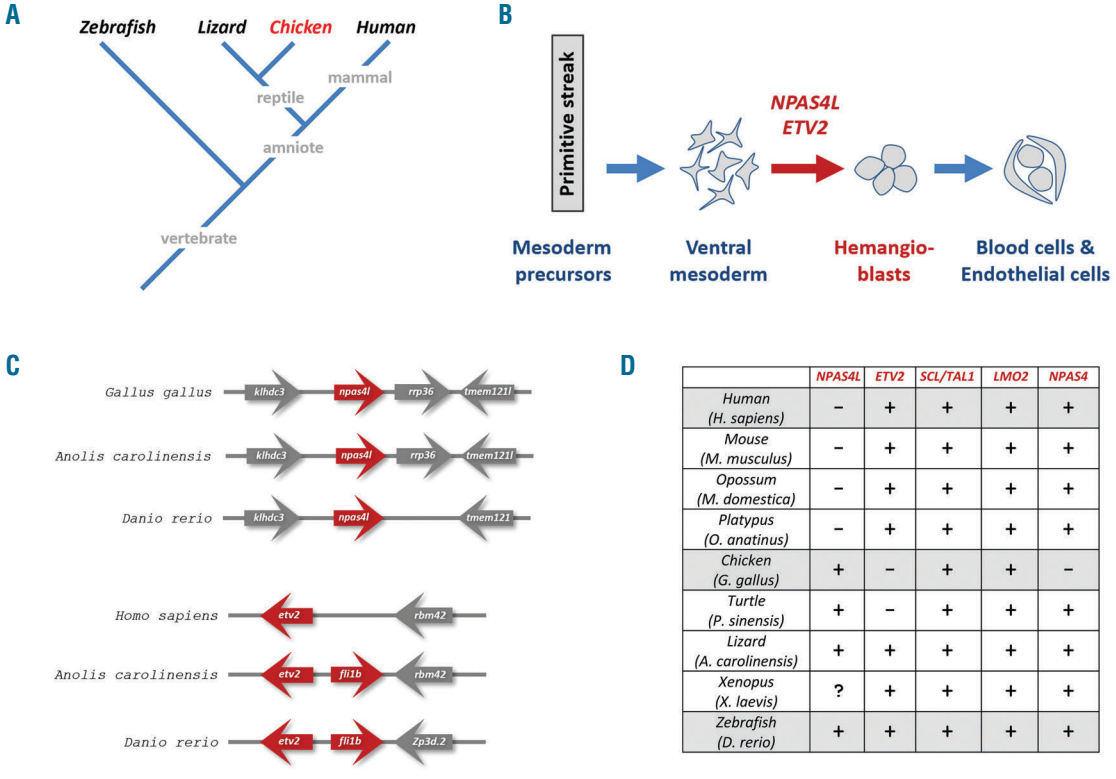
NPAS4L is involved in avian hemangioblast specification

Vertebrate primitive hematopoietic and vascular development is regulated by a conserved set of transcription factors. Their common precursors, the hemangioblasts, express Stem cell leukemia/T-cell acute lymphoblastic leukemia 1 (*SCL/TAL1*)[1] and Lim only protein 2 (*LMO2*)[2] in all vertebrate groups examined. Hemangioblast specification from nascent mesoderm was reported to be less conserved, with Ets variant 2 (*ETV2*) and Neuronal PAS-domain containing protein 4-like (*NPAS4L*) identified as its master regulator in mammals[3] and zebrafish,[4] respectively. We show here that the ortholog of *NPAS4L*, but not of *ETV2*, is present in the avian genome. Chicken *NPAS4L* is expressed in hemangioblasts prior to *SCL/TAL1* and *LMO2*. CRISPR-on mediated ectopic expression of endogenous *NPAS4L* leads to ectopic *SCL/TAL1* and *LMO2*, as with ectopic expression of zebrafish *NPAS4L*. We propose that the ancestral amniote genome had both *NPAS4L* and *ETV2* genes. The *ETV2* gene was lost in the avian lineage without affecting direct transcriptional regulation of *SCL/TAL1* and *LMO2* by *NPAS4L*.[5,6] The *NPAS4L* gene was lost in the mammalian lineage, with its roles partially replaced by *ETV2*.

Vertebrate primitive hematopoietic and vascular systems are derived from the mesoderm germ layer.[7,8] Lineage specification events taking place between gastrulation and the onset of circulation are controlled by a set of evolutionarily-conserved transcription regulators.[8,9] In birds,[10-12] as in fish, amphibians and mammals,[13-17] common progenitors of blood and endothelial cells (the hemangioblasts) start to express transcription factors *SCL/TAL1* and *LMO2* at Hamburger and Hamilton stage 4^{+} (HH4^{+}),[18] soon after their exit from posterior primitive streak where ventral mesoderm cells originate. This is followed by FGFR-mediated segregation of blood and endothelial lineages and functional differentiation of blood cells starting from HH7,[10] mediated by a conserved set of transcription factors including SCL/TAL1, LMO2, GATA-binding factor 2 (GATA2), LIM domain-binding protein 1 (LDB1) and transcription factor E2A (E2A).[19] After the onset of circulation from HH12/13, the hemangioblast markers *SCL/TAL1* and *LMO2* become restricted to the blood and endothelial lineages, respectively.



	NPAS4L	ETV2	SCL/TAL1	LMO2	NPAS4
Human (H. sapiens)	–	+	+	+	+
Mouse (M. musculus)	–	+	+	+	+
Opossum (M. domestica)	–	+	+	+	+
Platypus (O. anatinus)	–	+	+	+	+
Chicken (G. gallus)	+	–	+	+	–
Turtle (P. sinensis)	+	–	+	+	+
Lizard (A. carolinensis)	+	+	+	+	+
Xenopus (X. laevis)	?	+	+	+	+
Zebrafish (D. rerio)	+	+	+	+	+

Figure 1. The chicken genome has *NPAS4L*, but not *ETV2* ortholog. (A) A simplified vertebrate phylogenetic tree. (B) Schematic view of blood and endothelial cell differentiation from mesoderm precursors in the streak. *NPAS4L* and *ETV2* are proposed to function during hemangioblast specification in the ventral mesoderm. (C) The chicken genome has the *NPAS4L* orthologous gene flanked by *KLHDC3* gene on one side and *RRP36* and *TMEM121L* genes on the other. Similar syntenic organization is seen in lizard A. carolinensis and zebrafish D. rerio. These genes are missing in mammalian genomes. The chicken genome does not have *ETV2* ortholog. The lizard genome has *ETV2* and *FLI1B* as in the zebrafish genome. Mammals have *ETV2*, but not *FLI1B*. (It is to be noted that vertebrate genomes have three copies of such tandemly duplicated ETS family genes; not shown). In addition to the *ETV2-FLI1B* couplet which is the least conserved, the other two couplets (*ETS1-FLI1* and *ETS2-ERG*) are well-conserved. (D) Summary of presence and absence of *NPAS4L, ETV2, SCL/TAL1, LMO2* and *NPAS4* genes in various vertebrate groups. The following protein sequences were used for comparison. For SCL/TAL1: NP_001274276.1 (human), NP_001274317.1 (mouse), XP_001374963.3 (opossum), DNA clone XX-200B24 (platypus), NP_990683.1 (chicken), XP_030427307.1 (desert turtle; sequence in Chinese soft-shell turtle is incomplete), XP_008114556.1 (lizard), NP_001081746.1 (Xenopus) and NP_998402.1 (zebrafish); for LMO2: AAH42426.1 (human), AAH57880.1 (mouse), XP_027693653.1 (opossum), XP_028917173.1 (platypus), AAL78036.1 (chicken), XP_030415938.1 (turtle), XP_003225211.1 (lizard), NP_001081112.1 (Xenopus) and AAH93136.1 (zebrafish); for ETV2: NP_055024.2 (human), NP_031985.2 (mouse), XP_007491908.1 (opossum), XP_028921116.1 (platypus), XP_008119144.1 (lizard), NP_001089600.1 (Xenopus) and NP_001032452.1 (zebrafish); for NPAS4L: EntrezID 101750093 (chicken), XP_008103134.1 (lizard), XP_008165306.1 (turtle) and NP_001316841.1 (zebrafish).

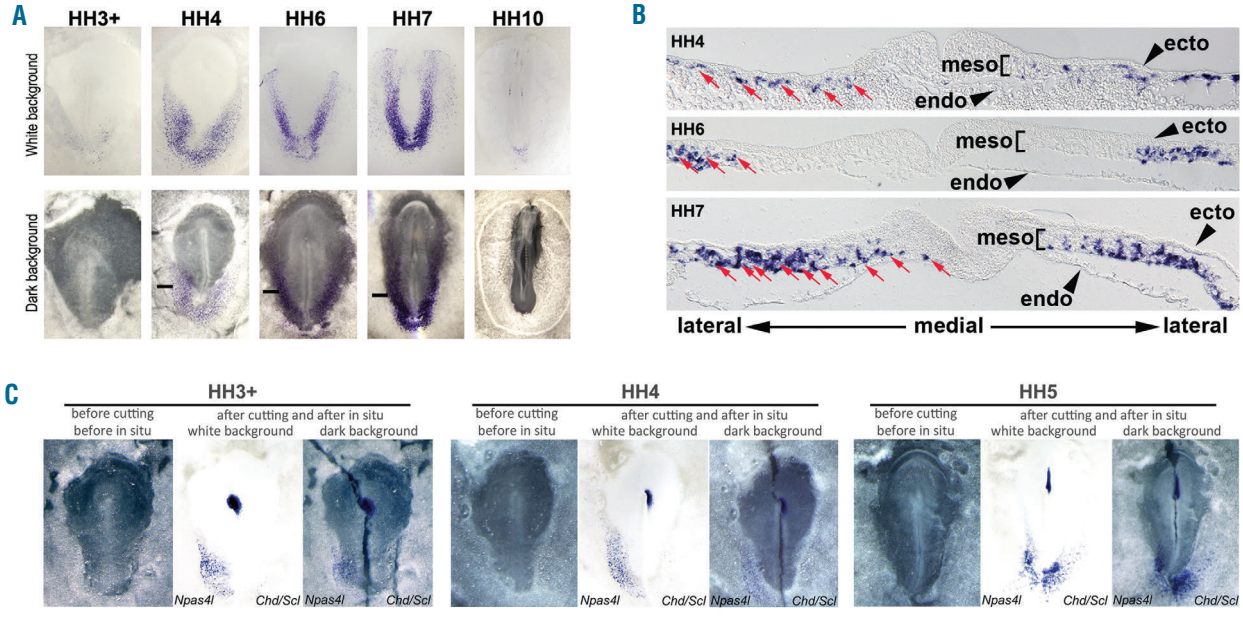


Figure 2. Chicken *NPAS4L* gene expression during hemangioblast specification. (A) Whole-mount *in situ* hybridization (WISH) of *NPAS4L* from HH3 to HH10. Fertilized hen's eggs were purchased from Takamoriryo in Aso (Kumamoto, Japan). (Top) White background for expression visualization; (bottom) dark background for stage visualization. Black lines indicate section levels shown in (B). (B) Section of embryos in (A). *NPAS4L*-expressing cells indicated by red arrows. Germ layers marked by black arrowheads (ectoderm and endoderm) and brackets (mesoderm). (C) Chicken *NPAS4L* is expressed starting from HH3$^+$, earlier than *SCL/TAL1*. Embryos were fixed and processed to the pre-hybridization step (left panels) and were cut into left (stained for NPAS4L) and right (stained for *SCL/TAL1* and *Chordin* together) halves. Stained half embryos were then photographed together (middle panels: white background showing both halves; right panels: dark background showing both halves). *Chordin* expression was used to mark precise embryo stages. At HH3$^+$ (top row) and HH4 (middle row), NPAS4L is expressed and SCL/TAL1 is not expressed. At HH5 (bottom row), both *NPAS4L* and *SCL/TAL1* are expressed.

Hemangioblast specification from their mesoderm precursors was reported to involve divergent transcriptional regulation, with *ETV2* in mammals[3,20] and *NPAS4L* in zebrafish[4] as the main driver. *ETV2* ortholog is present in the zebrafish genome, but its function was reported to be under the control of *NPAS4L*.[4,6] No *NPAS4L* ortholog has been identified in any mammalian species, suggesting that this gene is not involved in hemangioblast specification in mammals. Mammalian *NPAS4*, a homolog of *NPAS4L*, was able to rescue fish *cloche* (*npas4l*) mutant phenotypes.[4] Duplication of the *NPAS4* and *NPAS4L* genes, however, took place before the divergence of Actinopterygians (ray-finned fish, including the teleosts) and Sarcopterygians (lobe-finned fish, including the tetrapods) and *NPAS4* has not been associated so far with any aspect of vertebrate hematopoietic development, suggesting that these two genes have different biological functions involving separate molecular regulatory networks.

Since the mammals and birds are closely related both phylogenetically (Figure 1A) and ontogenetically (Figure 1B), we investigated whether avian *NPAS4* and *ETV2* genes are involved in early hematopoietic and vascular development. Molecular phylogenetic analysis indicated that an *NPAS4L* ortholog was present in the chicken (*G. gallus*) genome (in both galGal5 and galGal6 assemblies) (Figure 1C). Although this gene is annotated as *NPAS4* in the current assembly, syntenic analysis (Figure 1C) clearly indicated that it was the ortholog of *NPAS4L* in fish and other vertebrate groups (viewable through search term "npas4" in the NCBI genome data browser *https://www.ncbi.nlm.nih.gov/genome/gdv/?org=gallus-gallus* or the chicken FANTOM dataset browser *http://fantom.gsc.riken.jp/zenbu/gLyphs/#config=b1zZI1gUFZ6mHX6-4Gvxr*). Phylogenetic analyses also showed that the *NPAS4L* gene is present in all other bird species with their genomes fully or partially assembled and in non-avian reptiles with their genomes assembled (*Anolis* lizard shown as an example in Figure 1C). In contrast, the *ETV2* ortholog is missing in the entire avian lineage, and also in crocodiles and turtles, suggesting a loss of this gene before avian evolution. The *ETV2* ortholog, however, was found in some of the reptilian lineages (e.g., lizards and snakes) (Figure 1C and D). Taken together, our phylogenetic analyses suggest that birds have the *NPAS4L*, but not the *ETV2*, gene in their genomes.

We next asked whether *NPAS4L* plays a role in early hemangioblast specification in chick as was shown in zebrafish. For this purpose, we generated an RNA whole-mount *in situ* hybridization (WISH) probe for chicken *NPAS4L* and performed WISH using embryos from stage HH3 (early gastrulation) to stage HH12 (onset of circulation). Expression of chicken *NPAS4L* was detected in territories marking nascent hemangioblasts in ventral mesoderm (Figure 2A) from stage HH3$^+$, the earliest among all hemangioblast-specific genes (e.g., *SCL/TAL1* and *LMO2* expression starts from stage HH4$^+$). This observation was confirmed by WISH using left-right bisected embryos, with the left half stained for *NPAS4L* and the right half stained for *SCL/TAL1* and *Chordin* (Figure 2C). Paraffin-sectioning of stained embryos (Figure 2B) showed that *NPAS4L*-positive cells are located in a subset of the mesoderm germ layer that will give rise to blood and endothelial cell lineages (red arrows; germ layers marked by arrowheads and brackets), as we had previously reported.[10,21] *NPAS4L* expression levels peaked at HH7 and declined soon afterwards (Figure 2A), suggesting that this gene is specifically and transiently involved in hemangioblast formation, but not in their differentiation.

We have previously generated the chicken promoterome database, spanning the entire 21-day period of embryonic development.[22] When we searched this data-

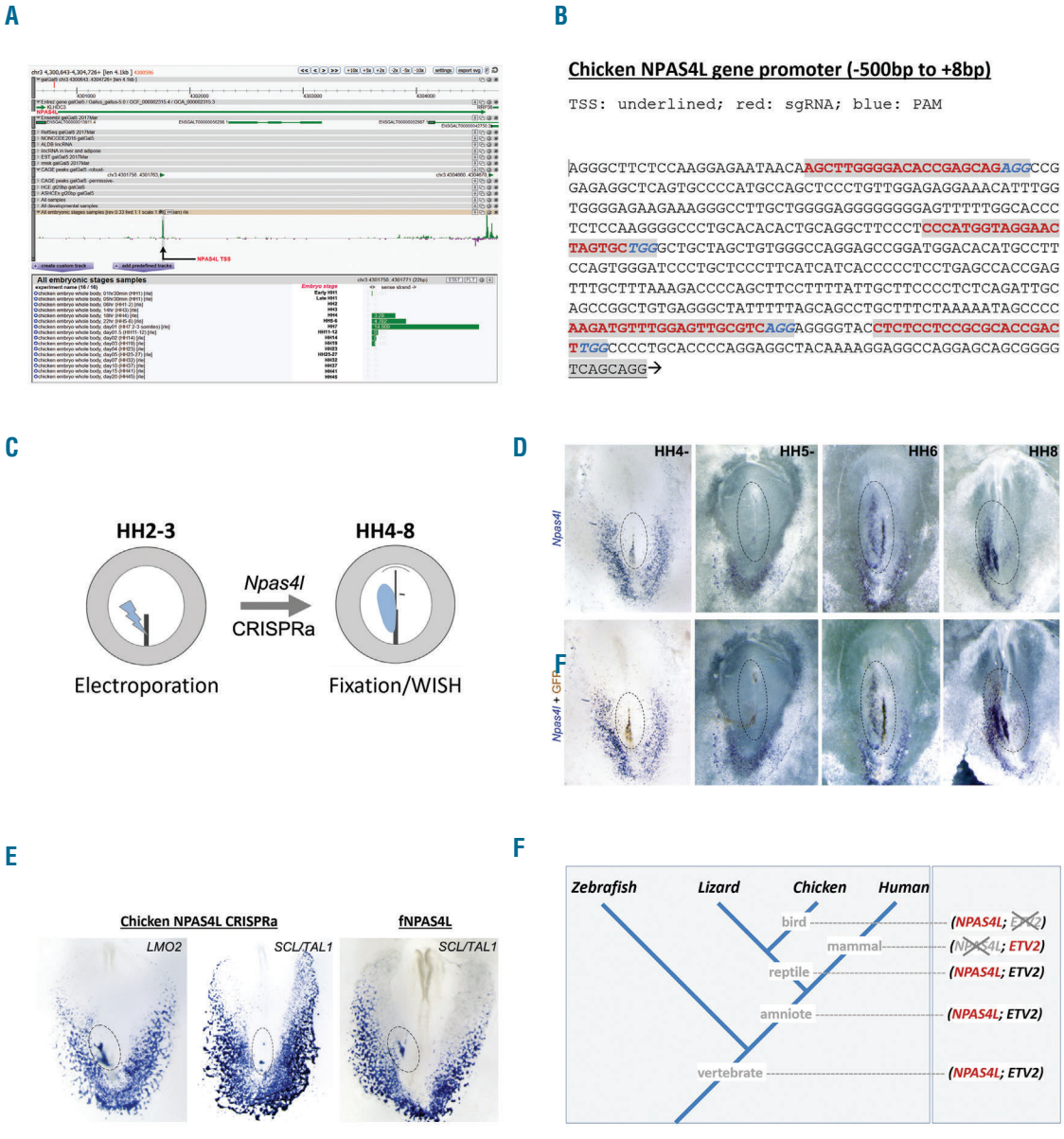


Figure 3. NPAS4L is involved in chicken hemangioblast specification. (A) Screenshot of *NPAS4L* locus in chicken promoterome database (see text for web link). *NPAS4L* transcription start site (TSS) is indicated by red label and black arrow. TSS activity levels at different developmental stages are shown at the bottom. The highest expression is seen at HH7. (B) Design of sgRNA for chicken *NPAS4L* CRISPRa. Mapped TSS is TCAGCAGG (underlined). Preceding 500 bp of promoter region is shown, with sgRNA sequences highlighted in red and PAM sequences in blue. (C) Schematic diagram of how embryos are electroporated and cultured. (D) *NPAS4L* CRISPRa constructs activate endogenous *NPAS4L* expression ectopically (oval). (Top) *NPAS4L* expression only; (bottom) *NPAS4L* expression together with anti-GFP staining marking the electroporated territories (brown). (E) *NPAS4L* CRISPRa constructs activate endogenous *LMO2* (left) and *SCL/TAL1* (middle) ectopically (oval). Zebrafish *NPAS4L* is also capable of activating hemangioblast markers (SCL/TAL1 shown in right panel, oval). (F) Hypothetic scenario of hemangioblast specification in the ancestral amniote and ancestral reptile. It is proposed that the zebrafish scenario (both *NPAS4L* and *ETV2* genes are present, with *NPAS4L* functioning upstream of *ETV2*) is the default one. Mammals have lost *NPAS4L* and birds have lost *ETV2*.

base (*http://fantom.gsc.riken.jp/zenbu/*), *NPAS4L* was shown (Figure 3A) to be only expressed in a narrow time window with its peak expression at HH7, consistent with the WISH data. To evaluate its molecular function, we used CRISPRa (CRISPR-mediated gene activation; also known as CRISPR-on)[23] to ectopically express this gene. CRISPRa utilizes a modified Cas9 protein (with dead nuclease activity and fused with ten copies of VP16 transactivation domain) to recruit transcriptional machinery to targeted promoters mediated by single guide RNA (sgRNA). We had previously confirmed the effectiveness of CRISPRa system in the avian model by taking advantage of the single-nucleotide level resolution in transcription start site (TSS) mapping.[22] Four sgRNA sequences located within the 500-base pair region preceding the *NPAS4L* TSS were selected (Figure 3) (for interactive view of *NPAS4L* TSS, use the link *http://fantom.gsc.riken.jp/zenbu/gLyphs/#config=b1zZI1gUFZ6mHX6-4Gvxr;loc=galGal5::chr3:4300587..4304021+*) and cloned into expression construct pAC154-dual-dCas9VP160-sgExpression (Addgene #48240). Mesoderm precursors in the streak in HH2/3 embryos were targeted for electroporation (see Weng and Sheng[19] for electroporation protocol) with these four sgRNA expression constructs together with marker GFP expression construct (Figure 3C), and

electroporated embryos were assessed for ectopic expression of endogenous *NPAS4L* and of two hemangioblast markers *SCL/TAL1* and *LMO2*. *NPAS4L* CRISPRa constructs were able to ectopically activate endogenous *NPAS4L* (Figure 3D, oval areas) (11 of 12; 92%) in regions that are normally *NPAS4L*-negative (Figure 2A), as well as hemangioblast markers *SCL/TAL1* (9 of 25; 36%) and *LMO2* (6 of 21; 29%) (Figure 3E, oval areas in left two panels), albeit with reduced efficiency. Interestingly, similar inductive effect (5 of 13 for *LMO2* and 4 of 9 for *SCL/TAL1*) was observed when we used zebrafish *NPAS4L* expression construct[4] (cloned into the pCAGGS expression vector) (Figure 3E, oval area in right panel), supporting partial molecular conservation between the zebrafish and chicken *NPAS4L* genes.

In conclusion, we present evidence that during early chicken development, *NPAS4L*, instead of *ETV2*, is involved in hemangioblast formation. Data from our molecular phylogenetic analyses support the hypothesis that both the *NPAS4L* and *ETV2* genes were present in the common reptilian ancestor and likely also in the common amniote ancestor (Figure 3F). A conclusive confirmation of their epistatic relationship, however, requires additional evidence from gain-of-function of *ETV2* (e.g., using a reptilian *ETV2* ortholog) and loss-of-function of *NPAS4L* (e.g., through CRISPR-mediated transcription inhibition) studies. In birds and other reptilian lineages which lack the *ETV2* ortholog in their genome, it is possible that other ETS family genes have been co-opted to play hemangioblast-specific roles of *ETV2*. Because ETV2 and NPAS4L are transcription factors with different DNA binding specificities and co-factor requirements, it remains to be shown how *ETV2* took over molecular functions of *NPAS4L* during early mammalian evolution.

Wei Weng, Hiroki Nagai, Sofiane Hamidi and Guojun Sheng

International Research Center for Medical Sciences (IRCMS), Kumamoto University, Kumamoto, Japan

Correspondence:
GUOJUN SHENG -sheng@kumamoto-u.ac.jp

doi:10.3324/haematol.2019.239434

References

1. Porcher C, Chagraoui H, Kristiansen MS. SCL/TAL1: a multifaceted regulator from blood development to disease. Blood. 2017; 129(15):2051-2060.
2. Morishima T, Krahl AC, Nasri M, et al. LMO2 activation by deacetylation is indispensable for hematopoiesis and T-ALL leukemogenesis. Blood. 2019;134(14):1159-1175.
3. Kataoka H, Hayashi M, Nakagawa R, et al. Etv2/ER71 induces vascular mesoderm from Flk1+PDGFRalpha+ primitive mesoderm. Blood. 2011;118(26):6975-6986.
4. Reischauer S, Stone OA, Villasenor A, et al. Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification. Nature. 2016;535(7611):294-298.
5. Liao EC, Paw BH, Oates AC, Pratt SJ, Postlethwait JH, Zon LI. SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish. Genes Dev. 1998;12(5):621-626.
6. Marass M, Beisaw A, Gerri C, et al. Genome-wide strategies reveal target genes of Npas4l associated with vascular development in zebrafish. Development. 2019;146(11).
7. Nagai H, Shin M, Weng W, et al. Early hematopoietic and vascular development in the chick. Int J Dev Biol. 2018;62(1-2-3):137-144.
8. Zon LI. Developmental biology of hematopoiesis. Blood. 1995; 86(8):2876-2891.
9. Shivdasani RA, Orkin SH. The transcriptional control of hematopoiesis. Blood. 1996;87(10):4025-4039.
10. Nakazawa F, Nagai H, Shin M, Sheng G. Negative regulation of primitive hematopoiesis by the FGF signaling pathway. Blood. 2006; 108(10):3335-3343.
11. Shin M, Nagai H, Sheng G. Notch mediates Wnt and BMP signals in the early separation of smooth muscle progenitors and blood/endothelial common progenitors. Development. 2009; 136(4):595-603.
12. Minko K, Bollerot K, Drevon C, Hallais MF, Jaffredo T. From mesoderm to blood islands: patterns of key molecules during yolk sac erythropoiesis. Gene Expr Patterns. 2003;3(3):261-272.
13. Davidson AJ, Zon LI. The 'definitive' (and 'primitive') guide to zebrafish hematopoiesis. Oncogene. 2004;23(43):7233-7246.
14. Walmsley M, Cleaver D, Patient R. Fibroblast growth factor controls the timing of Scl, Lmo2, and Runx1 expression during embryonic blood development. Blood. 2008;111(3):1157-1166.
15. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. Cell. 2008;132(4):631-644.
16. Manaia A, Lemarchandel V, Klaine M, et al. Lmo2 and GATA-3 associated expression in intraembryonic hemogenic sites. Development. 2000;127(3):643-653.
17. Drake CJ, Fleming PA. Vasculogenesis in the day 6.5 to 9.5 mouse embryo. Blood. 2000;95(5):1671-1679.
18. Hamburger V, Hamilton HL. A series of normal stages in the development of the chick embryo. J Morphol. 1951;88(1):49-92.
19. Weng W, Sheng G. Five transcription factors and FGF pathway inhibition efficiently induce erythroid differentiation in the epiblast. Stem Cell Reports. 2014;2(3):262-270.
20. Liu F, Li D, Yu YY, et al. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. EMBO Rep. 2015;16(5):654-669.
21. Weng W, Sukowati EW, Sheng G. On hemangioblasts in chicken. PLoS One. 2007;2(11):e1228.
22. Lizio M, Deviatiiarov R, Nagai H, et al. Systematic analysis of transcription start sites in avian development. PLoS Biol. 2017;15(9):e2002887.
23. Cheng AW, Wang H, Yang H, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013;23(10):1163-1171.