	*H. sapiens*	*B. floridae*	*C. elegans*
	AF-4	1	1	1	1	1	1	8	1	0
	ARID	4	5	7	4	4	2	15	4	5
	BTD	2	2	2	2	2	3	6	2	1
bZIP	bZIP_1	13	16	12	21	15	7	37	14	11
	bZIP_2	6	7	7	8	6	2	12	8	17
	CBF_beta	1	1	1	1	1	1	1	1	1
NF-Y	CBFB_NFYA	1	1	0	1	1	0	1	1	2
	CBFD_NFYB_HMF	8	9	8	13	8	4	45	16	9
	CG-1	0	1	1	0	0	0	2	1	1
	CP2	2	3	2	2	2	3	8	2	1
	CSD	9	5	4	7	4	4	17	2	5
	CSRNP_N	1	1	0	2	1	2	3	1	1

Table 1. *Cont.*

TF Super-family	TF Family	*S. clava*	*C. robusta*	*M. oculata*	*O. dioica*	*B. leachii*	*B. schlosseri*	*H. sapiens*	*B. floridae*	*C. elegans*
MH1	CTF_NFI	2	1	0	0	1	0	4	1	0
	MH1	4	5	7	11	4	5	11	4	7
	E2F_TDP	2	4	4	3	3	2	20	4	5
ETS	Ets	12	15	12	11	14	12	29	12	10
	ETS_PEA3_N	0	0	0	0	0	0	4	1	0
	Forkhead box	23	25	24	26	27	27	55	28	16
	GCFC	2	2	2	1	2	2	3	1	2
	GCM	0	0	0	0	0	0	2	1	0
	GTF2I	0	1	0	0	0	0	15	0	0
	bHLH	41	39	36	22	37	20	102	73	38
bHLH	Myc_N	0	1	1	0	0	1	4	1	0
	SIM_C	0	0	0	0	0	0	2	0	0
	HMG_box	20	20	14	19	18	21	102	29	16
	Homeodomain	73	76	74	70	74	45	264	108	85
Homeobox	CUT	2	1	3	3	3	2	8	3	7
	PBC	1	1	1	1	1	5	11	1	2
	Pou	2	3	3	5	3	2	17	6	3
	HPD	2	2	1	1	1	1	3	0	1
	HSF_DNA-binding	1	1	0	4	1	2	9	5	1
	HTH_psq	6	2	3	0	9	0	2	5	1
IRF	IRF	5	5	6	2	8	5	3	4	0
	IRF-3	1	3	2	0	4	9	9	4	0
	LAG1-DNAbinding	0	0	0	0	0	1	0	0	0
	LRRFIP	0	1	3	1	1	1	5	1	1
	MBD	3	2	4	0	3	2	7	3	2
	Myb_DNA-binding	12	13	11	7	13	11	25	16	7
	NCU-G1	1	0	1	0	1	2	1	1	0
	NDT80_PhoG	1	1	3	0	1	2	2	1	2
	Nrf1_DNA-binding	1	1	1	0	1	1	1	3	0
	P53	3	2	1	1	3	5	3	2	1
	PAX	7	6	5	8	5	2	9	5	9
	PC4	1	1	0	1	1	1	3	2	1
	RFX_DNA_binding	2	3	3	1	3	2	10	5	1
	RHD_DNA_binding	3	2	3	2	4	5	13	2	0
	Runt	1	1	1	1	1	2	3	1	1
	SAND	3	2	1	2	1	2	7	4	4
	SRF-TF	3	2	2	3	2	1	5	3	2
	STAT_alpha	0	0	0	1	0	2	2	0	0
STAT	STAT_binding	2	2	2	2	2	2	5	1	2
	STAT_int	0	1	0	0	0	0	0	2	0
	TAFH	2	1	1	1	1	1	6	3	1
	T-box	10	8	8	8	8	13	19	10	19
	TEA	1	1	1	2	1	1	4	1	1
	TF_AP-2	1	2	1	1	1	2	5	1	4
	TF_Otx	0	0	0	0	0	0	1	0	0
	TIG	11	7	6	12	16	29	18	15	4
	TSC22	2	1	1	0	2	1	4	1	5
	Tub	1	1	1	2	1	1	5	2	2
	Vert_HS_TF	0	0	0	0	0	0	3	0	0

Table 1. *Cont.*

TF Super-family		TF Family	*S. clava*	*C. robusta*	*M. oculata*	*O. dioica*	*B. leachii*	*B. schlosseri*	*H. sapiens*	*B. floridae*	*C. elegans*
zinc finger		THAP	21	8	27	0	40	49	9	7	3
		GATA	5	4	4	5	5	2	17	7	12
		DM	5	2	0	2	7	3	7	8	11
		Hormone_receptor	6	6	6	7	5	9	28	10	151
		Androgen_receptor	0	0	0	0	0	0	1	0	0
	Nuclear Receptor	Oest_receptor	0	0	0	0	0	0	1	0	0
		Prog_receptor	0	0	0	0	0	0	1	0	0
		GCR	0	0	0	0	0	0	1	0	0
		zf-C4	15	12	14	33	11	21	32	21	124
		zf-BED	17	2	6	0	8	9	3	0	9
		zf-C2H2	154	95	193	72	110	310	748	742	58
		zf-C2HC	5	3	2	2	4	5	8	2	1
		zf-LITAF-like	4	1	2	5	3	2	4	11	14
		zf-MIZ	2	2	3	2	1	0	6	2	2
		zf-NF-X1	1	1	2	0	1	1	1	1	1
Total TFs			553	456	557	425	522	693	1867	1240	703
Total TF families			60	64	57	51	61	60	74	64	55

2.2. Comparison Analysis of S. clava TFs

To compare the TFs in *S. clava* with other species, the same approach was utilized to identify TFs in the species genomes. Three species, including human (*H. sapiens*), cephalo-chordate (the lancelet, *Branchiostoma floridae*), nematode (*Caenorhabditis elegans*), and five tunicates, including *C. robusta*, *Molgula oculata*, *Oikopleura dioica*, *Botrylloides leachii*, and *Botryllus schlosseri* were selected for comparison analysis. The results showed that there were 51 families shared by *S. clava*, *H. sapiens*, *B. floridae*, and *C. elegans* (Figure 1A, Table S1). There were also 51 families shared among the five ascidian species (Figure 1B, Table S1). There were 44 gene families shared between these two groups. Seven gene families including Transcriptional Coactivator p15 (PC4), Cysteine/serine-rich nuclear protein N-terminus (CSRNP_N), DM, CBFB_NFYA, Helix-turn-helix (HTH_psq), Heat shock factor type DNA binding (HSF_DNA-binding), and the MIZ-type zinc finger (zf-MIZ) were only shared among the species in Figure 1A, while the other seven gene families including RHD_DNA_binding, BEAF and DREF-type zinc finger protein (zf-BED), AF-4, Homeo-prospero domain (HPD), IRF-3, IRF, and NLS-bindingm and DNA-binding, and the dimer-ization domains of Nrf1 (Nrf1_DNA-binding) were only shared among the species in Figure 1B (Table S1).

To explore the expansion and contraction of TFs families, we compared the TFs families in tunicates with other species. Nine human TF families could not be found in tunicates and *C. elegans* (Table 1). Among these nine families, two were found in *B. floridae* (Table 1). The thanatos-associated proteins (THAP) family was increased in Stolidobranchia ascidians, including *S. clava*, *M. oculata*, *B. leachii*, and *B. schlosseri*, compared to *H. sapiens* (Table 1). Compared to *B. floridae* and *C. elegans*, general transcription factor IIi (GTF2I), Signal Transducer, and the Activator of Transcription binding (STAT_binding) families were only found in tunicates (Figure 1A, Table 1 and Table S1).

In the *S. clava* genome, six TF families had the largest number of TFs among the six tunicates, including the bHLH, the Cold shock domain (CSD), the CTF/NF-I family transcription modulation region (CTF_NFI), NHR1 homology to TAF (TAFH), zf-BED, and Sp100, AIRE-1, NucP41/75, and DEAF-1 (SAND) families (Table 1). Two TF families in the *S. clava* genome had the least number of TFs among the six tunicates, including the forkhead box (fox) and Leucine-Rich Repeat in the Flightless-Interaction Protein (LRRFIP) families (Table 1). The number of TFs in *S. clava* were more than that in *C. robusta*, *O. dioica*, and *B. leachii* (Table 1). The expansion mainly concentrated on the zf-C2H2 family, in which 59, 82, and 44 more TFs were identified in *S. clava* than that in *C. robusta*, *O. dioica*, and *B. leachii*, respectively (Table 1). Overall, comparison analysis revealed that the gene number

of the zf-C2H2 family showed the most significant change among tunicates, indicating that the zinc finger family genes were under adaptative changes.
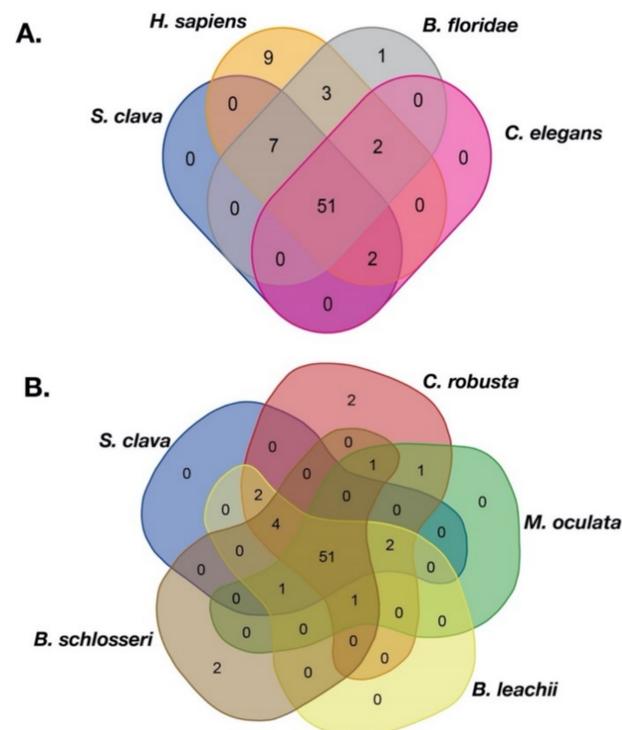

Figure 1. Veen of the TF families among different species. (**A**) TF families among *S. clava* (blue), *H. sapiens* (dark yellow), *B. floridae* (gary), and *C. elegans* (pink). (**B**) TF families among *S. clava* (blue), *C. robusta* (red), *M. oculata* (green), *B. leachii* (yellow), and *B. schlosseri* (brown). The transcription factor families contained in each part are listed in Table S1.

2.3. Expression of S. clava TFs at Different Developmental Stages

To explore the expression profiles of TFs during the development of *S. clava*, we acquired the fragments Per Kilobase of gene per million mapped reads (FPKM) value of each TF in 2-cell–8-cell embryos (2–8 cells), gastrula embryos (gast), neurula embryos (neu), tailbud-stage embryos (tb), hatched swimming larvae (hsl), tail-regressed larvae (trl), and metamorphic juveniles (mj) by RNA-sequencing [28]. These data were further validated through quantitative real-time PCR. The results showed that the relative expression levels of the randomly selected genes were coincident with the results of FPKM values (Figure S1). To analyze the expression profiles of TFs in *S. clava* during early development, we clustered 547 TFs by the weighted correlation network analysis (WGCNA) method [31] and visualized the expression level by a heat map (Figure 2). The TF genes were clustered into nine modules, labeled with different colors, based on the results of the WGCNA analysis. The turquoise module contained the most TFs, nearly one third of TFs, which were highly expressed at the 2–8 cells stage and the gast stage (Table 2). A total of 290 TF genes (52.44% of total TF genes) were expressed (with FPKM value >10) at the 2–8 cell stage in *S. clava*, about 161 of them were classified in the turquoise module. The expression level of these genes in this module decreased from the embryo to the larvae stage. The brown module contained nearly one fifth of TFs that were mainly expressed at the tb stage (Table 2, Figure 2).
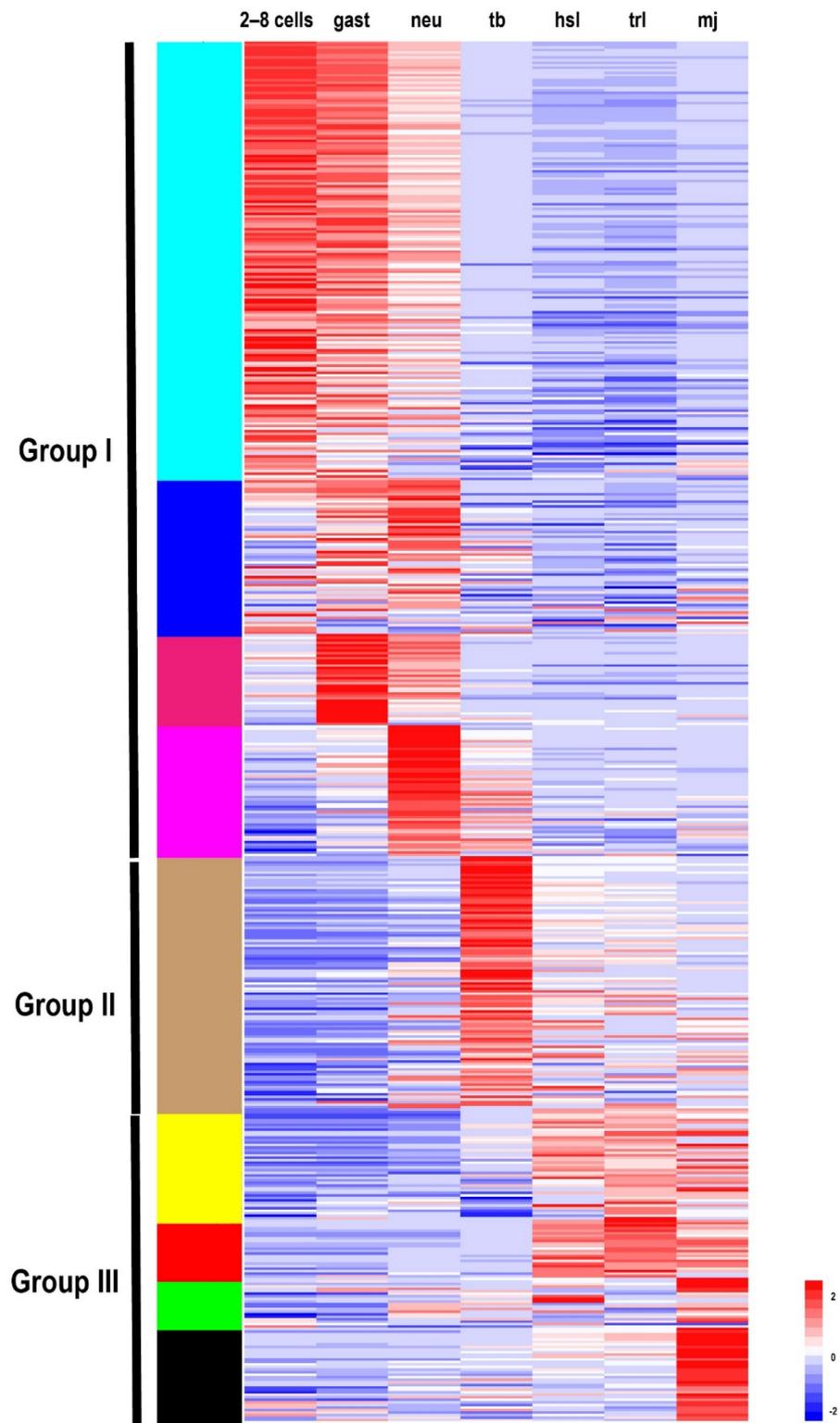
Figure 2. Expression patterns of TFs in *S. clava* genome—547 TFs in *S. clava* genome are shown in the heat map. They are classified into nine modules, including turquoise, blue, magenta, pink, brown, yellow, red, green, and black module through the WGCNA analysis. The turquoise, blue, magenta, and pink modules are classified into Group I, in which the TFs were highly expressed during 2–8 cells stage to the neural stage; the brown module is classified into Group II, in which TFs were highly expressed during the tailbud stage; the yellow, red, green, and black modules are classified into Group

III, in which TFs were highly expressed from the hatched swimming larvae stage to the metamorphosis juvenile stage. The scale bar indicates the centered FPKM values. The abbreviation of different developmental stages are as follows—2-cell–8-cell embryos (2–8 cells), gastrula embryos (gast), neurula embryos (neu), tailbud-stage embryos (tb), hatched swimming larvae (hsl), tail-regressed larvae (trl), and metamorphic juveniles (mj).

Table 2. Number of TFs in the WGCNA modules.

TF Family	Turquoise	Blue	Magenta	Pink	Brown	Yellow	Red	Green	Black	Total
AF-4	1	0	0	0	0	0	0	0	0	1
ARID	1	1	1	0	1	0	0	0	0	4
BTD	1	0	0	0	0	1	0	0	0	2
bZIP_1	3	0	0	0	2	3	0	0	5	13
bZIP_2	3	2	0	0	0	0	0	0	1	6
CBF_beta	0	0	0	0	0	0	0	1	0	1
CBFB_NFYA	1	0	0	0	0	0	0	0	0	1
CBFD_NFYB_HMF	2	1	1	2	0	2	0	0	0	8
CP2	1	0	0	0	0	0	0	0	1	2
CSD	2	1	1	0	1	1	3	0	0	9
CSRNP_N	0	0	0	1	0	0	0	0	0	1
CTF_NFI	2	0	0	0	0	0	0	0	0	2
CUT	0	1	0	0	1	0	0	0	0	2
DM	1	1	2	0	1	0	0	0	0	5
E2F_TDP	1	0	0	1	0	0	0	0	0	2
Ets	3	0	0	3	2	3	0	1	0	12
Forkhead box	4	1	0	3	5	4	3	1	2	23
GATA	4	0	0	0	0	1	0	0	0	5
GCFC	2	0	0	0	0	0	0	0	0	2
HLH	10	7	0	3	13	2	3	0	3	41
HMG_box	9	2	1	2	2	0	0	2	2	20
Homeodomain	5	4	2	6	37	8	3	4	3	72
Hormone_recepter	2	0	0	0	0	2	1	0	1	6
HPD	0	0	0	0	1	0	0	0	1	2
HSF_DNA-binding	1	0	0	0	0	0	0	0	0	1
HTH_psq	0	2	1	0	1	0	0	2	0	6
IRF	2	0	0	0	0	0	0	0	3	5
IRF-3	0	0	0	0	0	1	0	0	0	1
MBD	1	0	0	0	1	1	0	0	0	3
MH1	3	0	0	1	0	0	0	0	0	4
Myb_DNA-binding	6	3	1	0	1	0	0	0	0	11
NCU-G1	1	0	0	0	0	0	0	0	0	1
NDT80_PhoG	0	0	0	0	0	1	0	0	0	1
Nrf1_DNA-binding	1	0	0	0	0	0	0	0	0	1
P53	2	0	0	0	0	1	0	0	0	3
PAX	2	0	0	1	3	0	1	0	0	7
PBC	1	0	0	0	0	0	0	0	0	0
PC4	0	0	1	0	0	0	0	0	0	1
Pou	1	0	0	0	1	0	0	0	0	2
RFX_DNA_binding	1	0	0	1	0	0	0	0	0	2
RHD_DNA_binding	0	0	0	0	0	1	0	0	2	3
Runt	0	0	0	0	0	0	0	0	1	1
SAND	1	1	0	0	1	0	0	0	0	3
SRF-TF	2	1	0	0	0	0	0	0	0	3
STAT_alpha	0	0	0	0	0	0	0	0	0	0
STAT_binding	0	0	0	0	0	1	0	0	1	2

Table 2. *Cont.*

TF Family	Turquoise	Blue	Magenta	Pink	Brown	Yellow	Red	Green	Black	Total
TAFH	1	0	0	0	0	0	0	0	1	2
T-box	1	1	3	1	3	1	0	0	0	10
TEA	0	0	0	0	0	0	0	1	0	1
TF_AP-2	0	1	0	0	0	0	0	0	0	1
THAP	6	1	0	4	6	0	0	2	2	21
TIG	6	0	1	0	2	0	0	1	1	11
TSC22	2	0	0	0	0	0	0	0	0	2
Tub	0	0	0	0	1	0	0	0	0	1
zf-BED	3	2	0	7	2	2	0	1	0	17
zf-C2H2	67	24	20	12	7	5	7	3	5	150
zf-C2HC	2	2	0	0	1	0	0	0	0	5
zf-C4	0	1	1	3	2	3	2	1	2	15
zf-LITAF-like	1	1	0	1	1	0	0	0	0	4
zf-MIZ	1	0	0	0	1	0	0	0	0	2
zf-NF-X1	1	0	0	0	0	0	0	0	0	1
Total	174	61	36	52	100	44	23	20	37	

Based on the expression profiles of each module, we categorized nine modules into three groups. Group I contained TFs that were highly expressed during 2 cell stage to neu stage, including turquoise, blue, magenta, and pink modules. Group II contained TFs that were highly expressed during the tb stage, including the brown module. Group III contained TFs that were highly expressed during the hsl stage to mj stage, including the yellow, red, green, and black modules. More than half of the TFs (58.94%) were clustered into Group I, almost three times than that in Group II (18.25%) and Group III (22.81%) (Table 2). In Group I, zf-C2H2, homeodomain, and the bHLH family were the largest TFs (Table 2). In Group II, homeodomain and the bHLH families were the largest TFs (Table 2). In Group III, homeodomain, zf-C2H2, and fox families were the largest TFs (Table 2). Overall, the homeodomain TF family was dispersed into each group, while the Fox family was highly expressed during the larval development in Group III.

According to the expression correlation heat map of TFs, we found that five boxes of TFs showed significant correlation, which are indicated in box A to E (Figure 3). We found that the TFs in Group I were mainly distributed in box A, box D, and box E, TFs in Group II and Group III were mainly distributed in box C and box B, respectively (Table S2). The expression correlation results showed that the TFs expressed in the embryogenesis and larvae development had a low correlation, indicating the discrepancy of TFs in regulating the process of embryogenesis and larvae development.

2.4. Hox Genes

Hox genes are important members of the homeobox TF family and share common development mechanisms in regulating the anteroposterior body axis [8]. In vertebrates, there are 13 paralog groups (PGs) for *Hox* genes. Through phylogenetic analysis and blast annotation, we identified eight *Hox* genes in *S. clava*, including *ScHox1*, *ScHox2*, *ScHox3*, *ScHox4*, *ScHox5*, *ScHox10*, *ScHox12*, and *ScHox13*, which were classified into eight PGs (Figure S2). The molecular phylogenetic tree provided a convincing evidence that *ScHox1*, *ScHox2*, *ScHox3*, *ScHox10*, *ScHox12*, and *ScHox13* belonged to the PG1, PG2, PG3, PG10, PG12, and PG13, respectively (Figure S2). *ScHox4* and *ScHox5* were identified by sequence alignment. The results showed that they belonged to PG4 (*e*-value = 2×10^{-57}) and PG5 (*e*-value = 5×10^{-52}), respectively. All *Hox* genes identified in the *S. clava* genome were distributed on one single chromosome according to the Hi-C result [28]. *Hox* genes were also distributed on one single chromosome in *Halocynthia roretzi* with similar arrangement, and on two chromosomes in the *C. robusta* genome (Figure 4A). In other tunicates mentioned in this study, the *Hox* genes were distributed on several scaffolds (Figure 4A). In *H. sapiens*, four Hox clusters existed in the genome and were distributed on four chromosomes

(Figure 4A). *B. floridae* contained all 13 kinds of *Hox* genes and were distributed on one single chromosome (Figure 4A). In tunicates, there are nine *Hox* genes in *C. robusta*, *O. dioica*, *H. roretzi*, and *B. schlosseri* genome [32–35]. There are six, seven, and eight Hox genes in the *M. oculata*, *B. leachii*, and *S. clava* genome [27,35], respectively (Figure 4A).
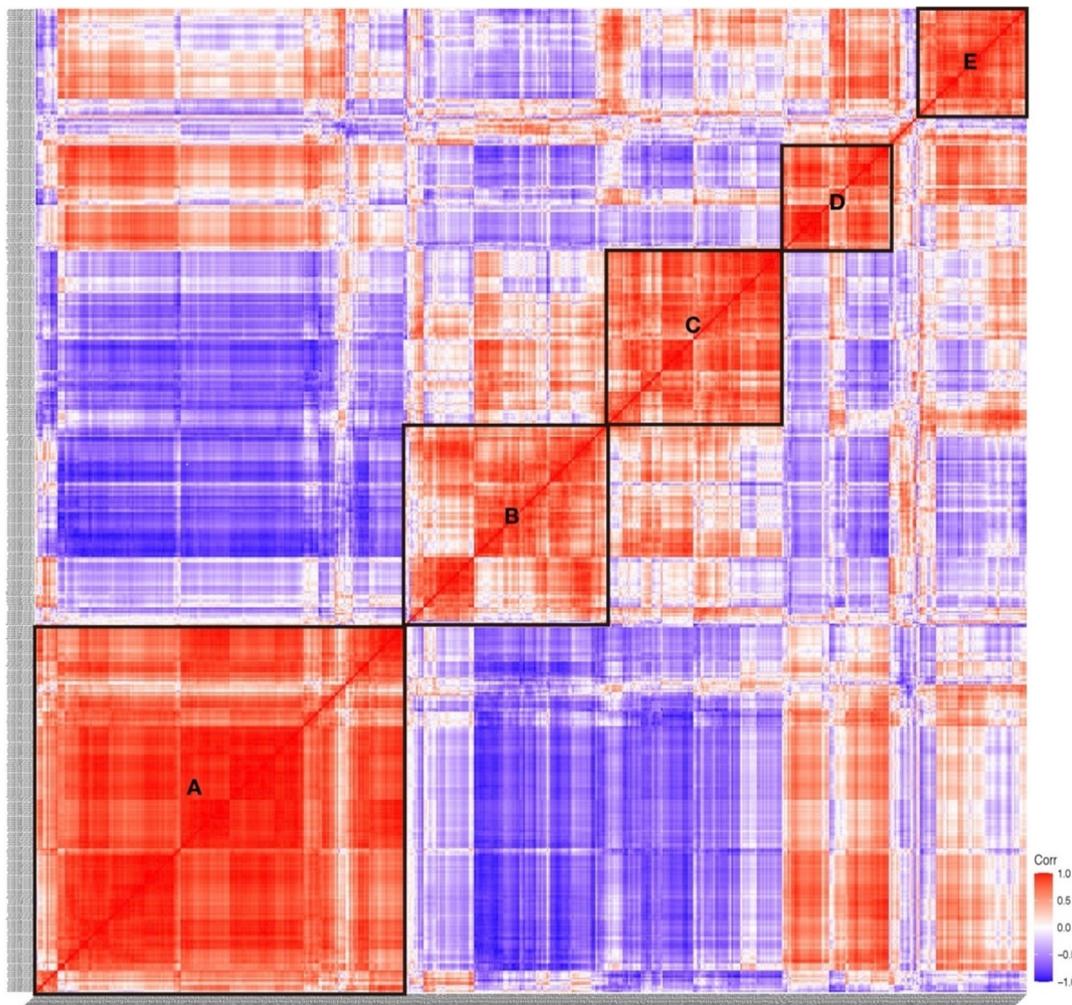

Figure 3. Expression correlation of TFs in *S. clava*. The correlation score showed between −1.0 (deep blue) to 1.0 (deep red). The red color indicated that these two TFs showed positive expression correlation, and the blue color showed negative expression correlation. Five significant correlation groups are indicated by box A to box E.

The expression profiles of *Hox* genes in *S. clava* showed that *ScHox4* and *ScHox12* were initially expressed at the neu stage, *ScHox1*, *ScHox10*, and *ScHox13* were initially expressed at the tb stage, *ScHox3* was initially expressed at the hsl stage, and expression values of *ScHox2* and *ScHox5* were low during early development (Figure 4B). Expression of *ScHox13* were restricted at the tb stage (Figure 4B). While *ScHox12* were not expressed at the hsl stage and the trl stage, after initial expression at the neu stage (Figure 4B). Among the sub-cluster of *ScHox2*, *ScHox3*, and *ScHox4*, *ScHox4* was initially expressed at 2–8 cells stage, earlier than the expression of *ScHox3* and *ScHox2* that were expressed at the gast stage and tb stage (Figure 4B).

2.5. Zinc Finger Family

Zinc finger superfamily contains the greatest number of TFs in the *S. clava* genome (Table 1). The quantity variation of zinc finger TFs was the main contributor to the different number of TFs between species that we mentioned. There were 11 families in the zinc finger superfamily, including THAP, zf-BED, zf-C2H2, etc. (Table 1). Among these families, the

THAP family was expanded in the tunicate genome and the zf-BED family was expanded in the *S. clava* genome, among the nine species we analyzed (Table 1).
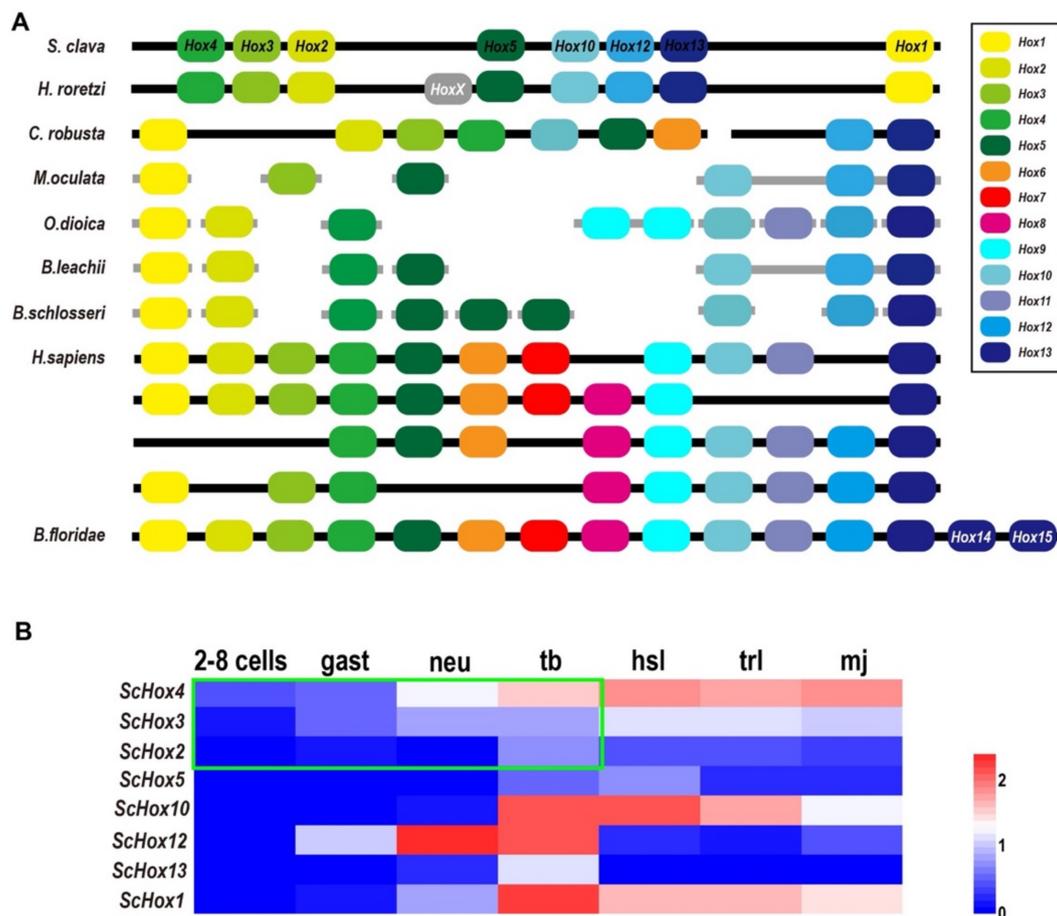
Figure 4. Comparison of *Hox* genes in different species and the expression patterns of *ScHox* genes. (**A**) Schema depicting *Hox* gene linkages retained in the genomes of seven tunicates, *H. sapiens* and *B. floridae*. Orthologous *Hox* genes are indicated by the same color of rounded rectangles and the legend is showed on the right. *HoxX* gene was an unclassified *Hox* gene in the *H. roretzi* genome. The spatial distribution of *Hox* genes in different species on the chromosome or scaffold are indicated by thick black line or thick gray line, respectively. (**B**) Heat map of the *ScHox* genes. Gene names are shown on the left. The scale bar indicates the FPKM values, which are dealt with $\log_{10}(\text{FPKM} + 1)$ but not centered. The green box shows a subcluster-level temporal co-linearity (STC) expression pattern among the *ScHox2*, *ScHox3*, and *ScHox4* cluster. Shorthand of different developmental stage, which is showed above the heat map, is mentioned above.

The Zf-C2H2 family was the largest family in the *S. clava* genome, and also in the other eight species we analyzed (Table 1). According to the domain analysis, there are two kinds of zf-C2H2 proteins in the *S. clava* genome, including ZBTB and zf-C2H2. The ZBTB proteins were characterized by containing two domains, the BTB domain and zf-C2H2 domain. According to the feature of ZBTB, we identified 12 ZBTB genes in the *S. clava* genome, more than the other six tunicates analyzed (Figure 5A).

The *zf-C2H2* genes were the majority of the zf-C2H2 family. There are 132 *zf-C2H2* genes mapped on 16 chromosomes of the *S. clava* genome (Figure 5B). The Chr 5 processed the greatest number of *zf-C2H2* genes (Figure 5B). Those *zf-C2H2* genes, which were highly expressed after the neu stage (red rectangles), were located on nine of 16 chromosomes, among which Chr 3 was the most common. The wide distribution of the zf-C2H2 genes indicates the crucial regulatory roles of the zf-C2H2 domain on activating the downstream gene expression in various biological processes.
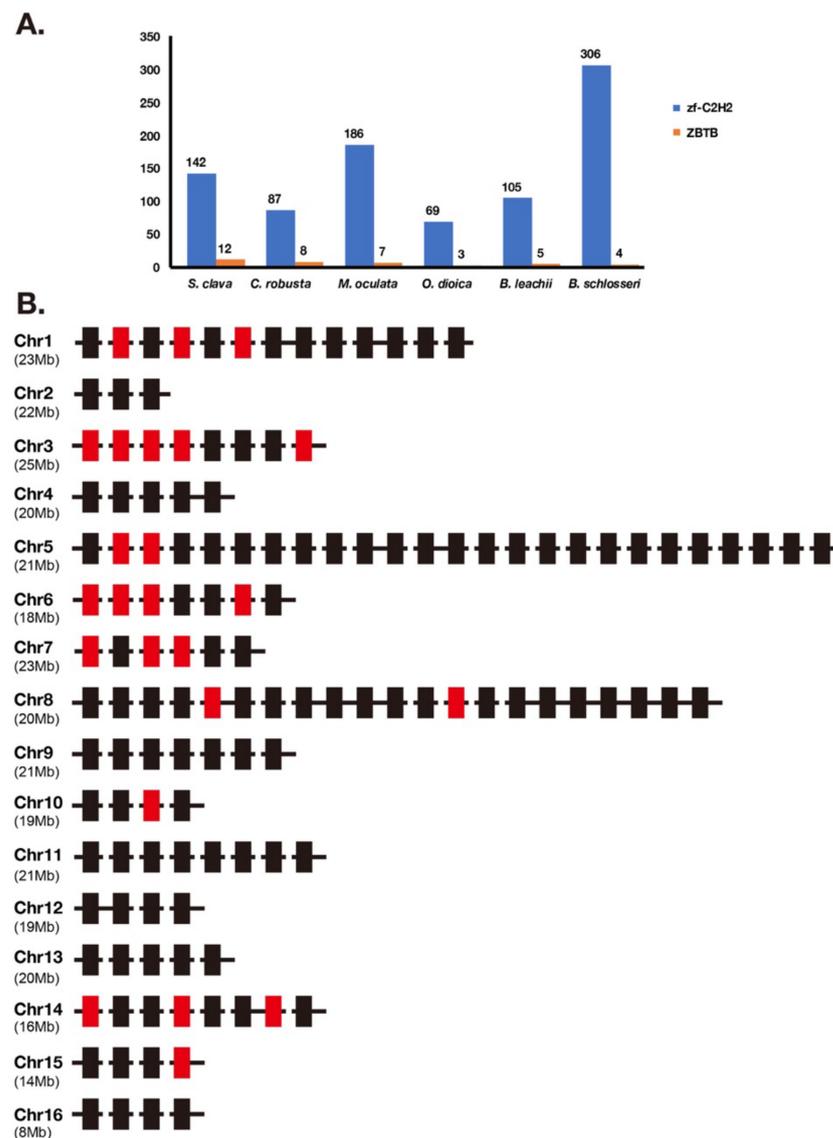
Figure 5. Classification of *zf-C2H2* genes. (**A**) Statistics of *zf-C2H2* genes and *ZBTB* genes in zf-C2H2 family in *S. clava*, *C. robusta*, *M. oculata*, *O. dioica*, *B. leachii*, and *B. schlosseri*. The blue volume indicates the number of *zf-C2H2* genes, and the orange volume indicates the number of *ZBTB* genes. (**B**) The distribution of *zf-C2H2* genes on chromosome. The rectangles indicated the *zf-C2H2* genes and the colors indicate the expression levels of each *zf-C2H2* genes, according to the WGCNA analysis (mentioned in Figure 2, the black and red rectangles indicate the high expression of the *zf-C2H2* genes during 2 cells to neu stages, and during the tbl to mj stages, respectively). The chromosomes are labeled on the left. The black horizontal line indicates scaffold.

2.6. Forkhead Box Family

A total of 23 *Fox* genes were identified in the *S. clava* genome and they could be grouped into 15 *Fox* subclasses (Figure 6 and Figure S3). The largest *Fox* subclass was FoxI, which contained four *ScFox* genes (Figure 6). The expression profiles of the *ScFox* genes were varied during early development (Figure 6, Table 2). Five of them were highly expressed before the tb stage, eight were highly expressed during the tb stage, and ten were highly expressed during larvae development (Figure 6). In the *S. clava* genome, we could not find the homologous genes of *FoxB*, *FoxK*, *FoxL*, and *FoxS*. The expression profiles showed that the *ScFox* genes were highly expressed in several stages during early development. This expression pattern was also presented in the Yesso scallop (*Patinopecten yessoensis*) [36] and sea urchin [37]. Similar patterns presented in early development in

these species indicated that *ScFox* genes are widely involved in regulating the processes of early development of embryogenesis.
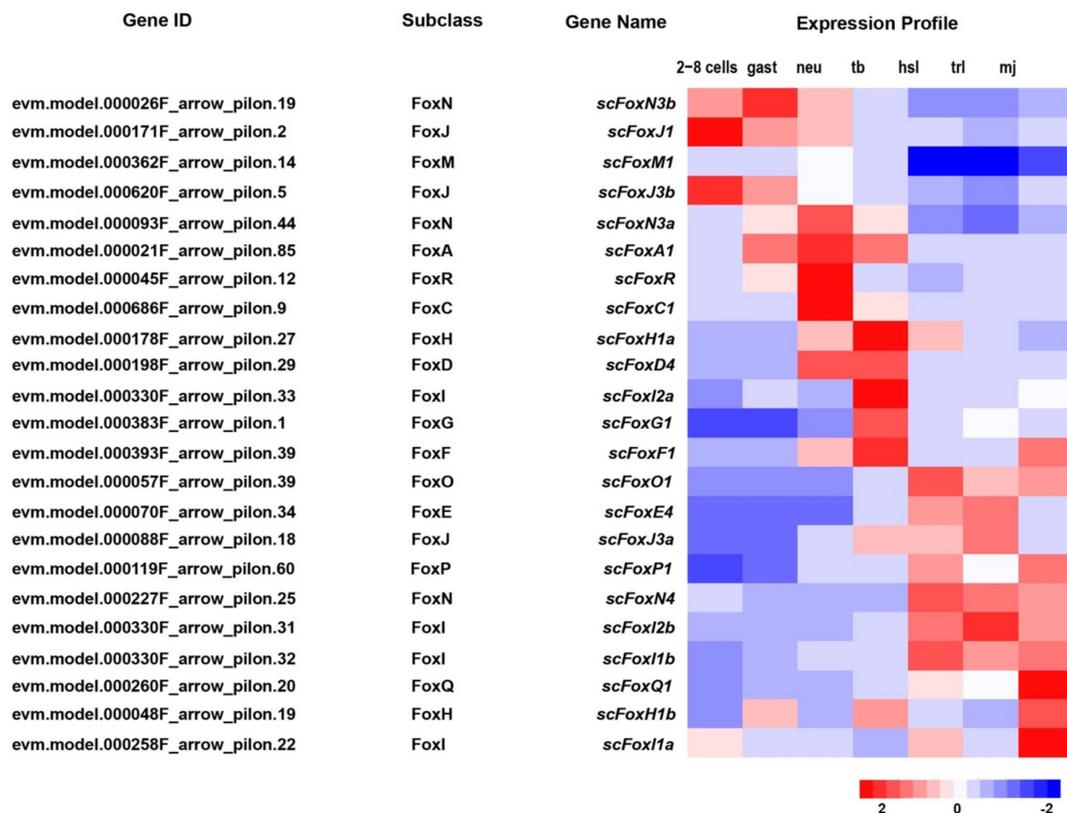
Figure 6. Identification and expression of *ScFox* genes in *S. clava*. The Gene IDs, subclasses, Gene names of 23 *scfox* genes are listed (left side). The expression profiles for each gene through different stages (2–8 cells, gast, neu, tb, hsl, trl, and mj stage) are shown on the right side. The scale bar indicates the centered FPKM values.

3. Discussion

Tunicates, as the closest relatives of vertebrate, show a special rates and patterns of molecular evolution [38]. Identification and analysis of important gene families were also performed, based on the sequenced genomes in several tunicate species [35,39]. In this study, we screened and identified 553 TFs in the *S. clava* genome and revealed their potential roles in environmental adaptation and early embryonic development, through their expression profiles.

S. clava showed a broad tolerance of environmental conditions [28]. Compared to *B. floridae* and *C. elegans*, there are more CSD TFs and immune response-related TFs, (such as *IRF* genes) in *S. clava*. CSD is a crucial characteristic of cold shock proteins (CSP), which were upregulated under the low temperature in *S. clava* [28]. *IRF* genes play an important role in immune response in *C. savignyi* [40,41]. The expansion of these TFs might help to improve the fitness to low temperature conditions and improve their ability of immune response to adapt to new environment.

TFs highly expressed during 2–8 cells stage is an important group that we identified in this study. Through gene ontology (GO) enrichment analysis, the biology processes of maternal TFs mainly concentrate in the response/cellular response, signaling, and regulation of biological process (Figure S4). More than half of TFs involved in these processes are annotated as nuclear receptors, which regulate the activation of genes and control many biological processes, such as cell proliferation, cell cycle, and metabolism [42]. For example, thyroid hormone receptor α1 (THRα1) controls the cell proliferation through the Wnt /β-catenin pathway [43]. In *C. robusta*, the combination of GATA and Ets induced FGF9/16/20 in the neural tissue [44–47]. TFs highly expressed during 2–8 cells stage in *S.*

clava might act as "responders" and "signal transmitters" to control the expression of the downstream genes and regulate early embryogenesis.

Proportion of the expressed TFs had increased from 2–8 cells stage to the tb stage (Figure S5) and was stable from the hsl stage to the mj stage. In the previous study, the different species showed very similar overall TF expression patterns, with TF expression increasing during the initial stages of development among *D. rerio*, *C. robusta*, *D. melanogaster*, and *C. elegans* [48]. We compared the TF expression between *S. clava* and *C. robusta* and found similar patterns, with TF expression increasing from 2–8 cells stage to tb stage (Figure S4). These results indicate that the most TFs might play a conserved role in early development.

S. clava larvae undergo metamorphosis after hatching. TFs, whose expression are significantly up-regulated after hatching might be involved into the regulation of this process. We classified those genes into two types. One is immune response TFs, which include cyclic AMP-dependent transcription factor (ATF-3), IRF1, IRF4, and B-cell lymphoma 6 (BCL6), which have strong activities in immune response regulation and maturation of the immune system [49–52]. High expression of these TFs indicates that the immune system and inflammatory reaction are involved in the metamorphosis of *S. clava*, and these immune responses also appeared in ascidian *Boltenia villosa* [53]. The others are the thyroid hormone, retinoic acid signaling TFs, and nuclear receptors, including thyroid hormone receptor alpha (THRα), vitamin D 25-hydroxylase (CYP2R1), and retinoic acid receptor alpha (RARα). Thyroid hormone and retinoic acid signaling pathways were demonstrated to be essential in the metamorphosis of *S. clava* [28,54]. Nuclear receptors are a superfamily of TFs that function in the regulation of various metabolic processes [42,55]. The metamorphosis process is regulated through interaction of hormone and nuclear receptors, which also existed in fishes [56]. Nuclear receptors, involved in the TH and RA signaling pathways, had potential in regulating metamorphosis in *S. clava*, according to the expression profiles.

Eight *ScHox* genes were identified in the *S. clava* genome and they were located in one chromosome. This was consistent with the location of *Hox* genes in *H. roretzi*, while *Hox* genes in genome of *C. robusta* were located in two chromosomes [27,33,35]. The similar location of *Hox* genes on the chromosome between *S. clava* and *H. roretzi* might be related to their close phylogenetic relationship [28]. Whole-cluster temporal collinearity (WTC) is a typical expression characteristic for *Hox* genes in vertebrate [57,58]. However, it is not applicable in tunicates. Expression profile of *Hox* genes in *S. clava* showed a subcluster temporal collinearity (STC), which was reflected in the *ScHox*2–4 cluster. STC was also presented in other invertebrate species, such as scallop *P. yessoensis* [59] and ascidian *C. robusta* [32]. In *C. robusta*, the *Hox* genes were spatially expressed in the central nervous system of larvae and the gut of the juvenile [32]. *CrHox10* and *CrHox12* play important roles in the neuronal and tail development in *C. robusta* [60]. These investigations suggest the potential roles of *Hox* genes in the development of the tail and the nervous system in *S. clava*.

The expression profiles of *Hox1*, *Hox3*, *Hox4*, *Hox10*, and *Hox12* genes in *S. clava* were similar to the expression profiles of these genes in *C. robusta* [32]. The expression of *Hox2* and *Hox5* were very low in *S. clava*. The expression profiles of *Hox13* were different between *S. clava* and *C. robusta*. *ScHox13* was expressed in the tb stage, but *CrHox13* was detectable only in the juvenile [32]. However, the number and distribution of *Hox* genes seemed not to be related to their body plan at the larval stage.

Fox genes are characterized by the highly conserved forkhead motif, which is known to be a "winged-helix" DNA-binding domain [61,62]. *Fox* genes play important roles in embryogenesis and metabolism [62]. A total of 22 *Fox* genes were identified in the genome of sea urchin [37]. Number of *Fox* genes in *S. clava* genome was nearly same as that of the sea urchin. Compared to the sea urchin, *FoxE*, *FoxH*, and *FoxR* subclasses in *S. clava* were found, and *FoxI* subclass was expanded, but *FoxB*, *FoxK*, and *FoxL* subclasses were not found. In *S. clava*, genes in the *FoxE* and *FoxI* subclasses were highly expressed after hatching. In vertebrates, *FoxE4*, *FoxI1*, and *FoxI2* play important roles in len, otic placode,

and retina formation and development, respectively [63–65]. *ScFoxE* and *ScFoxI* might have similar functions in the regulation of neuro and sensory organ formation.

In conclusion, we identified 553 TFs belonging to 60 TF families in the *S. clava* genome. The expression profiles provided a possible clue for the functions of different TF genes in embryogenesis, environmental adaptation, and metamorphosis in *S. clava* (Figure 7). Our study provides gene resources and a new perspective to understand the evolution and function of TFs in the leathery sea squirt.
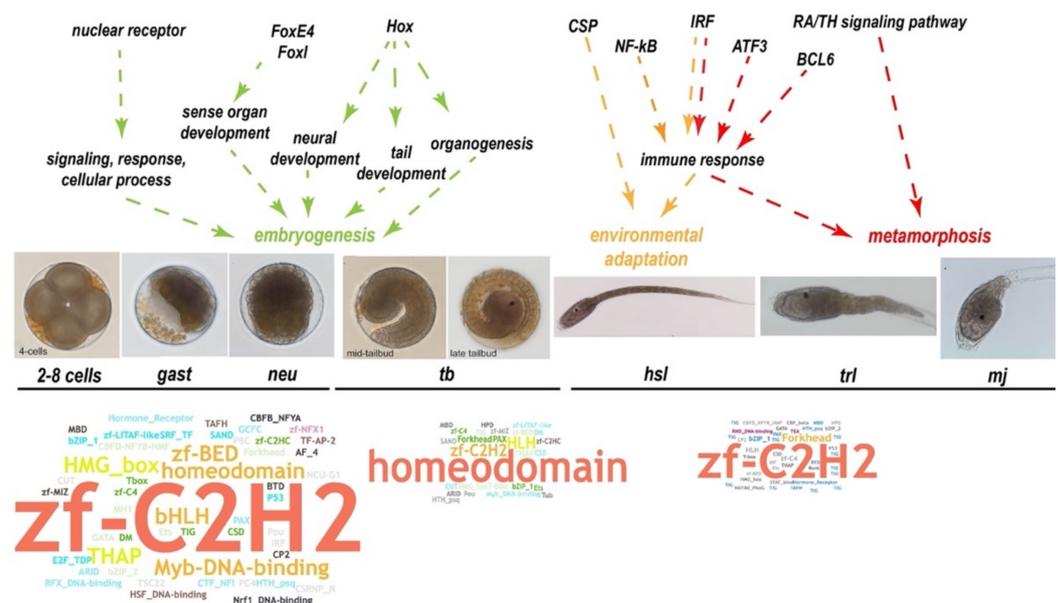
Figure 7. Summary of the TF expression and their potential roles in the embryogenesis and larval development of *S. clava*. Expression profiles indicates that TFs play essential roles during embryogenesis and larval development. We showed different stages of embryos at the bottom of the figure, including 2–8 cells stage, gast stage, neu stage, mid-tailbud stage, late-tailbud stage, hsl stage, trl stage, and mj stage. We summarized the predicted function of TFs genes into three parts and showed it in different colors, including embryogenesis (green), environmental adaptation (yellow), and metamorphosis (red). The word cloud pictures show the family of the expressed TF genes at different stages. The font size indicates the number of expressed TF genes in the TF families.

4. Materials and Methods

4.1. Animals and Embryos

The adults of *S. clava* were collected from Weihai City, China, and cultured in seawater at 18 °C in the laboratory. The eggs and sperm were collected separately from different individuals for fertilization at room temperature. The embryos and larvae were collected for RNA extraction at different stages. The study was approved by the Ocean University of China Institutional Animal Care and Use Committee (OUC-IACUC) prior to the initiation of the study (Approval number: "2021-0032-0012", 15 April 2019). All experiments and relevant methods were carried out in accordance with the approved guidelines and regulations of OUC-IACUC.

4.2. Transcriptome Sequencing and WGCNA Analysis

The transcriptome sequencing and data were described in a previous study [28]. The co-expression gene network for 21 transcriptomic datasets was constructed using the R package WGCNA, with the parameters of softPower = 12, minimum module size = 300, cutting height = 0.99, and deepSplit = F [31]. The genes that were expressed in at least one developmental stage were used for network construction.

4.3. Identification and Classification of TFs

The TFs were identified and classified, based on the conserved DBDs. After obtaining all protein sequences at the genome level, we removed random sequences through CD-HIT (v4.6.8) [66] and analyzed the protein domain according to the Hidden Markov Model (HMM) profiles from the Pfam database (version 31.0) [67], applying the hmmscan program in the HMMER package(v3.1b2) [68]. 85 Pfam ID of DBDs from REGULATOR database (http://www.bioinformatics.org/regulator/page.php?act=family, accessed on 8 December 2020) and the Animal TFDB database (http://bioinfo.life.hust.edu.cn/AnimalTFDB/#! /species, accessed on 8 December 2020) were used for analysis [9,10]. The Pfam IDs were showed in Table S3. Most genes containing DBDs were screened with the e-value threshold of 10^{-4} and were regarded as TFs. Some of thresholds of DBD were chosen according to the threshold in Animal TFDB 3.0 [10], including the threshold of the bHLH domain with 10^{-2}, thresholds of the HMG, Homeodomian, zf-BED, zf-C2H2 domains with 10^{-3}, and the thresholds of the zf-CCCH domains with 10^{-20}. The TFs were classified into different families, according to the DBDs.

4.4. Heat Map, Phylogenetic Analysis, Domain Analysis, and Expression Correlation Analysis

We normalized the FPKM values through log10(FPKM + 1) and imported the normalized values into the Heml software (v1.0) [69]. We performed molecular phylogenetic tree analysis using the MEGA software (v7.0) [70] through the Maximum-Likelihood (ML) method and beautified the figure through iTOL (https://itol.embl.de/, accessed on 13 December 2020) online [71]. The multiple sequence alignments were conducted by Clustal W [72]. The gene expression correlation Heat map were constructed by the corrplot package and the ggcorrplot package on R studio. The correlation coefficient indicated the correlation between two genes. The correlation coefficient was greater than 0 for positive correlation (shown in red), and less than 0 for negative correlation (shown in blue). (https://github.com/taiyun/corrplot, accessed on 21 December 2020).

4.5. Quantitative Real-Time PCR

Total RNA was extracted through the TRIZOL method. Reverse transcription experiments were performed according to HiScript II Q RT SuperMix for QPCR (+gDNA wiper) (Vazyme R223-01, Nanjing, China). 18S rRNA was chosen as an internal standard. The primers used in experiments are shown in Table S4. Quantitative real-time PCR was carried out by the ChamQ SYBR Color QPCR Master Mix (Vazyme Q411-01, Nanjing, China). The relative gene expression levels were calculated using the comparative Ct method with the formula $2^{-\Delta\Delta Ct}$ [73].

4.6. Gene Ontology (GO) Enrichment Analysis

The GO enrichment analysis was performed by the OmicShare tools online (https: //www.omicshare.com, accessed on 23 September 2020). The target genes were maternal expressed TF genes, the background genes were all TF genes. We obtained the top 20 of GO enrichment results for further analysis and made the bubble chart.

4.7. Data Availability

The genome sequences of *S. clava* were deposited in NCBI, under the BioProject number PRJNA523448. The transcriptome data of *S. clava* used in this study were also deposited in the NCBI SRA database, with the accession numbers SRR8599814 to SRR8599834.

The genome resources of *C. robusta*, *H. sapiens*, and *C. elegans* were downloaded from the Ensembl (https://asia.ensembl.org/index.html, version101, accessed on 1 July 2020) [74]. The genome resources of *O. dioica* and *B. floridae* were acquired from NCBI (https: //www.ncbi.nlm.nih.gov/, accessed on 1 July 2020), and the genome resources of *M. oculata*, *B. leachii*, and *B. schlosseri* were acquired from ANISEED (https://www.aniseed.cnrs.fr/, accessed on 1 July 2020) [75].

References

1. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [CrossRef] [PubMed]
2. Mitchell, P.J.; Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **1989**, *245*, 371–378. [CrossRef] [PubMed]
3. Bejerano, G.; Pheasant, M.; Makunin, I.; Stephen, S.; Kent, W.J.; Mattick, J.S.; Haussler, D. Ultraconserved elements in the human genome. *Science* **2004**, *304*, 1321–1325. [CrossRef] [PubMed]
4. Carroll, S.B. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* **2008**, *134*, 25–36. [CrossRef]
5. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237–1251. [CrossRef]
6. Arnold, H.-H.; Braun, T. The role of Myf-5 in somitogenesis and the development of skeletal muscles in vertebrates. *J. Cell Sci.* **1993**, *104*, 957–960.
7. Peter, J. Nuclear Factor-kB—A pivotal transcription factor in chronic inflammatory diseases. *Mech. Dis.* **1997**, *336*, 1066–1071.
8. Mallo, M. Reassessing the Role of Hox Genes during Vertebrate Development and Evolution. *Trends Genet.* **2018**, *34*, 209–217. [CrossRef]
9. Wang, K.; Nishida, H. REGULATOR: A database of metazoan transcription factors and maternal factors for developmental studies. *BMC Bioinform.* **2015**, *16*, 114. [CrossRef]
10. Hu, H.; Miao, Y.R.; Jia, L.H.; Yu, Q.Y.; Zhang, Q.; Guo, A.Y. AnimalTFDB 3.0: A comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **2019**, *47*, D33–D38. [CrossRef]
11. Chiba, S.; Awazu, S.; Itoh, M.; Chin-Bow, S.T.; Satoh, N.; Satou, Y.; Hastings, K.E. A genomewide survey of developmentally relevant genes in Ciona intestinalis. IX. Genes for muscle structural proteins. *Dev. Genes Evol.* **2003**, *213*, 291–302. [CrossRef]

12. Wada, S.; Tokuoka, M.; Shoguchi, E.; Kobayashi, K.; Di Gregorio, A.; Spagnuolo, A.; Branno, M.; Kohara, Y.; Rokhsar, D.; Levine, M.; et al. A genomewide survey of developmentally relevant genes in Ciona intestinalis. *Dev. Genes Evol.* **2003**, *213*, 222–234. [CrossRef]

13. Yagi, K.; Satou, Y.; Mazet, F.; Shimeld, S.M.; Degnan, B.; Rokhsar, D.; Levine, M.; Kohara, Y.; Satoh, N. A genomewide survey of developmentally relevant genes in Ciona intestinalis. III. Genes for Fox, ETS, nuclear receptors and NFkappaB. *Dev. Genes Evol.* **2003**, *213*, 235–244. [CrossRef]

14. Yamada, L.; Kobayashi, K.; Degnan, B.; Satoh, N.; Satou, Y. A genomewide survey of developmentally relevant genes in Ciona intestinalis. IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. *Dev. Genes Evol.* **2003**, *213*, 245–253. [CrossRef]

15. Imai, K.S.; Hino, K.; Yagi, K.; Satoh, N.; Satou, Y. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: Towards a comprehensive understanding of gene networks. *Development* **2004**, *131*, 4047–4058. [CrossRef]

16. Miwata, K.; Chiba, T.; Horii, R.; Yamada, L.; Kubo, A.; Miyamura, D.; Satoh, N.; Satou, Y. Systematic analysis of embryonic expression profiles of zinc finger genes in Ciona intestinalis. *Dev. Biol.* **2006**, *292*, 546–554. [CrossRef]

17. Katikala, L.; Aihara, H.; Passamaneck, Y.J.; Gazdoiu, S.; Jose-Edwards, D.S.; Kugler, J.E.; Oda-Ishii, I.; Imai, J.H.; Nibu, Y.; Di Gregorio, A. Functional Brachyury binding sites establish a temporal read-out of gene expression in the Ciona notochord. *PLoS Biol.* **2013**, *11*, e1001697. [CrossRef]

18. Di Gregorio, A. The notochord gene regulatory network in chordate evolution: Conservation and divergence from Ciona to vertebrates. *Curr. Top. Dev. Biol.* **2020**, *139*, 325–374. [CrossRef]

19. Yasuo, H.; Satoh, N. An ascidian homolog of the mouse Brachyury (T) gene is expressed exclusively in notochord cells at the fate restricted stage. *Dev. Growth Differ.* **1994**, *36*, 9–18. [CrossRef]

20. Corbo, J.C.; Levine, M.; Zeller, R.W. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. *Development* **1997**, *124*, 589–602.

21. Jose-Edwards, D.S.; Kerner, P.; Kugler, J.E.; Deng, W.; Jiang, D.; Di Gregorio, A. The identification of transcription factors expressed in the notochord of Ciona intestinalis adds new potential players to the brachyury gene regulatory network. *Dev. Dyn.* **2011**, *240*, 1793–1805. [CrossRef] [PubMed]

22. Jose-Edwards, D.S.; Oda-Ishii, I.; Nibu, Y.; Di Gregorio, A. Tbx2/3 is an essential mediator within the Brachyury gene network during Ciona notochord development. *Development* **2013**, *140*, 2422–2433. [CrossRef] [PubMed]

23. Imai, K.S.; Satoh, N.; Satou, Y. An essential role of a FoxD gene in notochord induction in Ciona embryos. *Development* **2002**, *129*, 3441–3453. [PubMed]

24. Yagi, K.; Satou, Y.; Satoh, N. A zinc finger transcription factor, ZicL, is a direct activator of Brachyury in the notochord specification of Ciona intestinalis. *Development* **2004**, *131*, 1279–1288. [CrossRef]

25. Imai, K.S.; Satou, Y.; Satoh, N. Multiple functions of a Zic-like gene in the differentiation of notochord, central nervous system and muscle in Ciona savignyi embryos. *Development* **2002**, *129*, 2723–2732.

26. Lemons, D.; McGinnis, W. Genomic Evolution of Hox Gene Clusters. *Science* **2006**, *313*, 1918–1922. [CrossRef]

27. DeBiasse, M.B.; Colgan, W.N.; Harris, L.; Davidson, B.; Ryan, J.F. Inferring Tunicate Relationships and the Evolution of the Tunicate Hox Cluster with the Genome of Corella inflata. *Genome Biol. Evol.* **2020**, *12*, 948–964. [CrossRef]

28. Wei, J.; Zhang, J.; Lu, Q.; Ren, P.; Guo, X.; Wang, J.; Li, X.; Chang, Y.; Duan, S.; Wang, S.; et al. Genomic basis of environmental adaptation in the leathery sea squirt (*Styela clava*). *Mol. Ecol. Resour.* **2020**. [CrossRef]

29. Goldstien, S.J.; Schiel, D.R.; Gemmell, N.J. Regional connectivity and coastal expansion: Differentiating pre-border and post-border vectors for the invasive tunicate *Styela clava*. *Mol. Ecol.* **2010**, *19*, 874–885. [CrossRef]

30. Goldstien, S.J.; Dupont, L.; Viard, F.; Hallas, P.J.; Nishikawa, T.; Schiel, D.R.; Gemmell, N.J.; Bishop, J.D. Global phylogeography of the widely introduced North West Pacific ascidian *Styela clava*. *PLoS ONE* **2011**, *6*, e16755. [CrossRef]

31. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef]

32. Ikuta, T.; Yoshida, N.; Satoh, N.; Saiga, H. Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15118–15123. [CrossRef] [PubMed]

33. Seo, H.-C.; Edvardsen, R.B.; Maeland, A.D. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* **2004**, *431*, 67–71. [CrossRef]

34. Sekigami, Y.; Kobayashi, T.; Omi, A.; Nishitsuji, K.; Ikuta, T.; Fujiyama, A.; Satoh, N.; Saiga, H. Hox gene cluster of the ascidian, Halocynthia roretzi, reveals multiple ancient steps of cluster disintegration during ascidian evolution. *Zoological. Lett.* **2017**, *3*, 17. [CrossRef]

35. Blanchoud, S.; Rutherford, K.; Zondag, L.; Gemmell, N.J.; Wilson, M.J. De novo draft assembly of the Botrylloides leachii genome provides further insight into tunicate evolution. *Sci. Rep.* **2018**, *8*, 5518. [CrossRef]

36. Wu, S.; Zhang, Y.; Li, Y.; Wei, H.; Guo, Z.; Wang, S.; Zhang, L.; Bao, Z. Identification and expression profiles of Fox transcription factors in the Yesso scallop (Patinopecten yessoensis). *Gene* **2020**, *733*, 144387. [CrossRef]

37. Tu, Q.; Brown, C.T.; Davidson, E.H.; Oliveri, P. Sea urchin Forkhead gene family: Phylogeny and embryonic expression. *Dev. Biol.* **2006**, *300*, 49–62. [CrossRef]

38. Berna, L.; Alvarez-Valin, F. Evolutionary Genomics of Fast Evolving Tunicates. *Genome Biol. Evol.* **2014**, *6*, 1724–1738. [CrossRef]

39. Dehal, P.; Satou, Y.; Campbell, R.K.; Chapman, J.; Degnan, B.; De Tomaso, A.; Davidson, B.; Di Gregorio, A.; Gelpke, M.; Goodstein, D.M.; et al. The draft genome of Ciona intestinalis: Insights into chordate and vertebrate origins. *Science* **2002**, *298*, 2157–2167. [CrossRef]

40. Ikushima, H.; Negishi, H.; Taniguchi, T. The IRF family transcription factors at the interface of innate and adaptive immune responses. *Cold Spring Harb. Symp. Quant. Biol.* **2013**, *78*, 105–116. [CrossRef] [PubMed]

41. Zhang, Z.; Wei, J.; Ren, R.; Zhang, X. Anti-virus effects of interferon regulatory factors (IRFs) identified in ascidian Ciona savignyi. *Fish Shellfish. Immunol.* **2020**, *106*, 273–282. [CrossRef] [PubMed]

42. Sever, R.; Glass, C.K. Signaling by Nuclear Receptors. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a016709. [CrossRef] [PubMed]

43. Plateroti, M.; Kress, E.; Mori, J.I.; Samarut, J. Thyroid Hormone Receptor α1 Directly Controls Transcription of the β-Catenin Gene in Intestinal Epithelial Cells. *Mol. Cell. Biol.* **2006**, *26*, 3204–3214. [CrossRef]

44. Vincent Bertrand, C.H.; Caillol, D.; Popovici, C.; Lemaire, P. Neural Tissue in Ascidian Embryos Is Induced by FGF9/16/20, Acting via a Combination of Maternal GATA and Ets Transcription Factors. *Cell* **2003**, *115*, 615–627. [CrossRef]

45. Miya, T.; Nishida, H. An Ets transcription factor, HrEts, is target of FGF signaling and involved in induction of notochord, mesenchyme, and brain in ascidian embryos. *Dev. Biol.* **2003**, *261*, 25–38. [CrossRef]

46. Hotta, K. Brachyury-downstream gene sets in a chordate, Ciona intestinalis: Integrating notochord specification, morphogenesis and chordate evolution. *Evol. Dev.* **2008**, *10*, 37–51. [CrossRef]

47. Gainous, T.B.; Wagner, E.; Levine, M. Diverse ETS transcription factors mediate FGF signaling in the Ciona anterior neural plate. *Dev. Biol.* **2015**, *399*, 218–225. [CrossRef]

48. Schep, A.N.; Adryan, B. A comparative analysis of transcription factor expression during metazoan embryonic development. *PLoS ONE* **2013**, *8*, e66826. [CrossRef]

49. Nurieva, R.I.; Chung, Y.; Martinez, G.J.; Yang, X.O.; Tanaka, S.; Matskevitch, T.D.; Wang, Y.H.; Dong, C. Bcl6 mediates the development of T follicular helper cells. *Science* **2009**, *325*, 1001–1005. [CrossRef]

50. Wang, H.; Mo, P.; Ren, S.; Yan, C. Activating transcription factor 3 activates p53 by preventing E6-associated protein from binding to E6. *J. Biol. Chem.* **2010**, *285*, 13201–13210. [CrossRef]

51. Dou, L.; Liang, H.F.; Geller, D.A.; Chen, Y.F.; Chen, X.P. The regulation role of interferon regulatory factor-1 gene and clinical relevance. *Hum. Immunol.* **2014**, *75*, 1110–1114. [CrossRef]

52. Nam, S.; Lim, J.S. Essential role of interferon regulatory factor 4 (IRF4) in immune cell development. *Arch. Pharm. Res.* **2016**, *39*, 1548–1555. [CrossRef]

53. Davidson, B.; Swalla, B.J. A molecular analysis of ascidian metamorphosis reveals activation of an innate immune response. *Development* **2002**, *129*, 4739–4751.

54. Ross, A.C.; Zolfaghari, R. Cytochrome P450s in the regulation of cellular retinoic acid metabolism. *Annu. Rev. Nutr.* **2011**, *31*, 65–87. [CrossRef]

55. Nikolenko, J.V.; Krasnov, A.N. Nuclear receptors: Structure and mechanisms of action. *Russ. J. Genet.* **2007**, *43*, 234–240. [CrossRef]

56. Shao, C.; Bao, B.; Xie, Z.; Chen, X.; Li, B.; Jia, X.; Yao, Q.; Ortí, G.; Li, W.; Li, X.; et al. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* **2016**, *49*, 119–124. [CrossRef]

57. Izpisu'a-Belmonte, J.-C.; Falkenstein, H.; Dollé, P.; Duboule, A.R.D. Murine genes related to the Drosophila AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *EMBO J.* **1991**, *10*, 2279–2289. [CrossRef]

58. Pascual-Anaya, J.; Sato, I.; Sugahara, F.; Higuchi, S.; Paps, J.; Ren, Y.; Takagi, W.; Ruiz-Villalba, A.; Ota, K.G.; Wang, W.; et al. Hagfish and lamprey Hox genes reveal conservation of temporal colinearity in vertebrates. *Nat. Ecol. Evol.* **2018**, *2*, 859–866. [CrossRef]

59. Wang, S.; Zhang, J.; Jiao, W.; Li, J.; Xun, X.; Sun, Y.; Guo, X.; Huan, P.; Dong, B.; Zhang, L.; et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat. Ecol. Evol.* **2017**, *1*, 120. [CrossRef]

60. Ikuta, T.; Satoh, N.; Saiga, H. Limited functions of Hox genes in the larval development of the ascidian Ciona intestinalis. *Development* **2010**, *137*, 1505–1513. [CrossRef]

61. Weigel, D.; Jürgens, G.; Küttner, F.; Seifert, E.; Jäckle, H. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell* **1989**, *57*, 645–658. [CrossRef]

62. Carlsson, P.; Mahlapuu, M. Forkhead transcription factors: Key players in development and metabolism. *Dev. Biol.* **2002**, *250*, 1–23. [CrossRef] [PubMed]

63. Solomon, K.S.; Kudoh, T.; Dawid, I.B.; Fritz, A. Zebrafish foxi1 mediates otic placode formation and jaw development. *Development* **2003**, *130*, 929–940. [CrossRef] [PubMed]

64. Solomon, K.S.; Logsdon, J.M., Jr.; Fritz, A. Expression and phylogenetic analyses of three zebrafish FoxI class genes. *Dev. Dyn.* **2003**, *228*, 301–307. [CrossRef]

65. Zilinski, C.; Brownell, I.; Hashimoto, R.; Medina-Martinez, O.; Swindell, E.C.; Jamrich, M. Expression of FoxE4 and Rx visualizes the timing and dynamics of critical processes taking place during initial stages of vertebrate eye development. *Dev. Neurosci.* **2004**, *26*, 294–307. [CrossRef]

66. Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* **2010**, *26*, 680–682. [CrossRef]

67. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids. Res.* **2016**, *44*, D279–D285. [CrossRef]

68. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucleic Acids. Res.* **2018**, *46*, W200–W204. [CrossRef]

69. Deng, W.; Wang, Y.; Liu, Z.; Cheng, H.; Xue, Y. HemI: A toolkit for illustrating heatmaps. *PLoS ONE* **2014**, *9*, e111988. [CrossRef]

70. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]

71. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]

72. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [CrossRef]

73. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [CrossRef]

74. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [CrossRef]

75. Dardaillon, J.; Dauga, D.; Simion, P.; Faure, E.; Onuma, T.A.; DeBiasse, M.B.; Louis, A.; Nitta, K.R.; Naville, M.; Besnardeau, L.; et al. ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates. *Nucleic Acids Res.* **2020**, *48*, D668–D675. [CrossRef]