Research Paper

Identification of Pluripotency Genes in the Fish Medaka

Danke Wang*, **Dwarakanath Manali, Tiansu Wang, Narayani Bhat, Ni Hong, Zhendong Li, Li Wang, Yan Yan, Rong Liu, Yunhan Hong** ✉

Department of Biological Sciences, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

* This work contains part of master thesis work of this author.

✉ Corresponding author: Prof. Yunhan Hong, Department of Biological Sciences, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260. Fax: +65 6779 2486; Tel: +65 6516 2915; Email: dbshyh@nus.edu.sg

Abstract

Stem cell cultures can be derived directly from early developing embryos and indirectly from differentiated cells by forced expression of pluripotency transcription factors. Pluripotency genes are routinely used to characterize mammalian stem cell cultures at the molecular level. However, such genes have remained unknown in lower vertebrates. In this regard, the laboratory fish medaka is uniquely suited because it has embryonic stem (ES) cells and genome sequence data. We identified seven medaka pluripotency genes by homology search and expression *in vivo* and *in vitro*. By RT-PCR analysis, the seven genes fall into three groups of expression pattern. Group I includes *nanog* and *oct4* showing gonad-specific expression; Group II contains *sall4* and *zfp281* displaying gonad-preferential expression; Group III has *klf4*, *ronin* and *tcf3* exhibiting expression also in several somatic tissues apart from the gonads. The transcripts of the seven genes are maternally supplied and persist at a high level during early embryogenesis. We made use of early embryos and adult gonads to examine expression in stem cells and differentiated derivatives by in situ hybridization. Strikingly, *nanog* and *oct4* are highly expressed in pluripotent blastomeres of 16-cell embryos. In the adult testis, *nanog* expression was specific to spermatogonia, the germ stem cells, whereas *tcf3* expression occurred in spermatogonia and differentiated cells. Most importantly, all the seven genes are pluripotency markers *in vitro*, because they have high expression in undifferentiated ES cells but dramatic down-regulation upon differentiation. Therefore, these genes have conserved their pluripotency-specific expression *in vitro* from mammals to lower vertebrates.

Key words: medaka, nanog, oct4, pluripotency, stem cells, tcf3

Introduction

Pluripotent stem cell cultures can be derived from developing embryos and adult tissues. In particular, embryonic stem (ES) cells from early developing embryos are an excellent model for analyzing vertebrate development *in vitro*, and provide a versatile resource for induced differentiation into many desired cell types in large quantity for regenerative medicine. In mouse, ES cells have widely been used also in gene targeting for generating knockout animals to study genes' functions and to model human diseases[1]. Since the production of first mouse ES cell lines in 1981[2,3], numerous attempts have been made towards ES cell derivation in different vertebrate organisms including important livestock species [4,5]. In 1996, our success in the production of ES cells from the fish medaka[6] demonstrated that the ability for ES cell derivation is present from fish to mammals. Indeed, human ES cells were successfully established in 1998[7]. At present, induced pluripotent stem (iPS) cells can be generated from differentiated cells by

introduced expression of reprogramming transcription factors[8].

Although interest in ES derivation has steadily been increasing and putative ES-like cells have been reported in several species[5,9], including short-term ES cell cultures capable of germline transmission in zebrafish[10], stable lines of *bona fide* ES cells have been limited to few organisms. One of the challenges is the lack of suitable molecular markers to monitor and regulate the pluripotency of putative ES cell cultures in organisms in which established ES cells are absent. Recent studies have established that mouse and human ES cells share a core transcriptional network consisting of Oct4, Nanog and Sox2 [11,12]. The mouse *oct4* gene (also known as *pou5f1*) encodes a POU domain-containing and octamer-binding protein and represents the prototype of pluripotency genes in mammals, because it is maternally supplied and expressed specifically throughout the totipotent cycle, including the inner cell mass (ICM), epiblast and primordial germ cells (PGCs) of early developing embryo and spermatogonia and oocytes in the adulthood [13]. These embryonic and adult cells – except oocytes – are all capable of producing stem cell cultures, in which quantitative *oct4* expression defines differentiation, dedifferentiation or self-renewal[14]. A similar expression pattern has been reported also for the mouse *sox2* encoding a HMG-containing transcription factor[15,16]. However, the human *Sox2* is absent in germ cells[17]. The gene *nanog* encodes a divergent protein that contains a homeobox and is not maternally deposited in mice. Instead its expression commences at the morula stage and later on occurs in the ICM, epiblast, PGCs and ES cell cultures, albeit its expression is absent in spermatogonia of the adult testis[18,19]. Accumulated data from mammals suggest a notion that although these three genes show a common pluripotency-specific expression in ES cell cultures, their expression in different species may exhibit obvious variations *in vivo*. In this paper, pluripotency genes refer to those whose expression is high in undifferentiated ES cells but dramatically reduced or even lost upon ES cell differentiation.

Additional pluripotency genes have recently been described in mammals. These include *klf4* encoding a Krüppel-like factor[8], *ronin* encoding the THAP domain containing 11 protein (Thap11) [20], *sall4* encoding the Sal-like protein 4 (Sall4)[21,22], *tcf3* (also called *tcf7l1*) encoding the T-cell transcription factor 3 [23,24] and *zfp281* encoding the zinc-finger protein 281[25]. These genes have been identified in mammalian ES cells. Their homologs/paralogs remain to be identified and characterized in lower vertebrates such as fish.

The medaka (*Oryzias latipes*) is well-suited for analyzing vertebrate development [26]. We use this laboratory fish as a lower vertebrate model for stem cell biology[5,9]. In this organism, we have previously established a feeder-free culture system that allowed for derivation of diploid ES cells[6,27] from fertilization midblastula embryos as the equivalent of the mammalian blastocysts, male germ stem cells from the adult testis [28] and most recently, even haploid ES cells from gynogenetic embryos [29,30]. Most recently, we have demonstrated in medaka that early embryos even up to the 32-cell stage are capable of cell culture initiation[31]. In addition, the medaka genome has been fully sequenced and partially annotated (http://www.ensembl.org/index.html). The availability of sequence data and well-characterized stem cell lines makes medaka a unique model organism to identify pluripotency genes *in vivo* and *in vitro*. Here we have identified seven medaka pluripotency genes and established their suitability as pluripotency markers by analyzing RNA expression patterns in adult tissues, developing embryos and ES cell culture. Our results demonstrate that these homologs/paralogs are molecular markers of stem cell cultures in fish and perhaps also other lower vertebrates.

Results

Gene identification

We adopted a homology approach to identify medaka pluripotency genes by taking the advantageous availability of genome sequence data. Through BLAST searches against public database, we have previously identified homologs/paralogs of several mammalian pluripotency genes including *nanog, oct4, klf4, sall4, zfp281, ronin* and *tcf3 [29]*. In literature and several annotated genomes, *tcf3* (also *tcf7l1*) is also confused with *e2a* (also called *e12/e47*) encoding transcription factor E2A (E2A immunoglobulin enhancer binding factor). The improvement in the genome sequence and the availability of more cDNA sequences provoked us to re-visit gene identification in medaka. Extensive reciprocal BLAST searches and sequence alignments revealed that the two previous *tcf3* genes actually encode E2A proteins instead (Fig. 1).

We found a single *tcf3* gene in the medaka genome, compared to two *tcf3* genes, namely *tcf3a/headless* and *tcf3b*, in zebrafish[32,33]. To see whether the medaka *tcf3* is homologous or paralogous to the mammalian gene, we compared its protein sequence and its chromosome location. On sequence alignment, the medaka Tcf3 protein shows a high overall conservation and has a beta-catenin binding

domain and a class-I HMG box (Fig. 2A). By pairwise comparison, the medaka Tcf3 is 65.7% and 67.5% identical to the mouse and human Tcf3 protein, respectively (Fig. 2A). On a phylogenetic tree, the medaka Tcf3 clusters with known zebrafish Tcf3 proteins, while the mouse and human Tcf3 proteins forms a second cluster, conforming to the organism phylogeny (Fig. 2B). More importantly, medaka *tcf3* exhibits a syntenic relationship to its human counterpart(Fig. 2C). These data suggest that the medaka *tcf3* identified in this study is a homolog of the mammalian *tcf3/tcf7l1* gene.
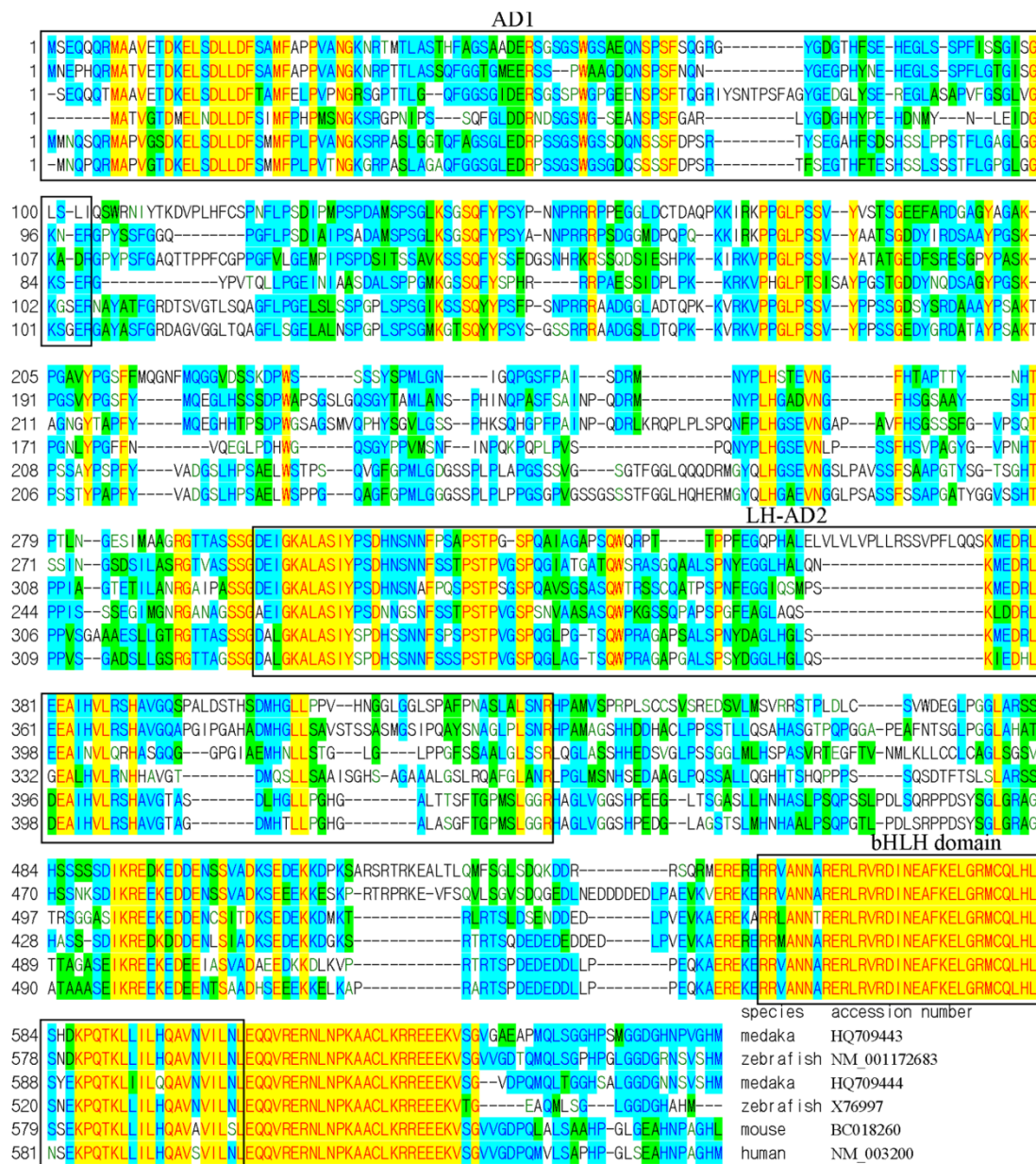
Fig. 1. Sequence comparison of E2A (E12/E47) proteins on alignment. Species names and gene accession numbers are given at the end of alignment. The two medaka e2a genes, e2a1 (HQ709443) and e2a2 (HQ709444), have been submitted to GenBank. The activation domain 1 (AD1), the loop-helix activation domain 2 (LH-AD2) and the basic helix–loop–helix (bHLH) are highlighted in frame, respectively.

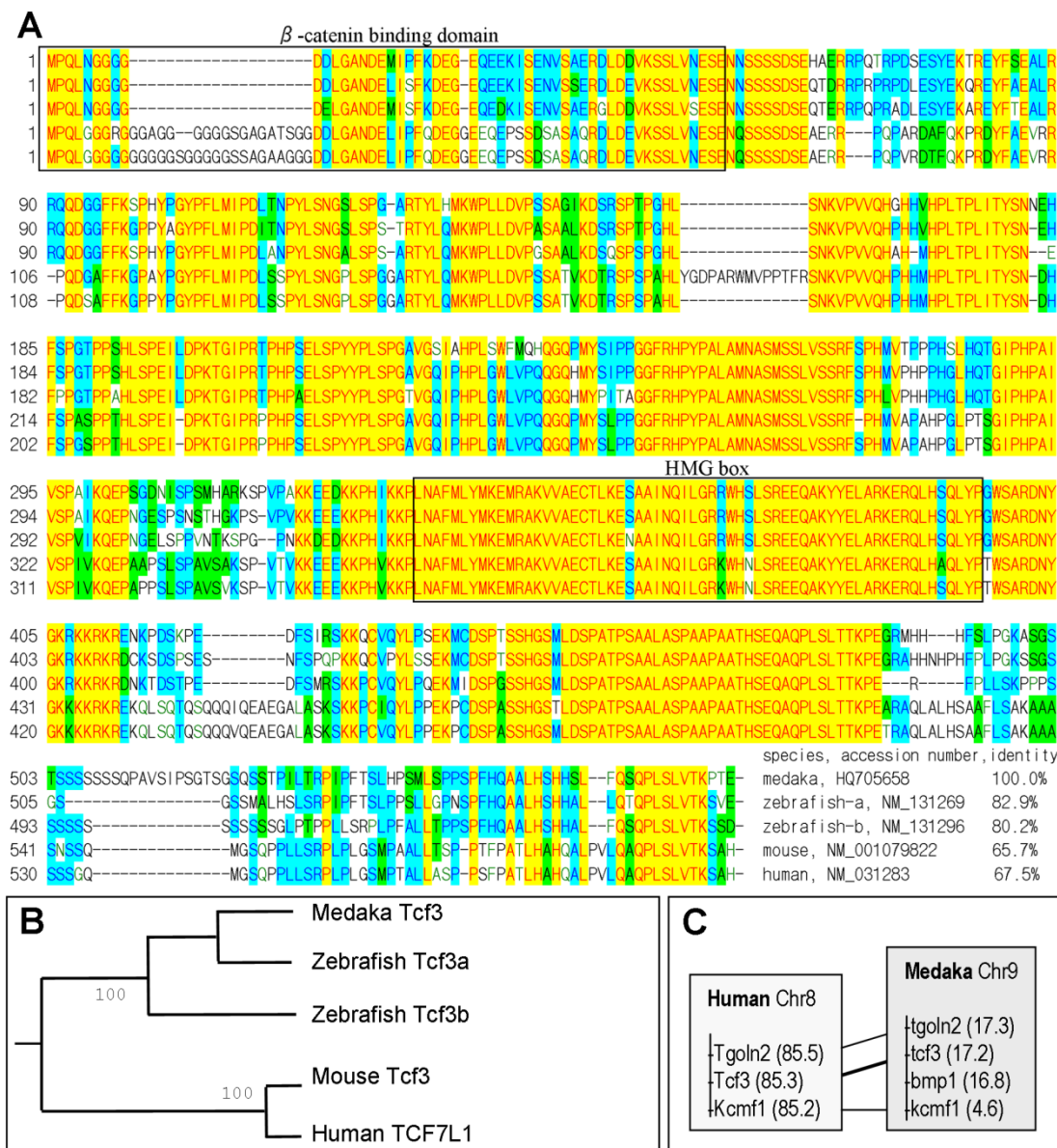| species	accession number
medaka	HQ709443
zebrafish	NM_001172683
medaka	HQ709444
zebrafish	X76997
mouse	BC018260
human	NM_003200

Fig. 2. Phylogenetic comparison of Tcf3/Tcf7ll proteins. (A) Sequence alignment. At the end of the alignment are species, gene accession numbers and amino acid identity values. The β-catenin binding domain and high mobility group (HMG) box are highlighted in frame. (B) Phylogenetic tree of Tcf3/TCF7ll proteins. (C) Chromosome location of the *tcf3/tcf7ll* gene in human and medaka. Chromosomal positions are in parenthesis.

RT-PCR analysis of expression in tissues and embryos

The addition of this newly recognized *tcf3* led to a set of seven medaka genes. By using primers listed in Table 1, we performed RT-PCR analyses to survey their expression patterns in eight representative tissues and seven critical stages of embryos. According to expression patterns in adult tissues, the seven genes fall into three groups (Fig. 3A): Group I contains *nanog* and *oct4*, which show gonad-specific expression; Group II comprises *sall4* and *zfp281*, which display high expression in the gonads and detectable expression in one or two somatic tissues; Group III consists of *klf4, ronin* and *tcf3*, which exhibit wide expression at variable levels in 6 to 7 tissues. The expression patterns of the seven genes are compared in more detail in Table 2.
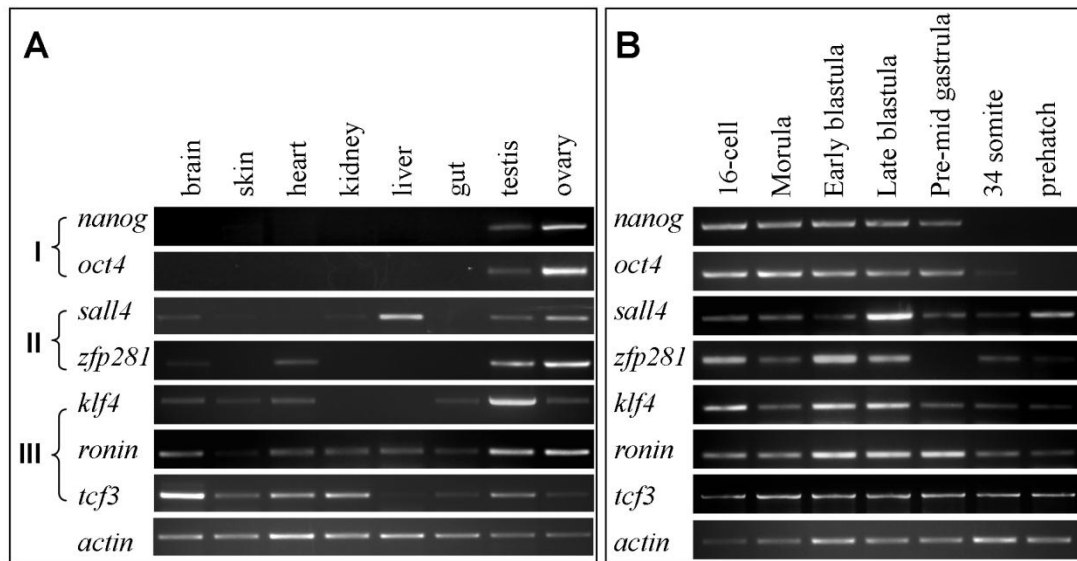
Fig. 3. RT-PCR analyses of RNA expression in adult tissues and embryos. (A) Adult tissues. (B) Embryos.

Table 1 Genes and Primers used in RT-PCR

Gene		Primer sequence		size (bp)
Name	Accession	Forward primer	Reverse primer	
nanog	FJ436046	CTCCACATGTCCCCCCTTATC	AGGATAGAATAGTCACATCAC	591
oct4	AY639946	GCTTTCTTTGGCGTAAACTCGTC	TCATCCTGTCAGGTGACCTACC	777
sall4	ENSORLG00000016130	ATGTCGAGGCGCAAACAAG	AGCCACTTTAGCGTCAGGTATG	501
zfp281a	ENSORLG00000002799	ATGAGTATTATCCAAGACAAGATA	TGTGTCCTTTTGTGTCGCTCC	854
klf4	ENSORLG00000005643	CATCCTCTCACCCAGATGC	TCATAAGTGCCTCTTCATGTGG	447
ronin	ENSORLG00000008903	AACTGAGAAGCGACGAGTACTC	CATTTTCTTTCTGAAACCAAC	302
tcf3	HQ705658	ATGCCTCAACTGAACGGAGG	CTGCAGAGCTGGGAACATCC	433

Table 2 Summary of RNA expression in medaka adult tissues

gene	brain	skin	heart	kidney	liver	gut	testis	ovary
nanog	-	-	-	-	-	-	++	+++
oct4	-	-	-	-	-	-	++	+++
sall4	+	+	-	+	+++		+	++
zfp281a	+	-	++	-	-	-	++	+++
klf4	+	+	+			+	++++	+
ronin	++	+	++	++	++	+	+++	+++
tcf3	+++	++	+++	+++	+	++	+++	++

Relative levels of expression are indicated by -, +, ++ and +++ for barely detectable, faintly detectable, easily detectable and high expression

The transcripts of the seven genes were easily detected in developing embryos ranging from cleavages to the late blastula stage (Fig. 3B), suggesting that their RNAs are maternally supplied, since zygotic transcription in medaka does not occur until the mid-blastula transition [34]. When embryonic development proceeded through gastrulation to the pre-hatching stage, differences in RNA level became ap-

parent among the genes: *nanog* and *oct4* sharply declined and ultimately became undetectable, *sall4*, *zfp281* and *klf4* exhibited a dramatic decrease, whereas *ronin* and *tcf3* displayed little change. The midblastula stage in lower vertebrates such as fish is a critical stage for pluripotency. At this stage, the embryo is thought to consist of developmentally indeterminate cells allowing for ES cell derivation[6,29]. After the blastula stage, cell lineage restriction and differentiation occurs during and after gastrulation. The temporal expression patterns suggest that *nanog* and *oct4* appear to be expressed mainly in pluripotent cells of developing embryos, while the remainder are expressed and required also for other events of embryonic development.

Expression in embryos and the adult gonad

In the mouse, the totipotent cycle begins with the zygote, proceeds through a series of transient structures (morula, inner cell mass and epiblast) and ends with the germline, which is established as PGCs and ultimately produces eggs and sperm in the adult gonads, the ovary in the female and the testis in the male[35]. All the transient embryonic structures and PGCs have the potential to produce pluripotent stem cell cultures. Even adult spermatogonia, the male germ stem cells, are also capable of generating stem cell cultures in fish[28] and mouse[36]. Therefore, we made use of early embryos and adult gonads as *in vivo* systems to study gene expression in stem cells and their differentiated derivatives.

By using the two color fluorescent in situ hybridization procedure[37,38], the transcripts of both *nanog* and *oct4* were found to be present already in blastomeres of the 16-cell staged embryo (Fig. 4A-C), in accordance with a high level of their RNAs by RT-PCR analysis (Fig. 3B). Notably, the blastomeres at this stage are thought to be totipotent cells in which zygotic transcription has not yet commenced, reinforcing that both *nanog* and *oct4* are maternally supplied and expressed in totipotent cells. We observed that a majority (75%, n = 12) of 16-cell embryos exhibited a significantly more intense signal for both *nanog* and *oct4* in the four inside blastomeres compared to the 12 outside blastomeres (Fig. 4A-C). Therefore, expression of both *nanog* and *oct4* is associated with embryonic pluripotency.

In medaka, the female produces eggs every day, and the mature ovary comprises a small number of oogonia and oocytes at ten ontogenic stages [39]. The adult medaka testis is a highly ordered organ, which is composed of spermatogonia and differentiating spermatogenic cells within seminiferous cysts. Spermatogonia are single cells or form small clusters, and

are located at the outmost periphery. Seminiferous cysts comprise of germ cells at various stages of spermatogenesis. Spermatogenesis proceeds synchronously within each cyst through primary spermatocytes (meiosis I), secondary spermatocytes (meiosis II), spermatids (meiosis completion) and sperm (terminally morphological differentiation). Importantly, cysts containing germ cells at progressive stages of spermatogenesis are regularly arranged, with cysts consisting of germ cells at the most advanced stages of development residing closer to the efferent duct, which is in the central region. The unique architecture and bipolarity of spermatogenesis allow for unambiguous distinction of stem cells and their progressive differentiation.

We chose *nanog* as a representative of group-I genes for chromogenic in situ hybridization in the ovary and testis. In the ovary, *nanog* was found to be expressed at a high level in the female germ stem cells oogonia (Fig. 4D). A weaker signal was also detected in small oocytes (Fig. 4E). It has been shown in mouse that the oocyte is an exception to differentiated cell, because it expresses pluripotency genes such as *oct4* at a high level for maternal supply to early embryogenesis[35]. In the testis, the *nanog* signal is detectable only in spermatogonia (Fig. 4F). In both ovary and testis, *nanog* expression is absent in somatic cells (Fig. 4D-F). Therefore, *nanog* expression is a pluripotency marker in the adult gonads.

We then examined the expression of *tcf3* as a representative of group-III genes in the testis. The *tcf3* signal is moderate in spermatogonia, peaks in spermatocytes and declines in spermatids and sperm (Fig. 4G). The wide *tcf3* expression in various stages of spermatogenesis is coincident with its stronger expression in the testis than in the ovary as revealed by RT-PCR analysis (Fig. 3A). Apparently, *tcf3* expression in the testis is not limited to stem cells but may be important also for spermatogenesis progression.

Expression in ES cell culture

Finally, we made use of ES cell cultures to validate the candidature of the seven medaka genes as pluripotency markers. We took the advantageous availability of well characterized medaka ES cell lines. MES1 is a diploid ES cell line from fertilization blastulae[6] and capable of chimera formation at high efficacy[27,40]. HX1 is a haploid ES cell line from gynogenetic blastulae and capable of whole animal production by semicloning[29,30]. To this end, both cell lines were maintained under undifferentiated conditions or subjected to suspension culture for embryoid body formation to induce differentiation[6,29]. Undifferentiated ES cells and differentiated samples

were analyzed for gene expression by RT-PCR. We found that all the seven genes were highly expressed in undifferentiated cultures of both ES cell lines but dramatically reduced after 7 days of induced differentiation in suspension culture (Fig. 5A). Successful induction of differentiation was evaluated by the expression of *no tail* (*ntl*), a mesendodermal marker encoding the T-box containing protein Brachyury[41]. As we have previously reported[29], *ntl* expression was absent in ES cells before differentiation but high upon differentiation (Fig. 5A). Therefore, all the seven genes appear to be markers for pluripotency in undifferentiated ES cell cultures.

To investigate gene expression before and after ES cell differentiation in more detail, we analyzed expression of *nanog* and *oct4* by real-time RT-PCR in MES1 cells. We found that the transcripts of both genes were reduced by ≥4-fold (Fig. 5B), reinforcing the usefulness of the seven genes, in particular *nanog* and *oct4*, for monitoring pluripotency in fish stem cell culture.
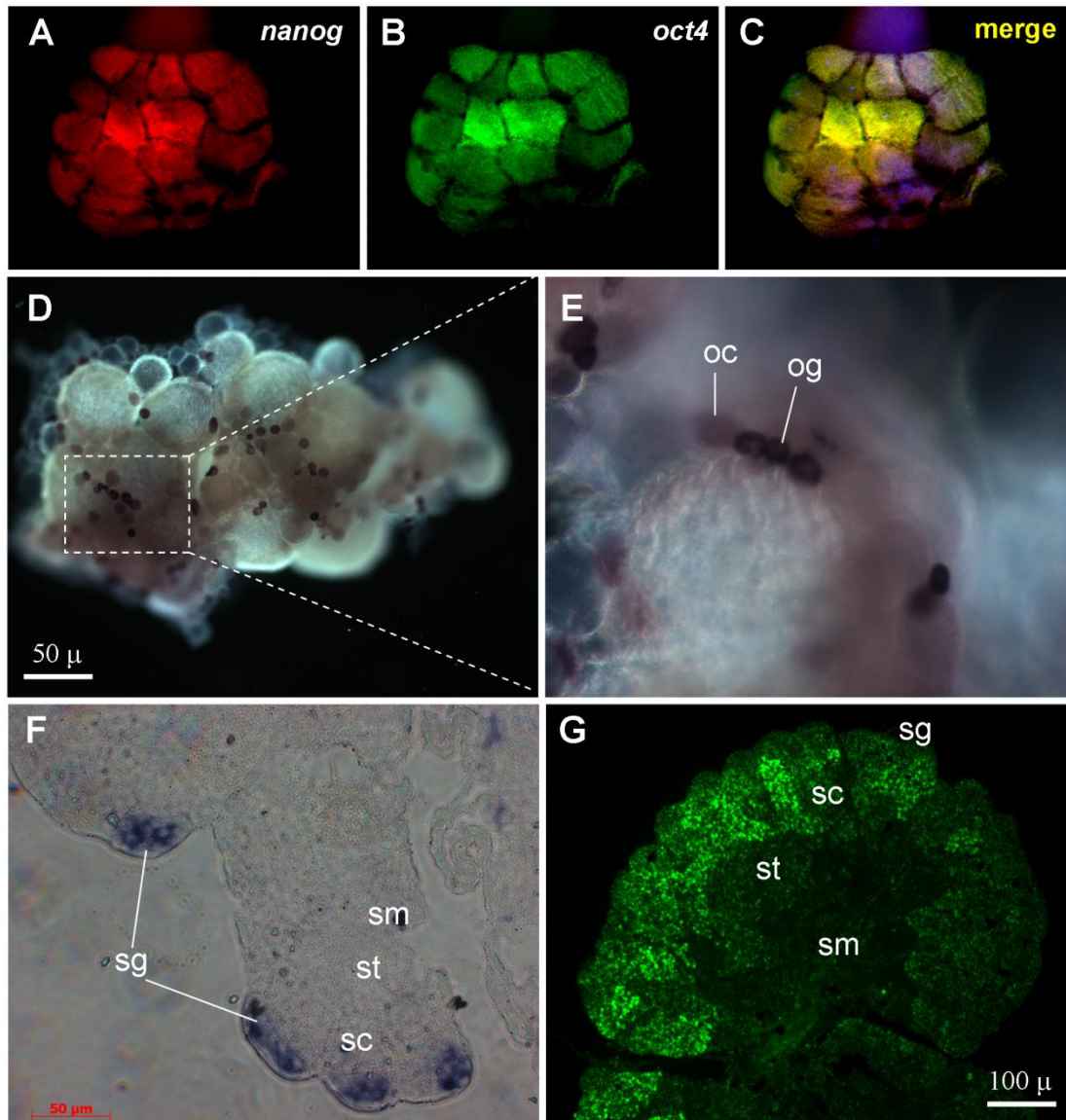

Fig. 4. RNA expression by in situ hybridization. (A-C) 16-cell embryo after two-color in situ hybridization showing *nanog* (red) and *oct4* (green) RNA expression. (D-E) Adult ovary after chemical in situ hybridization with a *nanog* riboprobe. (F) Adult testicular section after chemical in situ hybridization with a *nanog* riboprobe. (G) Adult testicular section after fluorescent in situ hybridization with a *tcf3* probe. og, oogonia; oc, oocytes; sg, spermatogonia; sc, spermatocytes; st, spermatids; sm, sperm.
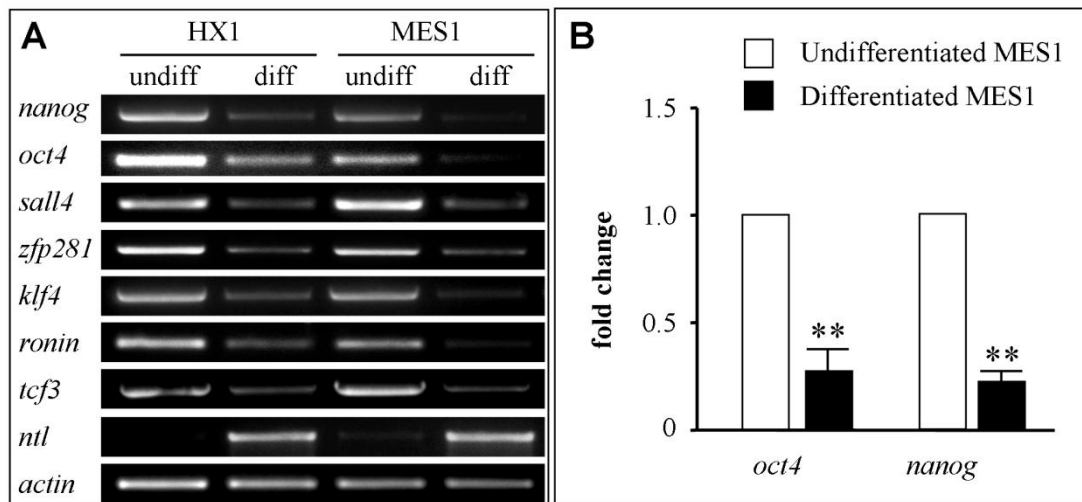
Fig. 5. RNA expression in ES cell culture. ES cells were maintained in adherent culture for undifferentiated (undiff) growth or in suspension culture for 10 days for induced differentiation (diff) by formation of embryoid bodies. (A) RT-PCR analysis in haploid (HX1) and diploid ES cell line (MES1). The mesendodermal marker *ntl* was used as differentiation marker. β-actin was used as a loading control. (B) Real-time RT-PCR analysis of oct4 and nanog RNA expression. Data are means ± s.d (bars above columns) of three samples; **, p ≤ 0.01.

Discussion

In this study we report the identification and expression of seven medaka homologs/paralogs of mammalian pluripotency genes. It has generally been accepted that fish has more genes than mammals because of one additional genome duplication (WGD) event in the common fish lineage after its separation ~450 million years ago from the tetrapod lineage leading to birds and mammals[42]. Duplicated fish genes from WGD may have different fates compared to their single copy gene counterparts in tetrapod vertebrates: One gene remains as homolog/paralog, whereas the other may have got lost, or become co-homolog or co-paralog (by loss of the chromosome synteny and original function). Many fish genes are said to be mammalian homologs on the basis of sequence comparisons and/or expression patterns. Examples are recent reports on medaka *nanog*[43] and *oct4*[44]. A phylogenetic sequence comparison and a conserved syntenic relationship together with the expression pattern suggest that the medaka *tcf3* identified in this study is homologous to the mammalian gene. However, in the absence of syntenic data, we cannot determine whether the other medaka genes are homologous or parologous to the mammalian counterparts.

According to their RNA expression patterns in adult tissues, the seven medaka genes fall into three groups. Specifically, *nanog* and *oct4* belong to group I, whose mouse counterparts are the best studied pluripotency genes in ES cell culture. Interestingly, we show for the first time that medaka *nanog* and *oct4* are expressed in 16-cell embryos in which they even show co-localization in certain blastomeres. We reveal in this study that medaka *nanog* expression delineates with pluripotency in early embryos, adult gonads and ES cell culture, an expression pattern reminiscent of that for the mouse *oct4*, the prototype pluripotency gene that shows expression throughout the totipotent cycle and ES cells[35]. It deserves to note that medaka *nanog* possesses a salient difference in RNA expression *in vivo* from the mouse *nanog*, which has been reported to lack maternal inheritance and expression in oocytes and spermatogonia of the adult gonads [19]. The *in vivo* RNA expression pattern of medaka *nanog* we observed in this study conforms in general to its protein expression detected by immunostaining, except for reported protein expression also in meiotic male germ cells[43]. The *in vivo* RNA expression pattern of medaka *oct4* obtained in the present study is similar to that reported[45]. Taken together, medaka *nanog* and *oct4* are pluripotency markers *in vivo* and *in vitro*.

The other five genes, namely *klf4, ronin, sall4, tcf3* and *zfp281a,* are of groups II and III, whose expression occurs also in one or more somatic tissues besides the adult gonads and persists beyond the gastrulation. These results suggest that expression of these genes is not restricted to pluripotency *in vivo* but extended for

other processes. The five medaka genes of groups II and III show similar RNA expression patterns to those of their mouse counterparts, suggesting their homology in expression pattern. On the other hand, these genes *in vitro* closely resemble *nanog* and *oct4*, in that they exhibit a high level of expression in undifferentiated ES cells and dramatic down-regulation upon ES cell differentiation. Accordingly, these five genes appear to be also pluripotency markers and perhaps even regulators for fish ES cell culture. Obviously, in medaka, the lack of pluripotency-specific expression *in vivo* is not exclusive for a gene in question as a pluripotency marker or even regulator *in vitro*. A similar situation is seen also in mouse, where *klf4* exhibits wide expression (see below) but is an integral component of the four factors capable of reprogramming differentiated cells into iPS cells[8].

Six of the seven medaka genes have previously been found to be expressed highly in undifferentiated cultures of two haploid ES cell lines (HX1 and HX2) and the diploid MES1 line, and two of them, *nanog* and *oct4* exhibited dramatic down-regulation upon haploid ES cell differentiation [29]. The present study corroborates and extends our previous work by demonstrating one and the same pluripotency-specific expression pattern in ES cells for a set of seven genes including *tcf3*, a new member identified in this study.

By identifying candidate pluripotency genes through the homology approach and by examining their expression *in vivo* and *in vitro* in this study, we have revealed two important findings in medaka. The first is that pluripotency-specific expression *in vivo* is strongly indicative of pluripotency-specific expression also *in vitro*, as is the case for *nanog* and *oct4*. The other is that pluripotency genes in ES cell cultures do not necessarily possess pluripotency-specific expression *in vivo*, as is the case for the five other genes. Taken together, our data underscore the requirement for established ES cell cultures as a unique *in vitro* system to ultimately test putative pluripotency genes for practical use in monitoring the pluripotency of putative ES cells and other stem cell cultures.

In broad sense, all the seven mammalian pluripotency genes show generally similar *in vivo* RNA expression patterns in adult somatic tissues compared to their medaka counterparts. In mice, for example, the *sall4* expression occurs in several somatic tissues such as the brain, heart and muscle[22], and the *tcf3* expression is also not limited to stem cells[23]. In medaka, out of six somatic tissues examined in this study, the transcripts of *sall4* and *tcf3* are detectable also in three (namely the brain, kidney and liver) and five somatic tissues, respectively. In this study, we have further demonstrated that *tcf3* is detectable throughout medaka spermatogenic stages, strengthening its wide expression *in vivo*. Expression besides in stem cells *in vivo* suggests additional functions. In zebrafish, for example, both *tcf3* genes are required for brain patterning [32,33], and *oct4* is essential for morphogenesis, brain development[46] and endodermal development[47]. In this regard, many pluripotency genes, such as those of groups II and III, appear to be pleiotropic *in vivo*, Taken together, these data reinforce our notion that a test in ES cell culture is most convincing for validating pluripotency genes.

Our observation that the medaka *klf4* has maternal inheritance and early embryonic expression, is similar to that in zebrafish, where expression of *klf4* in early embryos has also been documented[48]. In mouse, however, *klf4* lacks expression in early embryos until the blastocyst stage, leading to a notion that *klf5* might substitute *klf4* for blastocyst development [49]. Work is needed to determine whether this salient difference in embryonic *klf4* expression is due to the species difference or independent divergence between the fish and tetrapod lineages. A similar difference has also been observed for *sox2* expression even between mouse [15,16] and human [17].

Our work in medaka generally corroborates observations reported in mammals and extends these reports to fish. Importantly, we show that all the seven medaka genes identified by homology search are molecular markers for undifferentiated ES cells, as has been reported for their expression and use in mammalian ES cell research. This indicates the applicability of the homology approach to identify pluripotency genes also in other fish species and even other lower vertebrates. In this regard, our demonstration that the five genes of expression groups II and III are ES cell markers has an important implication in practice, because the protein products of these genes show a substantially higher sequence conservation for easier identification and cloning in organisms that lack sequence data, compared to *nanog* and *oct4*, both of which exhibit considerable divergence in protein sequence. Our data in the present study demonstrate that pluripotency genes – in spite of sequence divergence to varying extent - have conserved their pluripotency-specific expression *in vitro* from mammals to lower vertebrates. This notion is supported by our previous demonstration that the mouse *oct4* promoter is active in medaka to drive transgene expression of reporters in early embryos and drug selectable markers in undifferentiated ES cells for drug ablation of spontaneous differentiation[50]. We anticipate that the seven pluripotency genes and our approach for identifying and testing them as pluripotency markers

in vitro will provide a general tool for stem cell derivation in other fish species and other lower vertebrates [55, 56].

Materials and Methods

Fish. Work with fish followed the guidelines on the Care and Use of Animals for Scientific Purposes of the National Advisory Committee for Laboratory Animal Research in Singapore (permit number 27/09). Medaka was maintained under an artificial photoperiod of 14-h/10-h light/darkness at 26°C [51]. Embryogenesis was staged according to Iwamatsu[52]. Medaka strains HB32C and i^1 were used for gene expression analyses by RT-PCR and in situ hybridization. HB32C is a wild-type pigmentation strain from which diploid ES cell lines including MES1 were derived [6], whereas i^1 is an albino strain from which haploid ES cell lines including HX1 were generated [29,30].

Cell culture. Maintenance and induced differentiation by embryoid body formation in suspension culture were done essentially as described for the diploid ES cell line MES1 [6] and the haploid ES cell line[29,30]. Briefly, ES cells were cultured in medium ESM4 in gelatin-coated tissue culture plastic ware (BD biosciences, NJ) for undifferentiated growth. For induced differentiation, ES cells were trypsinized into single cells and small aggregates, seeded into non-adherent bacteriological Petri dishes for suspension culture in growth factor-depleted ESM4 containing all-trans retinoic acid (Sigma, final 5 µM) for 7 days before harvest for RNA isolation. In suspension culture, cells formed aggregates of varying sizes (mostly 100~20 µm in diameter) at day 1 and subsequently developed into spherical embryoid bodies. The dishes of suspension cultures were gently swirled twice a day to prevent any attachment.

Sequence Analysis

BLAST searches were run against public databases by using BLASTN for nucleotide sequences, BLASTP for protein sequences and tBLASTN from protein queries to nucleotide sequences. Multiple sequence alignment was conducted by using the Vector NTI suite 11 (Invitrogen, CA). Phylogenetic trees were constructed by using the DNAMAN package (Lynnon BioSoft, Canada). Chromosomal locations were investigated by examing corresponding genomic sequences.

Gene identification. Several medaka genes homologous/paralogous to the mammalian pluripotency genes have been previously described[29], including *oct4*, *nanog*, *klf4*, *ronin, sall4*, *zfp281a* and *tcf3*. However, the previous medaka *tcf3* according to the

genome annotation (ENSORLG00000004923 and ENSORLG00000015259) was found to encode transcription factor E2A instead (Fig. 1). A BLAST search by using the human TCF3 as a query against the medaka genome sequence (http://www.ensembl. org/index.html) led to the identification of a gene annotated as *tcf3* (ENSORLG00000011813), to which more than 30 expressed sequence tags (http://blast.ncbi.nlm.nih.gov/) displayed ≥96% identity in nucleotide sequence.

Cloning and Sequencing. A 433-bp fragment of the medaka *tcf3* cDNA was PCR-amplified by using the primer set (Table 1) and cloned into pGEM-T, resulting in pGEMtcf3. The insert was sequenced by using the Applied Biosystems 3130xl (Applied Biosystems, MA). The medaka *oct4* and *nanog* have previously been identified and cloned into pGEM-T, resulting in pGEMoct[53] and pGEMnanog[54].

RNA isolation and RT-PCR analysis. Total RNA was isolated by using the Trizol Reagent (Invitrogen, CA). Synthesis of cDNA templates was primed with oligo(dT)$_{25}$ by using M-MLV transcriptase (Promega, USA). PCR was run in a 25-µl volume containing 10 ng of cDNA reaction for 30 and 35 cycles (20 s at 94°C, 20 s at 58°C and 120 s at 72°C) for β-actin and other genes, respectively. Primers and gene accession numbers are listed in Table 2. The PCR products were separated on 1% agarose gels and documented with a bioimaging system (Synoptics, UK). Real-time RT-PCR was run in triplicate on the IQ5 system (BioRad, CA).

RNA *in situ* hybridization. RNA *in situ* hybridization by chromogenic staining with BCIP/NBT substrates on whole mount samples and sections were performed as described previously[51]. For probe preparation, pGEMnanog, pGEMoct4 and pGEMtcf3 were linearized with Sal I or Apa1 for the synthesis of sense and anti-sense RNA probes from SP6 or T7 promoter by using the digoxigenin (DIG) or FITC RNA Labeling Kit (Roche, Switzerland). The RNA probes were treated with RNase-free TURBO DNase (Ambion, MA). Fluorescent *in situ* hybridization was performed by using the TSA™ Plus Fluorescence Systems (NEN Life Science) as described[37,38]. Nuclear staining was done by using 4'-6-Diamidino-2-phenylindole (DAPI) in the Gold Antifade reagent (Invitrogen, CA).

Microscopy and Photography. Observation and photography on Leica MZFIII stereo microscope, Zeiss Axiovertinvert 2 invert microscope and Axiovert 200 upright microscope with a Zeiss AxioCam MRc5 digital camera (Zeiss Corp., Germany) were as described previously [29].

Statistics. Statistical analyses were calculated by

using GraphPad Prism v5.0.

Acknowledgment

Conflict of Interests

The authors have declared that no conflict of interest exists.

References

1. Wobus AM, Boheler KR. Embryonic stem cells: prospects for developmental biology and cell therapy. Physiol Rev 2005; 85: 635-678.
2. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature 1981; 292: 154-156.
3. Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 1981; 78: 7634-7638.
4. Munoz M, Trigal B, Molina I, Diez C, Caamano JN, et al. Constraints to progress in embryonic stem cells from domestic species. Stem Cell Rev 2009; 5: 6-9.
5. Yi M, Hong N, Li Z, Yan Y, Wang D, et al. Medaka fish stem cells and their applications. Sci China Life Sci 2010;53: 426-434.
6. Hong Y, Winkler C, Schartl M. Pluripotency and differentiation of embryonic stem cell lines from the medakafish (Oryzias latipes). Mech Dev 1996; 60: 33-44.
7. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, et al. Embryonic stem cell lines derived from human blastocysts. Science 1998; 282: 1145-1147.
8. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006; 126: 663-676.
9. Hong N, Li Z, Hong Y. Fish Stem Cell Cultures. Int J Biol Sci 2011; 7:392-402.
10. Ma C, Fan L, Ganassin R, Bols N, Collodi P. Production of zebrafish germ-line chimeras from embryo cell cultures. Proc Natl Acad Sci U S A 2001; 98: 2461-2466.
11. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 2005; 122: 947-956.
12. Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 2006; 38: 431-440.
13. Pesce M, Wang X, Wolgemuth DJ, Scholer H. Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. Mech Dev 1998; 71: 89-98.
14. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 2000; 24: 372-376.
15. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, et al. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 2003; 17: 126-140.
16. Western P, Maldonado-Saldivia J, van den Bergen J, Hajkova P, Saitou M, et al. Analysis of Esg1 expression in pluripotent cells and the germline reveals similarities with Oct4 and Sox2 and differences between human pluripotent cell lines. Stem Cells 2005; 23: 1436-1442.
17. Perrett RM, Turnpenny L, Eckert JJ, O'Shea M, Sonne SB, et al. The early human germ cell lineage does not express SOX2 during *in vivo* development or upon *in vitro* culture. Biol Reprod 2008; 78: 852-858.
18. Cavaleri F, Scholer HR. Nanog: a new recruit to the embryonic stem cell orchestra. Cell 2003; 113: 551-552.
19. Chambers I, Colby D, Robertson M, Nichols J, Lee S, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 2003; 113: 643-655.
20. Dejosez M, Krumenacker JS, Zitur LJ, Passeri M, Chu LF, et al. Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell 2008; 133: 1162-1174.
21. Lim CY, Tam WL, Zhang J, Ang HS, Jia H, et al. Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. Cell Stem Cell 2008; 3: 543-554.
22. Tsubooka N, Ichisaka T, Okita K, Takahashi K, Nakagawa M, et al. Roles of Sall4 in the generation of pluripotent stem cells from blastocysts and fibroblasts. Genes Cells 2009; 14: 683-694.
23. Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev 2008; 22: 746-755.
24. Tam WL, Lim CY, Han J, Zhang J, Ang YS, et al. T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. Stem Cells 2008; 26: 2019-2031.
25. Wang ZX, Teh CH, Chan CM, Chu C, Rossbach M, et al. The transcription factor Zfp281 controls embryonic stem cell pluripotency by direct activation and repression of target genes. Stem Cells 2008; 26: 2791-2799.
26. Wittbrodt J, Shima A, Schartl M. Medaka--a model organism from the far East. Nat Rev Genet 2002; 3: 53-64.
27. Hong Y, Winkler C, Schartl M. Production of medakafish chimeras from a stable embryonic stem cell line. Proc Natl Acad Sci U S A 1998; 95: 3679-3684.
28. Hong Y, Liu T, Zhao H, Xu H, Wang W, et al. Establishment of a normal medakafish spermatogonial cell line capable of sperm production *in vitro*. Proc Natl Acad Sci U S A 2004; 101: 8011-8016.
29. Yi M, Hong N, Hong Y. Generation of medaka fish haploid embryonic stem cells. Science 2009; 326: 430-433.
30. Yi M, Hong N, Hong Y. Derivation and characterization of haploid embryonic stem cell cultures in medaka fish. Nat Protoc 2010; 5: 1418-1430.
31. Li Z, Bhat N, Manali D, Wang D, Hong N, Yi M, Ge R, Hong Y. Medaka Cleavage Embryos Are Capable of Generating ES-Like Cell Cultures. Int J Biol Sci 2011; 7:418-425
32. Kim CH, Oda T, Itoh M, Jiang D, Artinger KB, et al. Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. Nature 2000; 407: 913-916.
33. Dorsky RI, Itoh M, Moon RT, Chitnis A. Two tcf3 genes cooperate to pattern the zebrafish brain. Development 2003; 130: 1937-1947.
34. Aizawa K, Shimada A, Naruse K, Mitani H, Shima A. The medaka midblastula transition as revealed by the expression of the paternal genome. Gene Expr Patterns 2003; 3: 43-47.
35. Pesce M, Gross MK, Scholer HR. In line with our ancestors: Oct-4 and the mammalian germ. Bioessays 1998; 20: 722-732.
36. Guan K, Nayernia K, Maier LS, Wagner S, Dressel R, et al. Pluripotency of spermatogonial stem cells from adult mouse testis. Nature 2006; 440: 1199-1203.
37. Liu L, Hong N, Xu H, Li M, Yan Y, et al. Medaka dead end encodes a cytoplasmic protein and identifies embryonic and adult germ cells. Gene Expr Patterns. 2009 Oct;9(7):541-8
38. Xu H, Li Z, Li M, Wang L, Hong Y. Boule is present in fish and bisexually expressed in adult and embryonic germ cells of medaka. PLoS One 2009;4: e6097.
39. Iwamatsu T, Ohta T, Oshima E, N S. Oogeneis in the Medaka Oryzias latipes -- Stage of Oocyte Development. Zoological science 1988; 5: 353-373.

40. Hong N, Li M, Zeng Z, Yi M, Deng J, et al. Accessibility of host cell lineages to medaka stem cells depends on genetic background and irradiation of recipient embryos. Cell Mol Life Sci. 2010;67(7):1189-1202.

41. Araki K, Okamoto H, Graveson AC, Nakayama I, Nagoya H. Analysis of haploid development based on expression patterns of developmental genes in the medaka Oryzias latipes. Dev Growth Differ 2001; 43: 591-599.

42. Amores A, Force A, Yan YL, Joly L, Amemiya C, et al. Zebrafish hox clusters and vertebrate genome evolution. Science 1998; 282: 1711-1714.

43. Camp E, Sanchez-Sanchez AV, Garcia-Espana A, Desalle R, Odqvist L, et al. Nanog regulates proliferation during early fish development. Stem Cells 2009; 27: 2081-2091.

44. Sanchez-Sanchez AV, Camp E, Garcia-Espana A, Leal-Tassias A, Mullor JL. Medaka Oct4 is expressed during early embryo development, and in primordial germ cells and adult gonads. Dev Dyn 2010;239: 672-679.

45. Sanchez-Sanchez AV, Camp E, Garcia-Espana A, Leal-Tassias A, Mullor JL. Medaka Oct4 is expressed during early embryo development, and in primordial germ cells and adult gonads. Dev Dyn 2010; 239: 672-679.

46. Burgess S, Reim G, Chen W, Hopkins N, Brand M. The zebrafish spiel-ohne-grenzen (spg) gene encodes the POU domain protein Pou2 related to mammalian Oct4 and is essential for formation of the midbrain and hindbrain, and for pre-gastrula morphogenesis. Development 2002; 129: 905-916.

47. Lunde K, Belting HG, Driever W. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. Curr Biol 2004; 14: 48-55.

48. Luo DJ, Hu W, Chen SP, Zhu ZY. Critical Developmental Stages for the Efficiency of Somatic Cell Nuclear Transfer in Zebrafish. Int J Biol Sci 2011; 7:476-486

49. Ema M, Mori D, Niwa H, Hasegawa Y, Yamanaka Y, et al. Kruppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. Cell Stem Cell 2008; 3: 555-567.

50. Hong Y, Winkler C, Liu T, Chai G, Schartl M. Activation of the mouse Oct4 promoter in medaka embryonic stem cells and its use for ablation of spontaneous differentiation. Mech Dev 2004; 121: 933-943.

51. Li M, Hong N, Xu H, Yi M, Li C, et al. Medaka vasa is required for migration but not survival of primordial germ cells. Mech Dev 2009; 126: 366-381.

52. Iwamatsu T. Stages of normal development in the medaka Oryzias latipes. Mech Dev 2004; 121: 605-618.

53. Liu R. Isolation and characterization of stem cell regulatory genes oct4 and stat3 from the model fish medaka (Oryzias latipes). Singapore: National University of Singapore Doctoral thesis. 2008

54. Li Z. Nanog in the twin fish models medaka and zebrafish: Functional divergence or pleiotropy of vertebrate pluripotency gene? Singapore: National University of Singapore Doctoral thesis. 2009

55. Sánchez-Sánchez AV, Camp E, Mullor JL. Fishing Pluripotency Mechanisms *In Vivo*. Int J Biol Sci 2011; 7:410-417

56. Rao F, Wang T, Li M, Li Z, et al. Medaka tert produces multiple variants with differential expression during differentiation *in vitro* and *in vivo*. Int J Biol Sci. 2011; 7(4): 426-439.