

RESEARCH ARTICLE

A novel role for *sox7* in *Xenopus* early primordial germ cell development: mining the PGC transcriptome

Amanda M. Butler¹, Dawn A. Owens¹, Lingyu Wang^{2,*} and Mary Lou King^{1,†}

ABSTRACT

Xenopus primordial germ cells (PGCs) are determined by the presence of maternally derived germ plasm. Germ plasm components both protect PGCs from somatic differentiation and begin a unique gene expression program. Segregation of the germline from the endodermal lineage occurs during gastrulation, and PGCs subsequently initiate zygotic transcription. However, the gene network(s) that operate to both preserve and promote germline differentiation are poorly understood. Here, we utilized RNA-sequencing analysis to comprehensively interrogate PGC and neighboring endoderm cell mRNAs after lineage segregation. We identified 1865 transcripts enriched in PGCs compared with endoderm cells. We next compared the PGC-enriched transcripts with previously identified maternal, vegetally enriched transcripts and found that ~38% of maternal transcripts were enriched in PGCs, including *sox7*. PGC-directed *sox7* knockdown and overexpression studies revealed an early requirement for *sox7* in germ plasm localization, zygotic transcription and PGC number. We identified *pou5f3.3* as the most highly expressed and enriched *POU5F1* homolog in PGCs. We compared the *Xenopus* PGC transcriptome with human PGC transcripts and showed that 80% of genes are conserved, underscoring the potential usefulness of *Xenopus* for understanding human germline specification.

KEY WORDS: *Xenopus*, RNA-seq, PGCs, *sox7*, *pou5f3.3* (*oct60*), Germ plasm

INTRODUCTION

The exclusion of germ cells from somatic cell fates in early development is an essential process in metazoans that ensures continuation of the species. Germline specification in *Xenopus* is established by inheritance of germ plasm, a subcellular matrix containing maternally derived RNAs and proteins. Germ plasm contains all the genetic information that protects primordial germ cells (PGCs) from somatic differentiation and initiates a unique gene expression program that preserves their potential for totipotency and differentiation. Furthermore, germ plasm has been shown to be both required and sufficient to determine germ cell fate in *Xenopus* (Tada et al., 2012). Germ plasm components are localized, along with somatic determinants, to the vegetal pole during oogenesis (Forristall et al., 1995; Heasman et al., 1984; Kloc and Etkin, 1995; Zhang et al.,

1998). During cleavage stages, cells containing germ plasm undergo asymmetric division so that the germ plasm is only inherited by one daughter cell termed the presumptive PGC (pPGC). Although somatic determinants are partitioned into pPGCs during cleavage stages, the genetic programs for somatic fate are not activated there because of translational repression and transient suppression of RNA polymerase II-regulated transcription (Lai and King, 2013; Venkatarama et al., 2010). Segregation of the germline occurs at gastrulation when the germ plasm moves to a perinuclear location and subsequent divisions result in both daughter cells, now termed PGCs, receiving germ plasm. PGCs then initiate their zygotic transcription program driven by unknown maternal transcription factors. However, the activated gene network necessary for proper PGC specification and development has not been characterized in *Xenopus*.

To begin to understand the gene network required for proper PGC development, we recently completed RNA-sequencing (RNA-seq) analysis of both animal and vegetal pole tips isolated from stage VI oocytes (Owens et al., 2017). This work defined 198 annotated RNAs highly enriched at the vegetal pole including several known germline and somatic determinants. Further analysis confirmed known germline components, such as *pgat* (also known as *xpat*), *sybu* and *otx1*, and identified novel components including the transcription factor (TF) *sox7* (Owens et al., 2017). The F-*sox* family member *sox7* has previously been shown to be an early downstream target of VegT and to induce expression of *nodal* genes necessary for somatic fates (Zhang, et al., 2005). Similar to *Xenopus* Sox7 (Hudson et al., 1997; Zhang et al., 2005), human SOX17, another F-*sox* family member, has also historically been reported as an essential transcription factor required for endoderm specification (Charney et al., 2017; Hudson et al., 1997; Irie et al., 2015). Interestingly, Irie and co-workers generated human primordial germ cell-like cells (hPGCLCs) from embryonic stem cells and identified *SOX17* as the primary regulator of human primordial germ cell-like fate (Irie et al., 2015).

In the present study, we utilized RNA-seq analysis to determine the zygotic PGC transcriptome in *Xenopus* by comprehensive interrogation of PGC and neighboring endoderm cell RNAs just after lineage segregation. We identified 1865 transcripts enriched in PGCs, and over a third of the 198 annotated, vegetally enriched transcripts (Owens et al., 2017) were among them, including *sox7*. To elucidate the specific role of *sox7* in PGCs, we directed *sox7* knockdown and overexpression constructs to the germline. Our results indicate that, prior to neurula, *sox7* is necessary for proper germ plasm localization, timely zygotic transcription and correct PGC number. These data provide further evidence that *sox7* is a crucial TF required for PGC development.

In addition to *sox7*, *pou5f3.3* (also known as *oct60*), which is in the same POU subclass as the human pluripotency factor *POU5F1* (also known as *OCT3/4*) (Frankenberg and Renfree, 2013; Hellsten et al., 2010; Hinkley et al., 1992), is enriched in PGCs. As *pou5f3.3* is not enriched at the vegetal pole of stage VI oocytes (Owens et al., 2017), along with other known germ plasm transcripts, it might

¹Department of Cell Biology, University of Miami Miller School of Medicine, 1011 NW 15th St, Miami, FL 33136, USA. ²Department of Biology, University of Miami, Coral Gables, FL 33124, USA.

*Present address: Department of Biology, Duke University, 125 Science Drive, Box 90338, Durham, NC 27708, USA.

†Author for correspondence (mking@med.miami.edu)

 M.L.K., 0000-0002-1923-4283

represent a zygotic germ plasm transcript required for proper PGC specification. In fact, *POU5F1* is considered a key gene necessary for human PGC (hPGC) specification (Tang et al., 2016), and *POU5F1* acts as a functional homolog for *pou5f3.3* in rescue experiments (Frankenberg and Renfree, 2013; Hellsten et al., 2010; Hinkley et al., 1992). In the present study, we show for the first time that *pou5f3.3* plays a crucial role in early development of *Xenopus* PGCs. Furthermore, we compared the *Xenopus* PGC transcriptome with the human PGC and hPGCLC transcriptomes (see supplementary information in Irie et al., 2015), and show that 80% of genes are conserved. Taken together, these data indicate that *Xenopus* is a genetically relevant system for modeling the gene networks necessary for human germline specification and development.

RESULTS

RNA-sequence analysis of PGCs after segregation from the endoderm

PGCs initiate zygotic transcription after they segregate from the endoderm at gastrulation (Venkatarama et al., 2010). To identify transcripts involved in PGC development, we took advantage of the large abundance of mitochondria in germ plasm to identify these cells. Four-cell embryos were stained briefly with the mitochondrial lipophilic dye DiOC₆, and when they reached stages (st.) 12.5-14, PGCs and embryo-matched endoderm cells were collected as described previously (Butler et al., 2017). In total, RNA was extracted from ~1040 PGCs and ~2100 paired endoderm cells from embryos generated by 28 female frogs. The respective RNA samples were split into three technical replicates each and then processed and sequenced as shown in Fig. S1 and described in Materials and Methods. Although version 9.1 of the *Xenopus laevis* genome (Xenbase.org, September, 2015) contained a higher number of transcripts and was better annotated, it did not contain all the transcripts in the previous version, 7.1. Therefore, the raw reads of

each sample were aligned to both v9.1 and v7.1 of the *Xenopus laevis* genome (Xenbase.org, September, 2015). The aligned reads for each genome version were sorted and counted, then merged together before normalization and data analysis. For transcripts identified in both v9.1 and v7.1, read counts from v9.1 were used. In total, 13,469 transcripts were identified (Table S1, transcripts unique to v7.1 are highlighted in yellow). Two-dimensional principal component analysis (PCA) showed that all three PGC samples cluster together and away from the three endoderm samples, which also cluster together (Fig. 1A). Differential expression analysis was performed and results are shown as a MA plot, in which M represents the log ratio and A represents mean average (Fig. 1B).

Transcripts were considered significantly enriched in PGCs or endoderm cells if their expression was increased at least threefold compared with endoderm cells or PGCs, respectively. Using these criteria, 1865 transcripts were enriched in PGCs (red) and 791 transcripts were enriched in endoderm cells (blue) (Table S1). A heat map was generated with the top 50 PGC-enriched transcripts (Fig. 1C). Several genes known to be expressed in PGCs, including *pgat*, *ddx25* (also known as *deadsouth*), *dazl*, *nanos1*, *grip2*, *velo1* and *dnd1* (Colozza and De Robertis, 2014; King et al., 2005; Tarbashevich et al., 2007), are among the top 50, confirming our sample identity. In addition to known germline RNAs, six of the top 50 PGC-enriched transcripts, namely *impad1*, *dsc3*, *fbxo34*, *fam168b*, *tspan1* and *cab39*, have not previously been shown to play a role in germ cell development (Fig. 1C).

Biological processes and network analysis of PGC-enriched transcripts

MetaCore (GeneGo) was used to interrogate potential gene networks operating in PGCs. Network analysis revealed five hubs (five or more connections) linking, in a direct-interaction network, 53 out of the top 150 PGC-enriched transcripts. Hubs included the

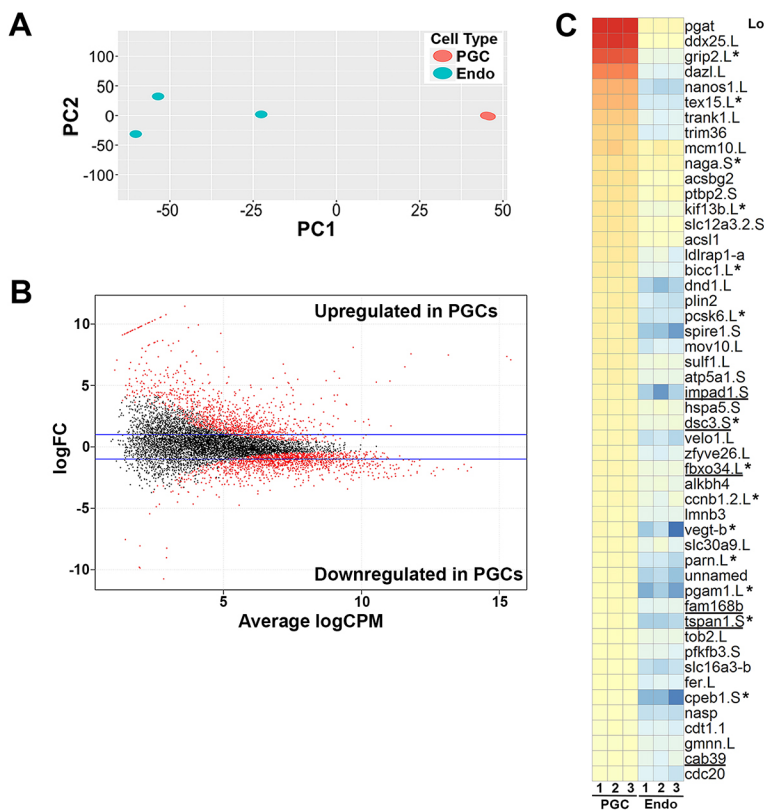


Fig. 1. RNA-seq analysis of PGC versus endoderm cell transcripts in *X. laevis* embryos. (A) Two-dimensional principal component analysis (PCA) of PGC and Endo transcripts from st. 12.5-14 embryos. Variance was 73.5% and 11.3% in PC1 and PC2, respectively. (B) Differential expression analysis of PGC versus Endo transcripts; red dots indicate differentially expressed transcripts. (C) Heat map of the top 50 PGC-enriched transcripts; novel germline transcripts are underlined. Endo, endoderm cell sample. Asterisks indicate transcripts that express more than one copy. 'L' and 'S' indicate long and short chromosomes, respectively. Note that only the more highly expressed copy in PGCs is shown.

TFs *e2f1*, *pou5f3.3/pou5f3.2* (two of three *POU5F1* *Xenopus* homologs; *pou5f3.2* is also known as *oct25*), *cyclin B2*, *chk1* (*chek1*) and *hells* (Fig. 2A). Interestingly, only 20 of the top 150 endoderm cell-enriched transcripts form a direct-interaction network (data not shown). This difference is probably because PGCs represent a single lineage, whereas several cell lineages are derived from endoderm cells (Chamey et al., 2017).

To identify potential biological processes necessary for PGC specification, the 300 most highly expressed, PGC-enriched transcripts were subjected to gene ontology analysis using GeneGo. The top 50 GeneGo processes are shown in Table S2. These processes were further categorized into four molecular functions including: cell cycle (93/300), cell division (53/300), reproduction (53/300) and apoptosis (31/300) (Fig. 2B). Considering that PGCs undergo few cell divisions before they infiltrate the somatic gonads much later in development, it was surprising that the most highly expressed PGC-enriched transcripts were primarily involved in cell cycle regulation and cell division.

We next identified 195 TFs that were enriched in PGCs, including *sox7*, *pou5f3.3*, *pou5f3.2*, *e2f1*, *mixer*, *otx1* and *twist1*, and 16 transcriptional co-factors (Table S1, tabs 3-5). Surprisingly, 89 of the identified TFs contain homeodomains. Homeodomain-containing TFs are crucial for proper developmental patterning in somatic tissues (Pearson et al., 2005; Philippidou et al., 2012), but their role in PGC development has not been characterized.

PGC transcriptome is conserved between *Xenopus* and human PGCs

Irie et al. (2015) recently performed RNA-seq analysis on human gonadal PGCs from week 7 human embryos (hPGCs), and PGC-like cells (hPGCLCs), created through differentiation of human embryonic stem cells. They identified *SOX17* as an early marker and necessary regulator of hPGCLC fate. Here, we performed RNA-seq analysis on *Xenopus* PGCs after segregation from the endoderm and showed that another F-sox family member, *sox7* is highly enriched in *Xenopus* PGCs (Table S1). To identify additional similarities between the *Xenopus* and human PGC transcriptomes, we performed a comparative analysis of hPGC and hPGCLC transcripts (Irie et al., 2015), versus our *Xenopus* PGC-enriched data set. Interestingly, we found that ~80% (1489/1865) of *Xenopus* PGC-enriched genes were also expressed in hPGCs (gray), hPGCLCs (green) or both (orange) (Table S3). Furthermore, of the top 50 *Xenopus* PGC-enriched transcripts, 90% are expressed in hPGCs and/or hPGCLCs (Table S3, tab 1). These results suggest that *Xenopus* is a relevant model system to study the gene networks required for human germline development.

Maternal, vegetally enriched transcripts are expressed in PGCs

During *Xenopus* oogenesis, hundreds of RNAs are selectively localized to the vegetal pole, including germ cell and endoderm

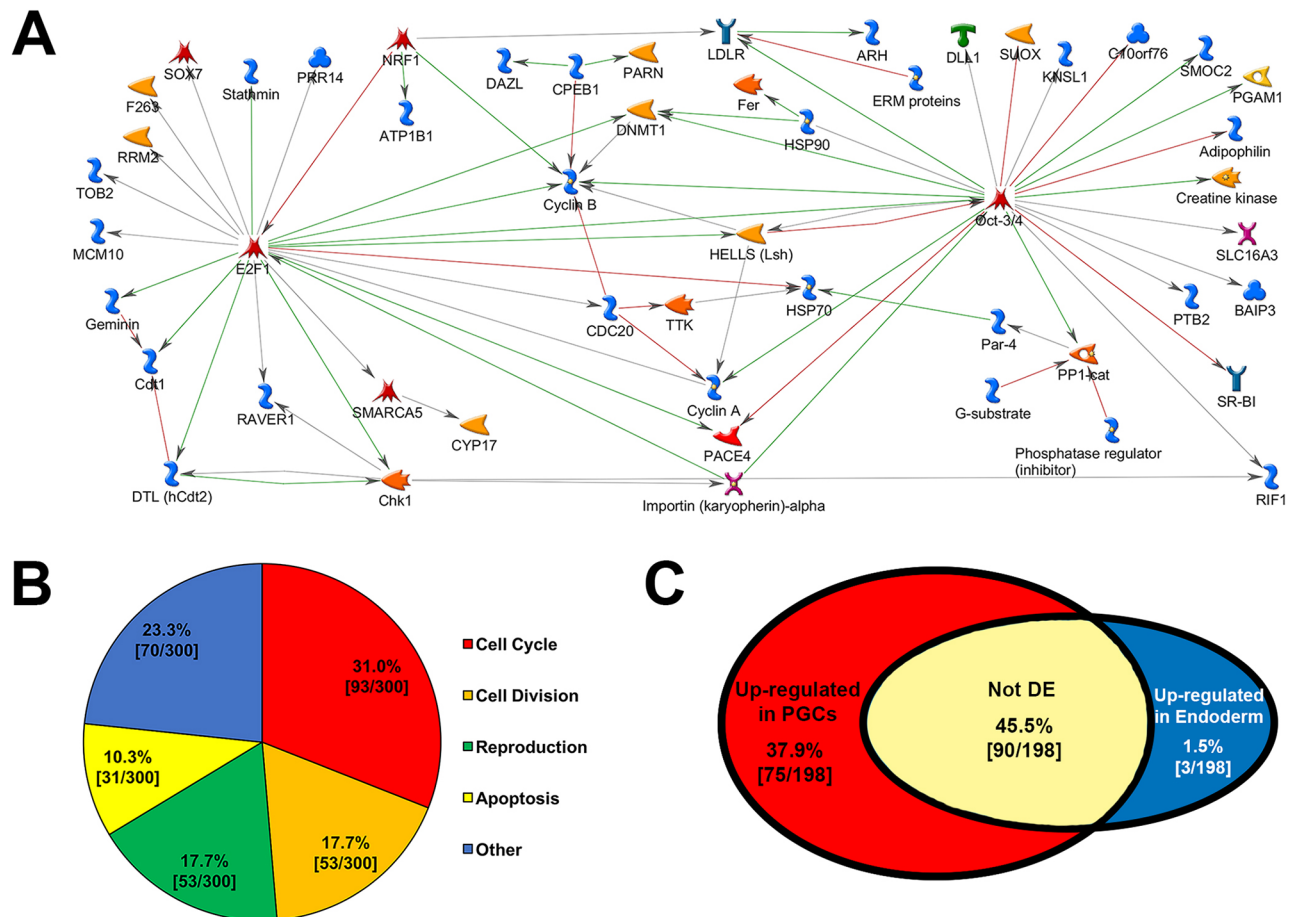


Fig. 2. Analysis of PGC-enriched transcripts. (A) GeneGo direct interaction network connecting 53 of the 150 most highly expressed PGC-enriched transcripts (network key: https://portal.genego.com/help/MC_legend.pdf). (B) Top 300 PGC-enriched transcripts categorized by molecular function. (C) Distribution of zygotically expressed maternal, vegetally enriched transcripts. DE, differentially expressed.

determinants. Recently, Owens et al. (2017) identified 411 RNAs enriched at the vegetal pole (compared with the animal pole), 198 of which were annotated. Considering that the maternal, vegetally enriched transcripts are known to contain both germ- and endoderm-cell determinants, we compared those annotated transcripts with the PGC/Endoderm cell data set. Of the 198 maternal, vegetally enriched transcripts, ~84% (167/198) were indeed expressed in either endoderm cells, PGCs, or both (Table S4). We therefore, considered those transcripts to be maternal, and the remaining transcripts were considered to be zygotically expressed. Only ~1.5% (3/198) of maternal mRNAs were enriched in endoderm cells compared with PGCs (blue), whereas ~45% (90/198) were expressed in both PGCs and endoderm cells, but not differentially expressed (white) (Fig. 2C; Table S4). However, over a third of the maternal transcripts (75/198) were enriched in PGCs compared with endoderm cells (red) (Fig. 2C; Table S4), which suggests that many of the annotated maternal, vegetally enriched RNAs are involved in germline specification. Whether they are essential will require further functional testing. Furthermore, although most PGC-enriched transcripts were zygotic, relative to germ plasm expression (i.e. not enriched at the vegetal pole of stage VI oocytes) (1760/1865), about half of the 50 most highly expressed PGC-enriched transcripts were maternal (vegetally enriched transcripts) (Table 1, gray).

Expression of PGC-enriched mRNAs during development

We next utilized whole-mount *in situ* hybridization (WISH) analysis to confirm mRNA expression in PGCs for a set of 15 highly expressed, PGC-enriched transcripts. Expression of selected transcripts was analyzed at gastrula (st. 11.5), neurula (st. 16) and tailbud (st. 33/34) stages. *pgat*, a PGC-specific transcript, was used as a positive control to reference expression in PGCs at each stage. Consistent with our RNA-seq data, all of the 15 transcripts tested (*fer*, *pphln1*, *prpsap2*, *cdc20*, *fam168b*, *lmbn3*, *nasp*, *rtn3*, *pgam1*, *tspan1*, *ppp1r2*, *pgat*, *h1foo*, *zfyve26* and *impad1*), were expressed in PGCs at neurula (12/15 are shown in Fig. 3), and all but four transcripts (*pphln1*, *cdc20*, *h1foo* and *impad1*) were also expressed at st. 11.5. However, of the 11 transcripts expressed at st. 11.5, only four (*fam168b*, *pgam1*, *ppp1r2* and *zfyve26*) were vegetally enriched, maternal transcripts (Owens et al., 2017). These data suggest that PGCs may have initiated zygotic transcription by st. 11.5. Furthermore, only *pgat* was still expressed, as detected by WISH, in PGCs at tailbud stages. At tailbud stages, all PGC-enriched transcripts tested, except *pgat*, were expressed only in somatic tissues. Interestingly, the most represented expression pattern at tailbud was in neural regions, including the eye, cranial ganglia, brain and neural tube (Fig. 3). This observation is consistent with what was previously found (Owens et al., 2017), and could suggest that these transcripts play dual roles in germline and neural specification during development.

sox7 knockdown or overexpression increased PGC number in early development

Historically, mice have been used as the primary model system to study human PGC development (Hayashi et al., 2007; Ohinata et al., 2009, 2005; Saitou et al., 2002; Saitou and Yamaji, 2012). Mouse studies have revealed specific gene sets required to specify the germ cell lineage, including the essential pluripotency genes: *NANOG*, *POU5F1* and *SOX2* (Tang et al., 2016). However, contrary to what has been predicted for human PGC development (based on mouse studies), Irie et al. (2015) identified the F-sox family member *SOX17*, as a primary regulator of human primordial germ cell-like

fate, not *SOX2* (a SOXB1 family member). Interestingly, although several *sox* transcripts were enriched in PGCs (namely *sox2*, *sox3*, *sox4*, *sox7* and *sox8*), *sox7*, another F-sox family member, is highly enriched (74-fold versus endoderm) and the most highly expressed [269 counts per million (CPM)] *sox* transcript in *Xenopus* PGCs (Table S1). Furthermore, *sox7* was recently shown to be expressed in germ plasm through gastrulation, and both morpholino (MO)-mediated *sox7* knockdown and overexpression of *sox7* in the whole embryo resulted in significantly reduced PGCs at tailbud stages (Owens et al., 2017). Therefore, we hypothesized that *sox7* is necessary for proper PGC development.

To test our hypothesis, *sox7* MO-mediated knockdown and overexpression of *sox7*-FL mRNA was targeted to PGCs as described in Materials and Methods (Fig. S2). Interestingly, PGC-directed *sox7* knockdown and overexpression resulted in significantly more PGCs during gastrulation (st. 11.5) (Fig. 4A, B). These effects were significantly rescued by co(PGC-directed)-injection of a *sox7* mRNA construct that cannot bind the *sox7*-MO (*sox7*-MO rescue; Fig. S3) (Fig. 4A,B), thereby confirming specificity of the observed phenotype. However, by tailbud stages (st. 33/34), normal PGC numbers were detected in PGC-directed *sox7* knockdown embryos whereas PGC-directed *sox7*-overexpressing embryos had significantly fewer PGCs (Fig. 4C, D). The observed increase in PGC number at gastrulation was both remarkable and unexpected as the amount of germ plasm assembled during oogenesis remains constant during development. One possible explanation for more germ plasm-bearing cells after *sox7* misexpression might be the premature migration of germ plasm to a perinuclear position, resulting in both daughter cells, rather than only one cell, receiving germ plasm.

Misexpression of *sox7* alters germ plasm localization

During *Xenopus* PGC development, germ plasm at blastula stages is localized in close apposition to the plasma membrane of pPGCs. Subsequent asymmetric divisions result in only one of the two daughter cells inheriting germ plasm (Aguero et al., 2017). At gastrula, the germ plasm in these cells moves to a perinuclear location, establishing them as PGCs. PGCs will subsequently undergo symmetric divisions at specific times during development. Although the timing for germ plasm migration has been described previously in *ex vivo* studies (Yamaguchi et al., 2013), it is currently unknown precisely when the germ plasm becomes perinuclear *in vivo*. To determine whether germ plasm migration is affected when *sox7* expression is altered in PGCs, we first set out to determine normal germ plasm patterning *in vivo* from late blastula to gastrula (st. 8-11).

Immunodetection of the nuclear protein PCNA was used to identify cell nuclei and, when overexposed, the outline of individual cells. Co-immunodetection of Piwi protein, a specific germ plasm marker, was used to identify germ plasm location in PGCs *in vivo* (Lau et al., 2009). We identified three distinct germ plasm phenotypes: membrane (germ plasm located along the plasma membrane), perinuclear and dispersed (germ plasm found in various regions of the PGC). The dispersed phenotype was rare (2.8-14.4%) at all stages analyzed (Fig. 5). Similar to what Taguchi et al. (2012) reported in their *ex vivo* studies, germ plasm was located at the membrane in 93.4% (99/106) of pPGCs at stage 8, and 11.6% (22/189) at stage 10. However, at stage 9 Taguchi et al. (2012) reported only 55% of PGCs with membrane-localized germ plasm, whereas we observed that the majority of PGCs (81.9%, 172/210) still exhibited membrane-localized germ plasm at stage 9. Taken together, these results indicate that germ plasm

Table 1. Maternal versus zygotic PGC-enriched transcripts

Top 50 PGC-enriched transcripts				
ID v9.1 (m)	Gene symbol	PGC average (CPM)	Endo average (CPM)	FC (PGC versus Endo)
v7.1 (m.g)				
Xelaev15011470m.g	<i>pgat</i>	89775	564	159
Xelaev18035291m	<i>ddx25.L</i>	71752	385	186
Xelaev18023956m	<i>grip2.L*</i>	38292	130	295
Xelaev18031029m	<i>dazl.L</i>	16463	82	202
Xelaev18034293m	<i>nanos1.L</i>	6457	30	216
Xelaev18006177m	<i>tex15.L*</i>	5499	53	104
Xelaev18031232m	<i>trank1.L</i>	3728	80	46
Xelaev15043509m.g	<i>trim36</i>	3005	67	45
Xelaev18017827m	<i>mcm10.L</i>	2998	810	4
Xelaev18047493m	<i>naga.S*</i>	2111	642	3
Xelaev15038666m.g	<i>acsbg2</i>	2088	379	6
Xelaev18025369m	<i>ptbp2.S</i>	1766	456	4
Xelaev18028249m	<i>kif13b.L*</i>	1648	189	9
Xelaev18010315m	<i>slc12a3.2.S</i>	1585	341	5
Xelaev15046920m.g	<i>acsl1</i>	1467	294	5
Xelaev15016832m.g	<i>ldlrap1-a</i>	1341	107	13
Xelaev18034461m	<i>bicc1.L*</i>	1169	93	13
Xelaev18017004m	<i>dnd1.L</i>	1051	19	54
Xelaev15069945m.g	<i>plin2</i>	1015	47	22
Xelaev18018420m	<i>pcsk6.L*</i>	971	46	21
Xelaev18046219m	<i>spire1.S</i>	930	11	81
Xelaev18012344m	<i>mov10.L</i>	890	60	15
Xelaev18032097m	<i>sulf1.L</i>	861	155	6
Xelaev18011128m	<i>atp5a1.S</i>	789	122	6
Xelaev18033736m	<i>impad1.S</i>	750	16	47
Xelaev18041802m	<i>hspa5.S</i>	690	193	4
Xelaev18033676m	<i>dsc3.S*</i>	649	161	4
Xelaev18031248m	<i>velo1.L</i>	639	36	18
Xelaev18040158m	<i>zfyve26.L</i>	633	78	8
Xelaev18039847m	<i>fbxo34.L*</i>	626	148	4
Xelaev15002965m.g	<i>alkbh4</i>	580	145	4
Xelaev18023131m	<i>ccnb1.2.L*</i>	562	148	4
Xelaev15070621m.g	<i>lmbn3</i>	536	99	5
Xelaev15071996m.g	<i>vegt-b*</i>	536	17	31
Xelaev18005239m	<i>slc30a9.L</i>	532	126	4
Xelaev18045154m	<i>parn.L*</i>	513	41	13
Xelaev15018513m.g	Unnamed	495	22	22
Xelaev18034686m	<i>pgam1.L*</i>	476	11	45
Xelaev15017463m.g	<i>fam168b</i>	462	94	5
Xelaev18025228m	<i>tspan1.S*</i>	438	23	19
Xelaev18023643m	<i>tob2.L</i>	427	121	4
Xelaev18020909m	<i>pfkfb3.S</i>	425	94	4
Xelaev15038059m.g	<i>slc16a3-b</i>	423	33	13
Xelaev18008022m	<i>fer.L</i>	415	70	6
Xelaev18020738m	<i>cpeb1.S*</i>	368	7	52
Xelaev15019487m.g	<i>nasp</i>	353	36	10
Xelaev18003587m	<i>cdt1.1</i>	351	71	5
Xelaev18031734m	<i>gmn.L</i>	343	102	3
Xelaev15049105m.g	<i>cab39</i>	339	96	4
Xelaev15061337m.g	<i>cdc20</i>	324	58	6

Maternal, vegetally enriched transcripts expressed in PGCs (gray), and zygotic transcripts (white).

CPM, counts per million; Endo, endoderm cell sample; FC, fold change; PGC, primordial germ cell sample.

*Most highly expressed copy is shown.

reorganization from membrane to perinuclear occurs late in stage 9, after the midblastula transition (MBT).

When *sox7* was knocked down or overexpressed, the dispersed germ plasm patterning was significantly more common at all stages tested (Fig. 5). In addition to dispersed germ plasm, significantly more PGCs exhibited perinuclear localized germ plasm prematurely, as early as stage 8 (Fig. 5). Movement of germ plasm to a perinuclear position requires microtubules (Taguchi et al., 2012). Germes (also known as *Loc779566*) is a germ plasm

component that binds to dynein light chains and plays a crucial role in germ plasm translocation from the plasma membrane to the perinuclear region in PGCs (Berekelya et al., 2007; Yamaguchi et al., 2013). In addition to Germes, *Ddx25* also affects germ plasm localization, but in a distinctly different manner, as its timely movement to a perinuclear position is delayed along with cell division (Yamaguchi et al., 2013). Thus, *Sox7* might play a role in germ plasm localization through transcriptional regulation of the genes necessary for this process.

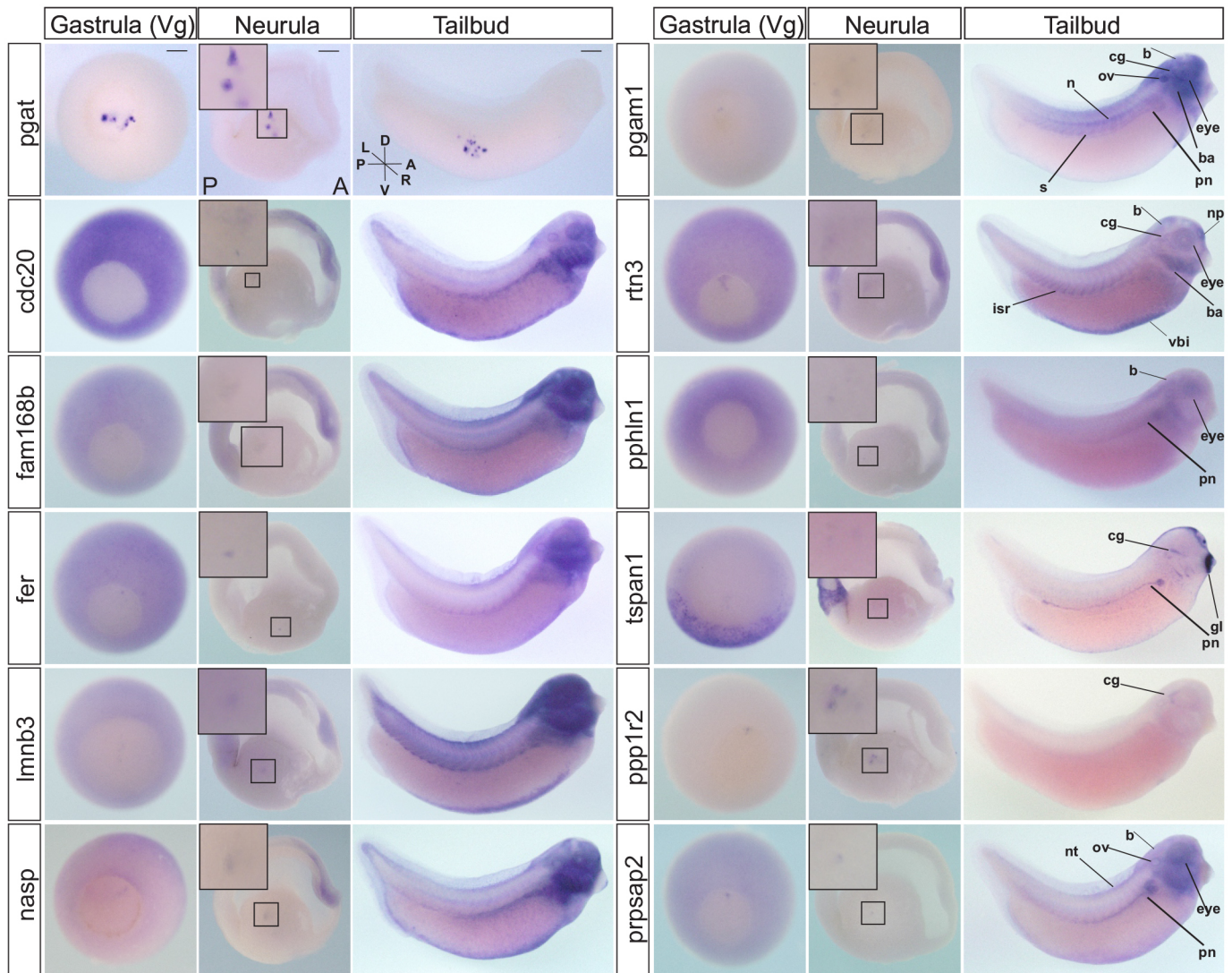


Fig. 3. WISH analysis of PGC transcripts. Expression of selected PGC-enriched mRNAs at post-MBT st. 11.5 (gastrula), 16 (neurula) and 33/34 (tailbud). *pgat* expression marks germ plasm in PGCs. Transcripts detected in PGCs at neurula are shown at higher magnification in insets. WISH analysis was performed on ≥ 21 total embryos from at least three adult female frogs. Insets show magnifications of the boxed areas. A, anterior; b, brain; ba, brachial arches; cg, cranial ganglia; D, dorsal; gl, cement gland; isr, intersegmental region; L, left; n, notochord; np, nasal placodes; nt, neural tube; ov, otic vesicle; P, posterior; pn, pronephros; R, right; s, somites; V, ventral; vbi, ventral blood islands; Vg, vegetal view. Scale bars: 200 μ m.

Altered expression of *sox7* in PGCs results in premature zygotic transcription

PGCs are transcriptionally silent until gastrulation, when they segregate from the endoderm and initiate zygotic transcription (Lai and King, 2013; Venkatarama et al., 2010). The transcription factor(s) within the germ plasm that initiate the PGC-specific gene program remain unknown, but are likely to act after the germ plasm has migrated to a perinuclear position. In this case, premature perinuclear localization would be predicted to result in premature transcriptional activity. To test whether *sox7* plays a role in PGC transcription, embryos with PGC-directed altered expression of *sox7* were collected at st. 10 and 11 and subjected to co-immunofluorescence with antibodies recognizing Piwi (to identify PGCs; Lau et al., 2009) and RNA Pol II CTD-pSer2 (to identify transcriptionally active nuclei) (Bensaude et al., 1999; Bregman et al., 1995; Seydoux and Dunn, 1997). Consistent with our prediction, *sox7* knockdown caused a significant increase in the number of transcriptionally active PGCs at both st. 10 and st. 11

compared with control (Fig. 6). This effect was significantly rescued by co(PGC-directed)-injection of *sox7*-MO with a *sox7* mRNA construct that the *sox7*-MO cannot bind (*sox7*-FL rescue) (Fig. 6), thereby confirming specificity of the observed phenotype. Furthermore, PGC-directed *sox7* overexpression also caused both premature germ plasm migration and a significant increase in the number of transcriptionally active PGCs at st. 10, albeit to a lesser degree than *sox7*-MO; this effect did not persist to st. 11 (Fig. 6). These results show that *sox7* is necessary for proper timing of PGC transcriptional activity.

PGC-directed *sox7* knockdown alters expression of germline genes

Our functional studies showed that both PGC-directed *sox7* knockdown and overexpression caused germ plasm to become dispersed or localize prematurely to the perinuclear region as early as st. 8 (Fig. 5), and caused PGCs to undergo premature transcription (Fig. 6). Thus, we hypothesized that Sox7 might

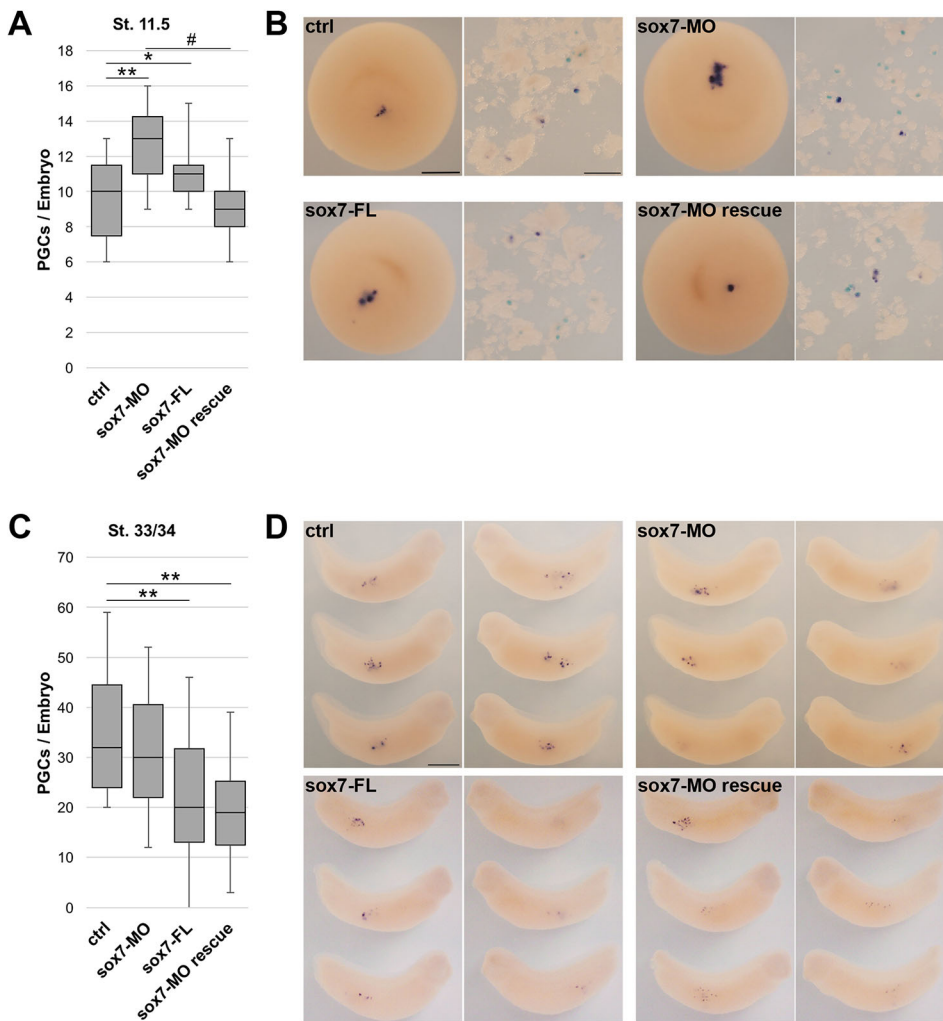


Fig. 4. PGC number increased after PGC-directed *sox7* knockdown or overexpression in early development.

(A-D) The number of PGCs per embryo was determined in control, *sox7*-MO, *sox7*-FL, and *sox7*-MO rescue embryos at gastrula (st. 11.5) (A,B) and tailbud (st. 33-34) (C,D) stages. Representative images are shown (B,D). In box and whisker plots, middle bar is the median, top of the box is the third quartile, bottom of the box is the first quartile, and the whiskers represent the minimum and maximum number of PGCs observed in one embryo of the set analyzed (i.e. the range). ctrl $n=23$, *sox7*-MO $n=24$, *sox7*-FL $n=23$, *sox7*-MO rescue $n=25$. * $P<0.05$, ** $P<0.005$ (compared with uninjected control); # $P<0.05$ (compared with *sox7*-MO). *pgat* expression by WISH was used to identify PGCs. Analysis shown for one of at least three independent experiments. ctrl, control; FL, full-length, flag-tagged; MO, morpholino. Scale bars: 500 μm (B); 1 mm (D).

regulate transcription of germline genes in PGCs. As there are over 1800 transcripts enriched in PGCs (Table S1), we first narrowed down possible candidates by selecting genes for which expression is both restricted to the germline and which were identified in preliminary chromatin immunoprecipitation (ChIP) experiments. The preliminary ChIP-seq data was kindly provided by Dr Ken Cho (University of California, Irvine, CA, USA; unpublished). In their study, ChIP-seq was performed on st. 8 *Xenopus tropicalis* embryos using an anti-Sox7 antibody. In addition to genes for which expression is restricted to the germline, we also looked at the human *POU5F1* homolog *pou5f3.3*, because it is a primary regulator of human PGC specification and is highly enriched in *Xenopus* PGCs. The preliminary ChIP-seq results revealed that of our selected candidates, *germes*, *dazl*, *grip2*, *ddx4* (also known as *vasa*), *dnd1* and *pou5f3.3* all had one or more peaks, indicating that Sox7 directly binds to these genes *in vivo*.

Because the observed effects of *sox7* misexpression we observed occurred as early as st. 8 (Fig. 5), and 99.6% of *Xenopus tropicalis* and *laevis* orthologous genes are syntenic between the two species (Riadi et al., 2016), we investigated whether PGC-directed *sox7* knockdown or overexpression alters transcription of the identified germline genes in st. 8, 9, 10 and 11 embryos by RT-qPCR. Interestingly, *sox7* knockdown caused a significant increase in *grip2* expression compared with scramble-MO control, and a decrease in *germes* expression as early as st. 8 (Fig. 7A). These data

indicate that *sox7* might play a role in germ plasm localization by directly regulating *germes* expression, and negatively regulating expression of *grip2* during early PGC development. Surprisingly, we saw no significant effects on expression of the germline genes when *sox7* was overexpressed (Fig. 7B). These data suggest that the increased PGC number observed in *sox7*-overexpressing embryos (Fig. 4A,B) might not be due to altered transcription of germline genes. Alternative explanations that could account for the observed phenotype are proposed in the Discussion section below.

Unlike *germes*, *dazl*, *grip2*, *ddx4* and *dnd1*, *pou5f3.3* expression is not restricted to the germline in st. 8, 9, 10 and 11 embryos. Therefore, although we did not observe a significant change in *pou5f3.3* expression in PGC-directed *sox7* knockdown or overexpressing embryos, this could be because any effect specific to PGCs was overshadowed by *pou5f3.3* expression in the rest of the embryo. Therefore, to determine whether *pou5f3.3* plays a role in PGC development, *pou5f3.3* MO-mediated knockdown and overexpression of *pou5f3.3*-FL mRNA (Fig. S4) was targeted to PGCs as described in Materials and Methods (Fig. S2). Interestingly, PGC-directed *pou5f3.3* knockdown resulted in significantly more PGCs, in five out of eight independent experiments, during gastrulation (st. 11.5; Fig. S5A,B); however, normal PGC numbers were recovered by tailbud stages (st. 33/34; Fig. S5C,D). These data suggest that *pou5f3.3* might play a role in early PGC development, consistent with what has been shown for

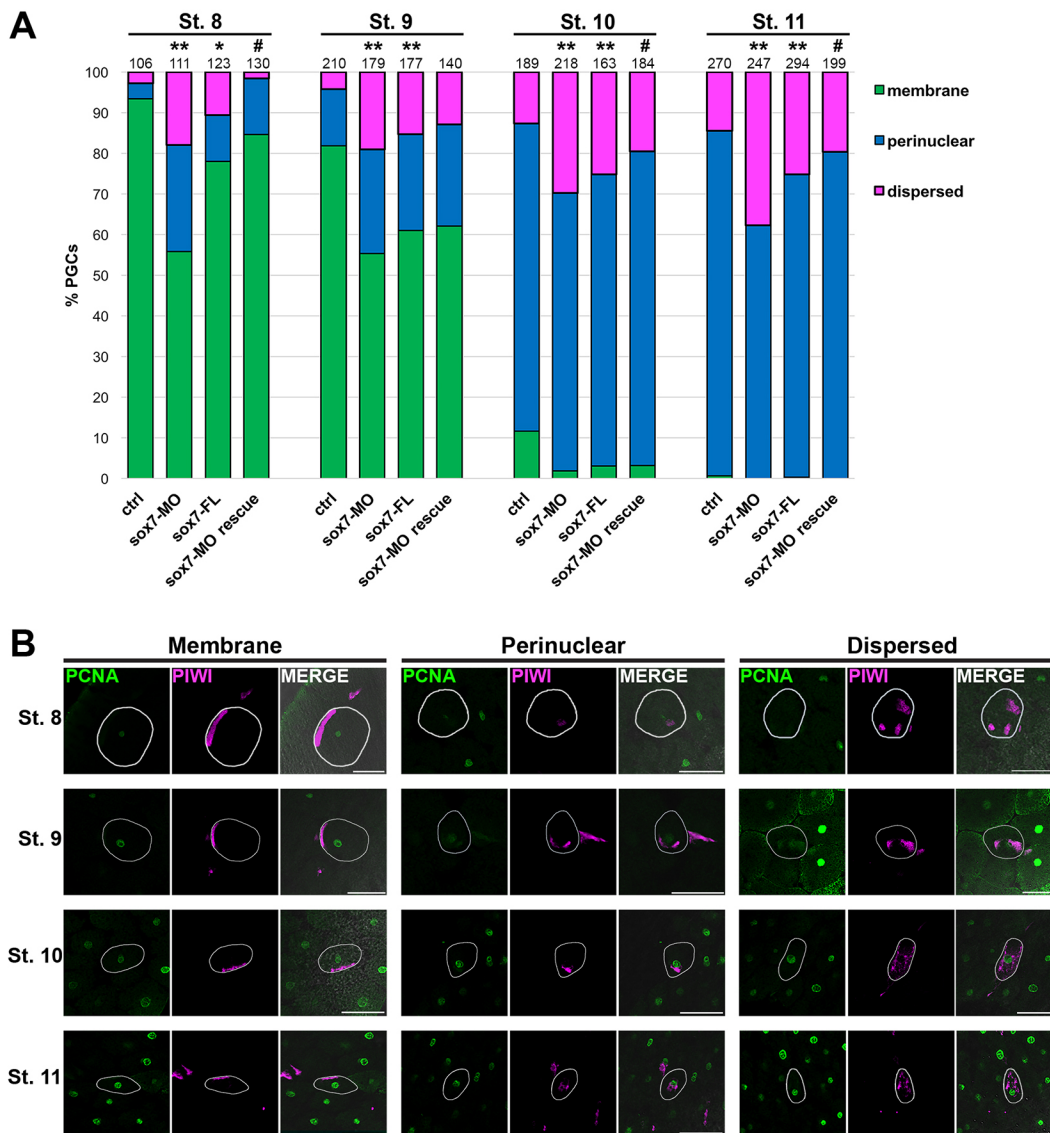


Fig. 5. PGC-directed *sox7* knockdown or overexpression altered germ plasm localization in early PGC development. Control, *sox7*-MO, *sox7*-FL, and *sox7*-MO rescue embryos were analyzed for Piwi (germ plasm; pink) and PCNA (nuclear; green) expression by immunofluorescence at the indicated stages. (A) Germ plasm localization in each PGC was classified as membrane, nuclear or dispersed and quantified; *n* numbers of PGCs analyzed are indicated. (B) Representative images of each classification are shown for all stages analyzed. Representative PGC in each panel is outlined in white. * $P < 0.05$, ** $P < 0.005$ (compared with ctrl); # $P < 0.05$ (compared with *sox7*-MO). ctrl, control; FL, full length, flag-tagged; MO, morpholino. All images were taken at 20 \times . Scale bars: 100 μ m.

POU5F1 in humans. Further functional studies are necessary to determine the exact role of *pou5f3.3* in *Xenopus* PGC development, and whether *Pou5f3.3* works in conjunction with *Sox7* to promote specification of this cell type.

DISCUSSION

In the present study, we identify for the first time the *Xenopus laevis* PGC transcriptome after its segregation from the endoderm lineage. In total, 1865 transcripts were enriched in PGCs (versus embryo- and stage-matched endoderm cells), and six of the top 50 PGC-enriched transcripts have never been shown to play a role in the germline. Cell cycle, cell division, and reproduction are the three most highly represented molecular functions in the top 300 PGC-enriched transcripts, and 53 of the 150 most highly expressed PGC-enriched transcripts form a direct interaction network. Interestingly, 84% of maternal, vegetally enriched transcripts are zygotically

expressed, and of those, ~38% were enriched in PGCs. Similar to humans, both *pou5f3.3* (*POU5F1* homolog) and an F-sox family member (*sox7* in *Xenopus* and *SOX17* in humans) are required for proper PGC development. Functional studies show that *sox7* is necessary for proper germ plasm localization and for timing of PGC transcription during early development of the *Xenopus* germline. Finally, 80% of *Xenopus laevis* PGC-enriched genes are expressed in human PGCs and/or hPGCLCs.

Identification of the zygotic PGC transcriptome

The majority (88%) of the 50 most highly expressed *Xenopus* PGC-enriched transcripts are known germline components, which confirms the specificity of our PGC isolation (Fig. 1C). Additionally, we discovered six (of the top 50) PGC-enriched transcripts that have not been previously shown to play a role in the germline: *impad1*, *cab39*, *dsc3*, *fam168b*, *tspan1* and *fbxo34*. These

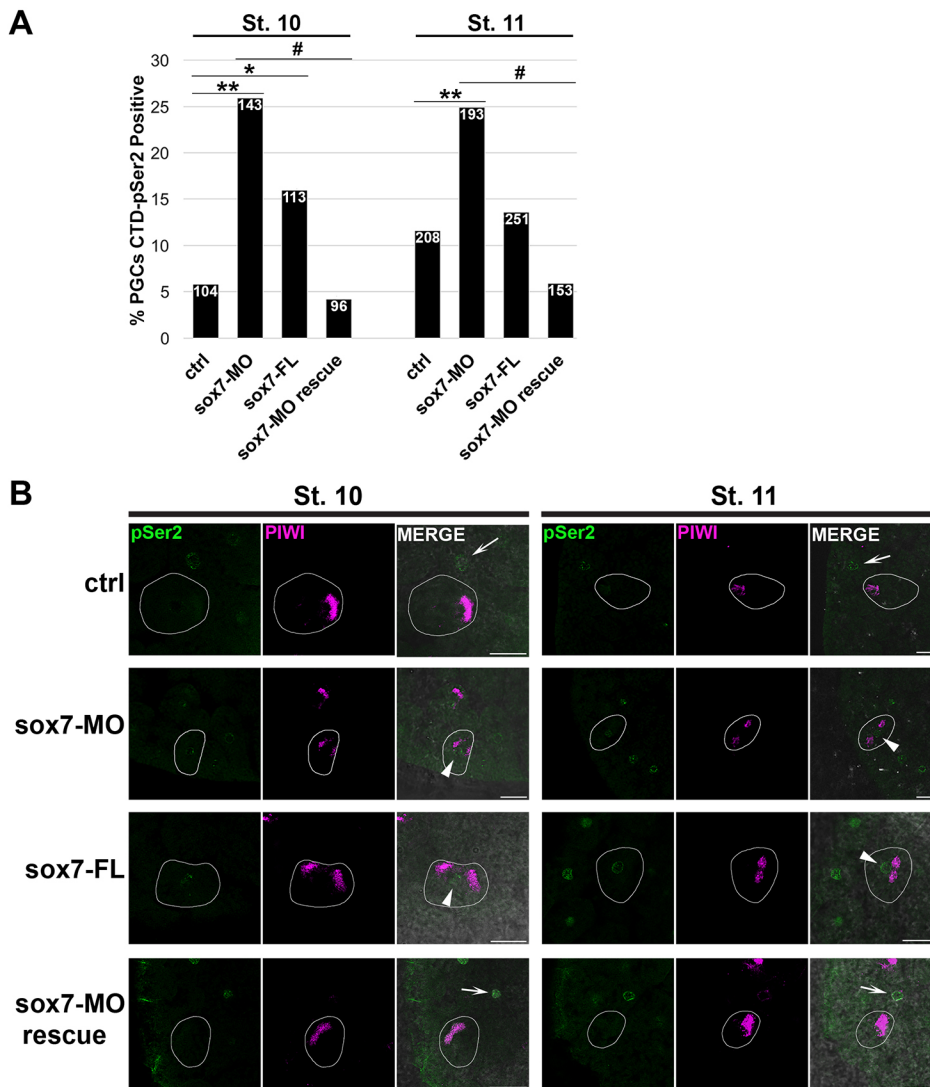


Fig. 6. PGC-directed *sox7* knockdown or overexpression increased the number of transcriptionally active PGCs in early development. Control, *sox7*-MO, *sox7*-FL, and *sox7*-MO rescue embryos were analyzed for Piwi (germ plasm; pink) and pSer2 (transcriptionally active nuclei; green) expression by immunofluorescence at the indicated stages. (A) The percentage of pSer2-positive PGCs was determined; *n* numbers of PGCs analyzed are indicated. (B) Representative images of PGCs from control and injected embryos are shown for each stage. Representative PGC in each panel is outlined in white; arrowheads indicate pSer2-positive PGCs; arrows indicate pSer2-positive somatic cell nuclei. * $P < 0.05$, ** $P < 0.005$ (compared with ctrl); # $P < 0.005$ (compared with *sox7*-MO). ctrl, control; MO, morpholino; FL, full length, flag-tagged; pSer2, RNA polymerase II carboxy-terminal-domain phosphorylated at Ser2. All images were taken at 20 \times . Scale bars: 50 μ m.

genes have, however, been shown to play roles in cellular processes necessary for PGC development. *impad1* is involved in inositol phosphate and sulfur metabolism, and might be necessary for maintaining pluripotency, which is required for proper PGC development (Fagerberg et al., 2014). Furthermore, *cab39* (Guo et al., 2017; Zhao et al., 2017), *dsc3* (Getsios et al., 2004) and *fam168b* (also known as *mani*) (Mishra et al., 2011, 2012) have been shown to mediate cell migration, and thus may be enriched in PGCs to prime them for future migration through the endoderm to the presumptive gonads.

Members of the transmembrane 4 superfamily, which includes *tspan1*, have been shown to form complexes with EPCAM to mediate pluripotency reprogramming by upregulation of OCT4 (Chen et al., 2010; Chuang et al., 2012; Jiang et al., 2015; Kanatsu-Shinohara et al., 2011), a well-established pluripotency gene necessary for PGC development. Similarly, members of the F-box protein family, such as Fbxo34, regulate proteins involved in differentiation and cell cycle progression including the pluripotency factor KLF4 and members of the Wnt signaling pathway (Randle and Laman, 2016), respectively. Tight regulation of the Wnt signaling pathway is necessary for proper PGC development (Chawengsaksophak et al., 2012; Kimura et al., 2006; Laird et al., 2011). Therefore, *fbxo34* might be necessary to maintain pluripotency and regulate the cell cycle of PGCs. In fact,

genes involved in cell cycle and cell division are very abundant in PGCs after segregation from the endoderm (Fig. 2B). Although PGCs undergo very few cell divisions prior to arriving in the gonads, of the top 300 PGC-enriched genes that are involved in cell cycle regulation, only about half are negative regulators (www.genecards.org). One possible explanation is that, similar to oogenesis, when products are made and stored for future use in the developing embryo, PGCs may make and store the products that they will need after they enter the gonads and undergo an intense period of proliferation. Functional studies are necessary to address this hypothesis and to deduce the specific roles of novel germline transcripts in PGCs.

About 10% of PGC-enriched genes are TFs and thus likely to be involved in PGC specification. Surprisingly, the gene encoding the maternal, vegetally enriched T-box transcription factor VegT, which is known to initiate mesoendodermal differentiation (Xanthos et al., 2001; Zhang et al., 1998; Zhang and King, 1996), is among the PGC-enriched TFs (Table S1). Heasman et al. (2001) have shown that *vegt* mRNA is necessary to anchor late pathway mRNAs to the vegetal cortex of the fully grown *Xenopus* oocyte. We have recently identified *sox7* as a late pathway mRNA (Owens et al., 2017), and in the present study show that *sox7* is necessary for proper PGC development (Figs 4-6). Thus, *vegt* RNA may be necessary to anchor *sox7* to the vegetal cortex and thereby concentrate it within

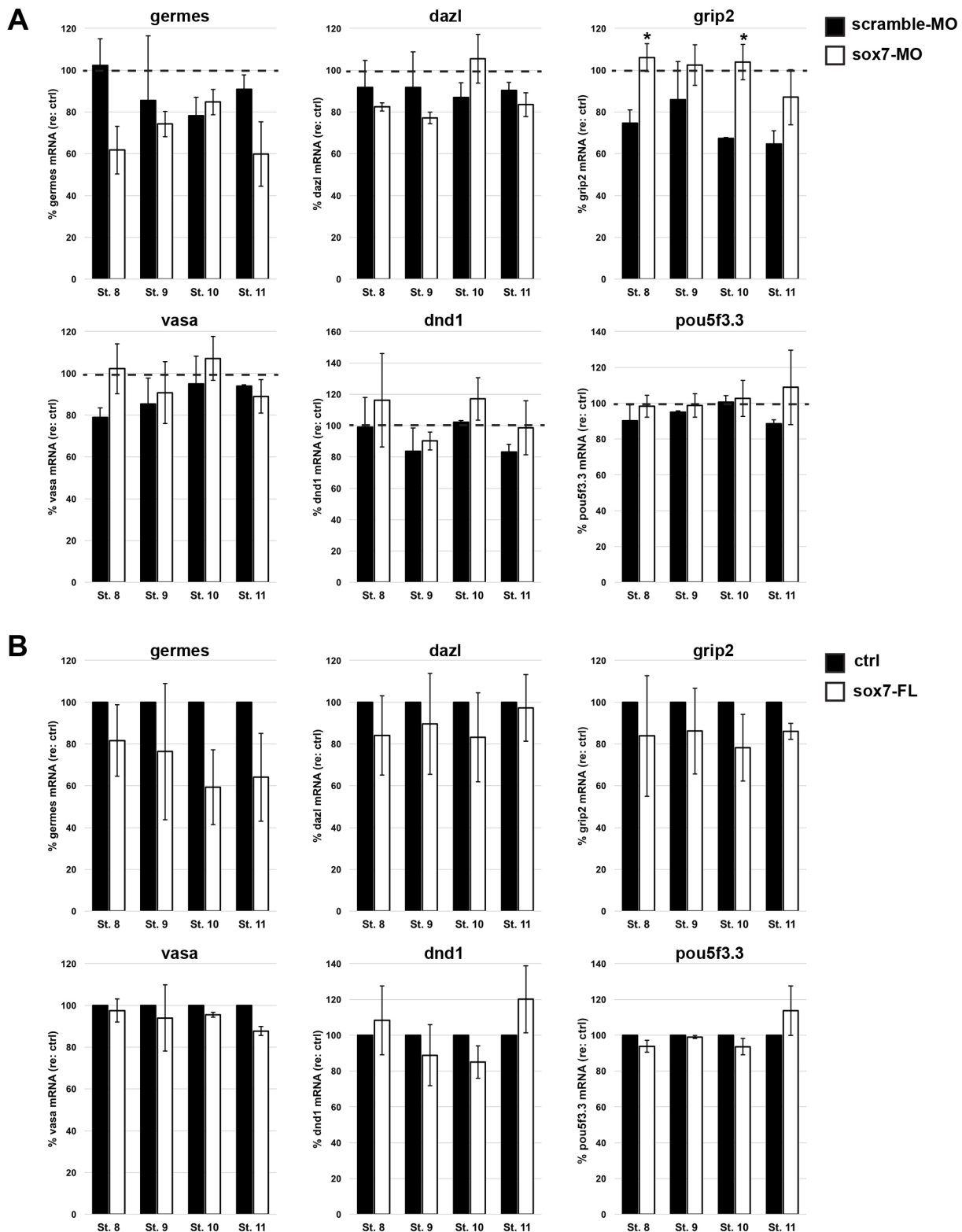


Fig. 7. PGC-directed *sox7* knockdown, but not overexpression, alters expression of germline genes. (A,B) RT-qPCR of expression of germline genes in PGC-directed scramble-MO control and *sox7* knockdown embryos (*sox7*-MO) relative to uninjected control embryos (dashed line) (A) and uninjected control (ctrl) and PGC-directed *sox7* overexpression (*sox7*-FL) embryos (B) at st. 8, 9, 10 and 11. * $P < 0.05$ (compared with relative control). *odc1* expression was used as the reference gene. Error bars represent s.d. of at least two experiments.

the germ plasm during oogenesis. If this is true, and maternal *sox7* mRNA persists in the germ plasm of PGCs until st. 12.5-14, then, presumably, maternal *vegt* mRNA might also persist longer in PGCs

than in surrounding somatic cells. To address this question, we took advantage of the fact that maternal (Zhang and King, 1996) and zygotic (Stennard et al., 1996) *vegt* are slightly different splice

isoforms (Stennard et al., 1999). We further analyzed the reads from our PGC and endoderm cell samples to determine whether the maternal, zygotic, or both *vegt* isoforms were expressed. Consistent with our hypothesis, maternal and not zygotic *vegt* is expressed in PGCs (Fig. S6). However, we were unable to detect distinct maternal or zygotic isoforms in our endoderm samples based on alignment to the unique regions of these two isoforms (Fig. S6A). Because *vegt* is not highly expressed in the endoderm samples (17 CPM, Table S1), deeper RNA sequencing would be required to decipher what *vegt* isoform(s) are expressed in endoderm cells. Additionally, Zhang et al. (2005) revealed that VegT directly regulates *sox7* expression, which induces the *nodal*-related genes *nodal1*, *nodal2*, *nodal*, *nodal5* and *nodal6* (also known as *xnr1*, *xnr2*, *xnr4*, *xnr5* and *xnr6*, respectively) and the pan-endodermal marker *a2m* (also known as *endodermin*), to specify the endoderm. If *sox7* is zygotically re-expressed in PGCs, then perhaps *vegt*, in addition to its role in endoderm specification, is enriched in PGCs to also induce new zygotic germline *sox7* expression.

Interestingly, almost half of the PGC-enriched TFs contain homeodomains (Table S1). Historically, homeodomain-containing TFs are known to be required for proper developmental patterning in somatic tissues (Pearson et al., 2005; Philippidou et al., 2012). Their role in PGCs is currently unknown; however, a recent report by Zheng et al. (2015) suggests a new function for a subset of homeobox-containing genes that may be relevant. Using touch receptor neurons, Zheng et al. (2015) revealed that the homeodomain-containing proteins CEH-13/Lab and EGL-5/Abd-B act as ‘guarantors’ to increase the likelihood of *mec-3* transcription by a POU-homeodomain TF (UNC-86), via the same Hox/Pbx binding site. These data lead us to suggest that at least some of the homeodomain TFs enriched in PGCs might act as ‘guarantors’ to ensure differentiation of the germline. It is tempting to speculate that, similar to what has been shown in touch receptor neurons, PGC-specific differentiation factors are activated by Pou5f3.3 (or other POU proteins) in conjunction with homeodomain-containing proteins to promote differentiation in the germline. This idea is also consistent with the hypothesis that, following lineage segregation from the endoderm, PGCs transcribe and store products for later use, after entry into the somatic gonads.

PGC transcriptome is conserved between *Xenopus* and human PGCs

Complete analysis of the *Xenopus laevis* PGC transcriptome showed that ~80% of genes were also expressed in hPGC and/or hPGCLCs, including *nanos*, *pou5f3.3*, *grip2* and *dazl* (Table S3). It is interesting that such a large percentage of genes expressed in *Xenopus* PGCs by late gastrula/early neurula are conserved in human PGCs considering that these cell types are generated through unique mechanisms (inheritance and induction, respectively). We therefore investigated whether transcripts enriched at the vegetal pole of fully mature *Xenopus laevis* oocytes (Owens et al., 2017) are also conserved in mature human oocytes (Yan et al., 2013). Just over half (60%; 119/198) of the transcripts enriched at the vegetal pole of stage VI *Xenopus laevis* oocytes are also expressed in mature human oocytes (Table S5) indicating that there are more differences in gene expression in *Xenopus* and human oocytes compared with PGCs. These data suggest that, although the *Xenopus* and human PGC transcriptomes are well conserved, the unique mechanisms used to generate this cell type could account for differences in gene expression profiles during PGC development.

sox7 is required for PGC development

Typically, manipulation of PGC-specific genes results in fewer PGCs (Horvay et al., 2006; Houston and King, 2000; Lai et al., 2012). However, PGC-directed *sox7* knockdown and overexpression resulted in significantly more germ plasm-bearing cells during gastrulation (Fig. 4A,B). These results can probably be attributed to altered germ plasm localization and premature initiation of PGC transcription, which was also observed in both PGC-directed *sox7* knockdown and overexpressing embryos (Figs 5 and 6). Altered germ plasm location, whether it be dispersed throughout the cell or localized prematurely to the perinuclear region, would cause symmetric (rather than asymmetric) division of pPGCs and thus account for the increased number of PGCs observed in *sox7*-misexpressing embryos. However, except for the slight differences in the level of altered transcription in *sox7*-misexpressing embryos (Fig. 6), it was difficult to distinguish how both knockdown and overexpression of the same gene in PGCs generated similar phenotypes.

To begin to shed light on potential mechanisms that could explain these similarities, we investigated expression of known germline genes that Sox7 was shown (in preliminary ChIP-seq studies) to bind directly. Interestingly, no differences were observed in PGC-directed *sox7*-overexpressing embryos, but PGC-directed *sox7* knockdown significantly increased *grip2* expression but decreased expression of *germes* (Fig. 7). These results indicate a dual role for Sox7 in PGCs as both a negative and positive regulator of transcription, both of which have been previously reported for Sox7 (Charney et al., 2017; Ko et al., 2017; Wang et al., 2014; Futaki et al., 2004; Séguin et al., 2008; Zhang et al., 2005).

As mentioned previously, *germes* plays an essential role in microtubule-based germ plasm translocation from the plasma membrane to a perinuclear location (Yamaguchi et al., 2013). Similarly, Grip2 has been shown to play a crucial role in microtubule alignment and bundling at the vegetal cortex in zebrafish (Ge et al., 2014). Interestingly, mutation of the zebrafish gene *grip2* (also known as *hecate*) causes defects in the asymmetric movement of *wnt8a* mRNA, and disrupts anchoring of *sybu*, two mRNAs enriched in PGCs (Table S1). Furthermore, *hecate* mutants also have altered PGC number and location (Ge et al., 2014). Therefore, the effect on germ plasm location in PGC-directed *sox7* knockdown embryos is likely to be due to altered expression of *germes* and *grip2*.

As no differences in expression of PGC genes were observed when *sox7* was overexpressed, the effect of *sox7* overexpression on PGC number might be influenced by an alternative role of *sox7*. Recently, a new paradigm called coupling has emerged that links transcription with RNA stability (reviewed by Haimovich et al., 2013). One possibility is maternal Sox7 protein occupies specific promoters and facilitates the recruitment of RNA stability regulators, ‘imprinting’ respective genes. Therefore, *sox7* overexpression could enhance the stability of germline genes in this manner (rather than directly enhancing or reducing transcription). Whether Sox7 protein can bind and stabilize germ plasm RNAs is an interesting question to be resolved.

Although we focused on the role of *sox7* in early PGC development, it is worth noting that by st. 33/34, PGC numbers for PGC-directed *sox7*-overexpressing embryos were significantly decreased, and *sox7* knockdown embryos were indistinguishable from controls (Fig. 4C,D). For PGC-directed *sox7* morphant embryos this could be explained by the dispersed germ plasm pattern leading to some pPGCs not inheriting enough germ plasm to maintain PGC specification later in development. However, *sox7*

overexpression could be causing ectopic expression of somatic genes such as *sox17* in PGCs (Zhang et al., 2005). Misexpression of *sox17* in PGCs results in apoptosis during early tailbud stages (~st. 28) (Lai et al., 2012), which could account for the reduced number of PGCs observed at mid-tailbud stages (st. 33/34; Fig. 4C,D) in PGC-directed *sox7*-overexpressing embryos. Additionally, because PGC-directed injections also target the surrounding endoderm of PGCs (Fig. S2), an effect on the PGC niche cannot be overlooked.

***pou5f3.3* might play a role in PGC development**

The TF POU5F1 has been well-established as a pluripotency gene that is crucial for proper PGC development in animals from zebrafish to humans (Lacerda et al., 2014; Tang et al., 2016). We previously isolated PGCs and, using semi-quantitative PCR, identified *pou5f3.1* (also known as *oct91*) as a zygotic PGC transcript (Venkatarama et al., 2010). Since then, we modified our PGC isolation procedure to achieve more precise PGC selection (Butler et al., 2017). Using this modified method, we isolated over ten times more PGCs than the original study to perform highly sensitive RNA-sequence analysis. In the current study, all three *POU5F1* homologs, *pou5f3.3*, *pou5f3.2* and *pou5f3.1*, were expressed in PGCs; however, only *pou5f3.3* and *pou5f3.2* were enriched in PGCs compared with embryo- and stage-matched endoderm cells (Table S1). *pou5f3.3* was the predominant POU member in PGCs and, similar to *sox7* knockdown, PGC-directed *pou5f3.3* knockdown resulted in significantly more PGCs in early development (Fig. S5). However, this increase in PGC number was observed in only five out of eight experiments, suggesting that, in contrast to *sox7* knockdown, *pou5f3.3* knockdown alone was not sufficient to cause the phenotype 100% of the time. These results suggest that in the absence of *pou5f3.3*, *pou5f3.2* (and/or *pou5f3.1*) might serve a redundant role in PGCs.

Interestingly, *pou5f3.3* contains octamer and Sox binding motifs within the *pou5f3.3* regulatory (promoter) region, which are crucial for *pou5f3.3* transcription in oocytes (Morichika et al., 2014). Furthermore, although *pou5f3.3* is expressed in the oocyte, it is not enriched at the vegetal pole, where germ plasm and its maternal components are localized (Owens et al., 2017); therefore, we consider *pou5f3.3* to be a zygotic PGC transcript (with respect to germ plasm). Unlike *pou5f3.3*, *sox7* is enriched at the vegetal pole of stage VI oocytes along with germ plasm components, and is thus considered a maternal PGC transcript (Table S4). As *pou5f3.3* is considered a zygotic PGC transcript that contains an octamer-sox binding motif, it is reasonable to suggest that *sox7* initiates zygotic *pou5f3.3* expression, and together they turn on the zygotic PGC transcriptome to specify the germline. It is worth noting that, to date, the maternal TF(s) that activate zygotic transcription in PGCs remain unknown. Although we investigated *sox7* in the present study, *e2f1* is another maternal TF that is enriched in PGCs and represents a major hub in the PGC direct-interaction network (Table S1; Fig. 2A). Further studies are necessary to elucidate the specific roles of *e2f1*, and the *POU5F1* homologs *pou5f3.3*, *pou5f3.2* and *pou5f3.1*, and to investigate their possible regulation by *sox7* in *Xenopus* PGC development and specification.

MATERIALS AND METHODS

Isolation of PGC and endoderm cell samples

Primordial germ cells and stage-matched endoderm cells were isolated from st. 12–14 *Xenopus laevis* embryos and collected into RNA lysis buffer as described by Butler et al. (2017). Total RNA was extracted using the RNAqueous-Micro Total RNA Isolation Kit as per the manufacturer's protocol (Ambion). To acquire enough RNA for processing, RNA was

extracted from ~1040 PGCs and ~2100 endoderm cells isolated from ~245 total embryos that were generated by 28 different females and pooled. The total RNA was quantified and assessed for quality by electrophoresis on an Agilent 2100 Bioanalyzer RNA 6000 Pico Chip (Agilent Technologies).

RNA library preparation and sequencing

Ribosomal RNA (rRNA) was depleted from a total of 300 ng of each RNA sample using the RiboZero Gold (Human/Mouse/Rat) kit (Illumina, MRZG12324). The yield of rRNA-depleted RNA was 6.9 ng for the PGC sample and 10.2 ng for the endoderm cell sample. Three 1-ng aliquots of each rRNA-depleted sample were used as template for preparation of sequence-able template DNA molecules using the ScriptSeq v2 RNaseq Library Preparation Kit (Illumina, SSV21124). The quality and size distribution of the amplified libraries were determined utilizing an Agilent 2100 Bioanalyzer High Sensitivity DNA Chip. Libraries were quantified using the KAPA Library Quantification Kit (Kapa Biosystems), and equimolar concentrations were diluted prior to loading onto the flow cell of the Illumina cBot cluster station. The libraries were extended and bridge amplified to create sequence clusters using the Illumina HiSeq PE Cluster Kit v4, and sequenced on an Illumina HiSeq Flow Cell v4 with 100-bp paired-end reads plus index read using the Illumina HiSeq SBS Kit v4. Real-time image analysis and base calling were performed on the instrument using the HiSeq Sequencing Control Software version 2.2.58. Three DNA libraries were prepared from *Xenopus laevis* primordial germ cell RNA and three from matched endoderm cell RNA were sequenced on the HiSeq2500. All samples had a minimum of 46,313,867 passed-filter paired-end reads.

RNA-sequence data analysis

Bowtie 2 v2.2.6 was used to create index reference genomes for *Xenopus laevis* v7.1 and v9.1 (xenbase.org, September, 2015), then the raw reads from PGC and endoderm samples were aligned to each genome version using TopHat2 v2.1.0. Alignment to *X. laevis* v7.1 ranged from 45.9% to 62.5%, and from 45.9% to 62.6% for v9.1. Aligned reads were sorted and counted with Samtools v1.2 and HTSeq v0.6.0, respectively. Raw count data were filtered to remove transcripts with FDR>0.1, and CPM<3 in at least three samples. Filtered counts were TMM normalized and GLMTagwise dispersion was estimated. In total, 7902 transcripts from v7.1 and 10,364 from v9.1 were retained. Data from alignment to v7.1 and v9.1 were merged and normalized generating 13,469 total unique transcripts (Table S1). Note, for transcripts identified in both genome versions, the counts from v9.1 were used. These transcripts were then subjected to principal component analysis (PCA), dispersion, and differential expression analysis using RStudio. EdgeR v3.12.0 was used for differential gene expression analysis. Transcripts were considered to be differentially expressed if the FDR corrected (Benjamini–Hochberg) *P*-value was <0.05. Average CPM from the three PGC samples was calculated and compared with that from the three endoderm cell samples. Transcripts at least threefold more highly expressed in PGCs compared with endoderm cells were considered enriched in PGCs, and those threefold more highly expressed in endoderm cells compared with PGCs were considered enriched in endoderm cells. Using these criteria, 1865 transcripts were enriched in PGCs, and 791 in endoderm cells (Table S1).

Gene ontology, biological process, and pathway analysis

Human homologs of the top 300 PGC-enriched transcripts were submitted to GeneGo for gene ontology and biological process analysis. Additionally, human homologs of the top 150 PGC-enriched transcripts were subject to pathway analysis via GeneGo (MetaCore Bioinformatics software from Thomson Reuters; <https://portal.genego.com>).

Whole-mount *in situ* hybridization (WISH)

X. laevis embryos were obtained as described in Sive et al. (2000). WISH was performed exactly as described by Owens et al. (2017). Plasmids containing full-length clones were purchased from GE Dharmacon (*rtn3*, *pgam1*, *ppp1r2*, *zfyve26*), Transomic Technologies (*fer*, *pphn1*, *cdc20*,

lmbn3, *nasp*, *tspan1*, *h1foo*, *impad1*) and ThermoScientific (*prpsap2*, *fam168b*). Inserts were verified by restriction digestion and sequencing. Inserts were PCR amplified using the following primer sets: T7/Sp6 (*pphln1*, *cde20*, *fer*, *h1foo*, *impad1*, *lmbn3*, *rtn3*, *fam168b*, *pgam1*, *zfyve26*, *prpsap2*) or T3/Sp6 (*ppp1r2*, *tspan1*, *nasp*). *pgat* clone was synthesized as described by Lai et al. (2012). Antisense probes containing digoxigenin-11-UTP were synthesized using the T7 or T3 RNA polymerase (NEB). Primer sequences: T7, 5'-TAATACGACTACTATAG-3'; T3, 5'-AATTAACCTT CACTAAAG-3'.

PGC-targeted injections

To show the ability to target the germline, FITC-dextran (0.2%) was injected in the germ plasm-bearing blastomeres of 16- to 32-cell embryos as shown in Fig. S2. Injected embryos were incubated in the dark at 18°C, then collected and fixed for WISH at st. 11.5 and 33/34. Embryos were staged according to Nieuwkoop and Faber (1956). WISH was performed as described by Owens et al. (2017). The germ plasm marker *pgat* was used to identify PGCs and FITC as a lineage tracer to identify progeny of injected blastomeres. It is important to note that although this method does restrict the injected material to the PGC region, the surrounding endoderm is also targeted to some degree (Fig. S2).

Functional analysis of *sox7* and *pou5f3.3* in PGCs

Morpholinos targeting the following RNAs were purchased from Gene Tools: *pou5f3.3*-MO (5'-ATTGGTCCATTTCAGCACTTGTC), and *sox7*-MO (Owens et al., 2017). To test *pou5f3.3*-MO efficiency, flag-tagged full-length *pou5f3.3* (*pou5f3.3*-FL) was synthesized and purchased from Genewiz. *sox7*-MO efficiency was previously determined using flag-tagged full-length *sox7* (*sox7*-FL) (Owens et al., 2017). *pou5f3.3*-FL-rescue and *sox7*-FL-rescue were generated from *pou5f3.3*-FL and *sox7*-FL, respectively, by introducing conservative mutations in the regions that bind the morpholinos (*pou5f3.3*-FL-rescue, 5'-GATcAGGTtCT aGagATGGATcAGt; *sox7*-FL-rescue, 5'-TaAGtAAaTCCgTGcAGcATcAT GAC), rendering the respective morpholinos ineffective (Fig. S4C and Fig. S3, respectively). Mutations were introduced using the Q5 Site-Directed Mutagenesis Kit (New England Biosciences, E0554) as per the manufacturer's protocol, with the following primer sets: *pou5f3.3*-FL-rescue-forward, 5'-agatggatcagTCCATATTGTACAGCC AAAG; *pou5f3.3*-FL-rescue-reverse: 5'-ctagaacctgaTCAGGGAATTCC AATCGATG; *sox7*-FL-rescue-forward, 5'-gtggcatcATGACTACCC TGATGGGATC; *sox7*-FL-rescue-reverse: 5'-ggattacttATGATCGATGGG ATCCTG.

pou5f3.3-MO was tested for knockdown efficiency, and the effectiveness of the *pou5f3.3*-FL rescue and *sox7*-FL rescue constructs in the presence of *pou5f3.3*-MO and *sox7*-MO, respectively, was tested using the Wheat Germ Extract Kit (Promega, L4380) as per the manufacturer's protocol. Translation of *sox7*-FL rescue, *pou5f3.3*-FL and *pou5f3.3*-FL rescue were detected by Flag immunoprecipitation followed by western blot analysis for Flag (Figs S3 and S4, respectively). A histogram was generated by first comparing each band intensity to respective loading controls using ImageJ, then calculating percentage inhibition relative to either *sox7*-FL rescue, *pou5f3.3*-FL or *pou5f3.3*-FL rescue expression in the absence of the respective MO. Primary antibody was monoclonal mouse anti-Flag (Sigma, F1804); secondary antibody was HRP-conjugated anti-mouse IgG (Promega, W402B).

Synthetic capped mRNAs for microinjection were obtained by *in vitro* transcription using the mMACHINE SP6 Kit (Ambion). DNA templates were generated via PCR using SP6/T3 primers, then transcribed as per the manufacturer's protocol. PGC-targeted injections were performed on 16- to 32-cell embryos with the following: *pou5f3.3*-MO (16 ng total); *pou5f3.3*-FL (200 pg); *pou5f3.3*-MO (16 ng total) and *pou5f3.3*-FL-rescue (200 pg) together; *sox7*-MO (16 ng total); *sox7*-FL (150 pg); *sox7*-MO (16 ng total) and *sox7*-FL-rescue (150 pg) together. Injected embryos were collected and fixed at st. 11.5 and 33/34 (staging according to Nieuwkoop and Faber, 1956), and PGCs were identified using WISH against *pgat* as previously described (Owens et al., 2017). PGC number per embryo was calculated by manually counting PGCs in each embryo. The yolk plugs along with the endodermal cores, which contain the

PGCs, were removed from st. 11.5 embryos and dissected into individual cells to ensure all PGCs were counted. Note: PGCs on the surface of the yolk plug stained purple, and internal PGCs stained bluish-purple. All results shown are representative of at least three independent experiments. The $P(t)$ -values were determined using a two-tailed unpaired Student's t -test. P -values <0.05 were considered significant.

Sample preparation for confocal analysis

These methods were modified from Lai et al. (2012). Staged embryos were fixed in Dent's fixative for at least 24 h at -20°C , then re-hydrated with 75%, 50% and 25% methanol in PBS-tween (PBST) for 5 min each at room temperature (RT) followed by 3×10 min washes with PBST at RT. Next, embryos were bleached in $0.5 \times$ standard saline citrate (SSC) solution with 5% formamide and 1% H_2O_2 for 1-2 h under fluorescent light. Embryos were then washed twice (10 min each wash) in PBST then twice (30 min each wash) in 0.3% Triton X-100 in PBS at RT. Embryos were blocked in blocking solution (2% bovine serum albumin, 0.15% Triton X-100 in PBS) for 15 min followed by blocking solution with 10% goat serum for 30 min at RT with rocking. Embryos were incubated with primary antibody in blocking solution overnight at 4°C . The following day, embryos were washed four times (1 h each wash) in PBST then incubated with Alexa Fluor-labeled secondary antibody in blocking solution overnight at 4°C in the dark. The following day, embryos were washed four times (1 h each wash) in PBST. Embryos were analyzed in Murray's Clear (2:1 benzylbenzoate: benzylalcohol) using an inverted Zeiss LSM-510 confocal laser scanning microscope equipped with argon ion, helium-neon and green-neon lasers after dehydration with ethanol.

Germ plasm analysis

Uninjected (control) embryos and embryos injected with *sox7*-MO (16 ng total), *sox7*-FL (150 pg), or *sox7*-MO (16 ng total) and *sox7*-FL-rescue (150 pg) together were collected at st. 8, 9, 10 and 11, then fixed and processed for confocal analysis. The following primary antibodies were used: rabbit anti-Piwi1 (1:1000; a gift from Dr Nelson Lau; Lau et al., 2009) and mouse anti-PCNA Clone 10 (1:200; Sigma, P8825). The following secondary antibodies were used: Alexa Fluor 488 goat anti-mouse IgG (H +L) (1:500; Molecular Probes, A10680) and Alexa Fluor 633 goat anti-rabbit IgG (H+L) (1:500; Molecular Probes, A21071). PCNA fluorescence was over-saturated to identify cell nuclei and to outline individual cells. Each embryo was scanned for Piwi expression to identify PGCs. Once PGCs were identified, the merged image was used to visualize the location of germ plasm (detected by Piwi immunofluorescence) in each individual PGC relative to cell membrane and nuclei (detected by PCNA immunofluorescence). Germ plasm location was classified as either membrane, perinuclear or dispersed. Over 100 PGCs were assessed in control and injected embryos at st. 8, 9, 10 and 11; embryos used were generated from four different female frogs. Germ plasm location in PGCs from control embryos was compared with that in injected embryos, and $P(t)$ -values were determined using a two-tailed unpaired Student's t -test. P -values <0.05 were considered significant.

PGC transcription analysis

Uninjected (control) embryos and embryos injected with *sox7*-MO (16 ng total), *sox7*-FL (150 pg), or *sox7*-MO (16 ng total) and *sox7*-FL-rescue (150 pg) together were collected at st. 10 and 11, then fixed and processed for confocal analysis. The following primary antibodies were used: rabbit anti-Piwi1 (1:1000; a gift from Dr Nelson Lau; Lau et al., 2009) and mouse anti-RNA polymerase II carboxy-terminal-domain phosphorylated at Ser2 (CTD-pSer2) (1:300; Biologend, 920204). The following secondary antibodies were used: Alexa Fluor 488 goat anti-mouse IgG (H+L) (1:500; Molecular Probes, A10680) and Alexa Fluor 633 goat anti-rabbit IgG (H +L) (1:500; Molecular Probes, A21071). Each embryo was scanned for Piwi and CTD-pSer2 expression to identify PGCs and transcriptionally active cells, respectively. Once PGCs were identified, the merged image was used to determine transcriptionally active PGCs. A total of 96-143 PGCs were assessed at st. 10, and 153-251 PGCs at st. 11, from embryos generated from three different female frogs. The percentages of transcriptionally active PGCs from control embryos were compared with

those of injected embryos, and *P*(*t*)-values were determined using a two-tailed unpaired Student's *t*-test. *P*-values <0.05 were considered significant.

RNA isolation and real-time quantitative PCR (RT-qPCR)

Total RNA was isolated using Trizol reagent exactly as described by Owens et al. (2017) from ten uninjected control, PGC-directed scramble-MO (5'-CCTCTTACCTCAGTTACAATTTATA), *sox7* knockdown (*sox7*-MO), and *sox7*-overexpressing (*sox7*-FL) embryos at st. 8, 9, 10 and 11. For each sample, total RNA was used to synthesize cDNA via the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) was used according to the manufacturer's protocol for quantitative, real-time PCR (qRT-PCR) analysis of *odc1* (also known as *ode*) (Xanthos, et al., 2001); *germes* (forward: 5'-CAAGATGAACATCAGGAGAGG; reverse: 5'-GCACA GCTTGATAACCAAAGG); *dazl* (forward: 5'-GTTTCAGGCTTGCC CATATCCAAG; reverse: 5'-TTGGATCCATATCACAGCAGTGG); *grip2* (forward: 5'-GACCTGAAACATGTGGACAGTCAG; reverse: 5'-TGTTGCTGCTGATGTGATGGCTCC); *ddx4* (forward: 5'-CATCAA CAAGCATTACGGTG; reverse: 5'-CCAATTCTATGGACGTACTC ATC); *dnd1* (forward: 5'-TGGTAATGCTCCAGTCAGTG; reverse: 5'-TAAGCGAACCTCGATTGAG); and *pou5f3.3* (forward: 5'-GGAAGT GAAGAGGATGGAATG; reverse: 5'-CAGTTGCAGGGACTCAAAGC). qRT-PCR analyses were carried out using 50 ng of cDNA on a Bio-Rad C1000 Thermal Cycler, CFX96 Real-Time System. *germes*, *dazl*, *grip2*, *ddx4*, *dnd1* and *pou5f3.3* expression was normalized to *odc1*. mRNA expression values were calculated as 2-[CT(target)-CT(*odc1*)]. Two (Fig. 7B) or three (Fig. 7A) independent experiments were performed and the average expression of each gene was calculated relative to uninjected control embryos (set to 100%) then plotted. Relative gene expression in *sox7* knockdown embryos was compared with scramble-MO control embryos, and relative gene expression in *sox7*-overexpressing embryos was compared with uninjected control embryos, and *P*(*t*)-values were determined using a two-tailed unpaired Student's *t*-test. *P*-values <0.05 were considered significant.

vegt analysis

The raw reads from PGC and endoderm samples were aligned to specific regions of the maternal (Zhang and King, 1996) and zygotic (Stennard et al., 1996) isoforms of *vegt*. Because the unique region of zygotic *vegt* is only 32 bp, and the size of our sample reads was 100 bp, we extended the zygotic *vegt* reference sequence to also include 70 bp of the *vegt* region that is conserved between maternal and zygotic isoforms (Fig. S6A). The reference for the maternal isoform was generated using the 131 bp unique region plus the same 70 bp conserved region used for the zygotic *vegt* isoform (Fig. S6A). The raw reads from PGC and endoderm samples were also aligned to a *vegt* region that is 100% conserved between maternal and zygotic isoforms (Fig. S6A), which served as a positive control for *vegt* alignment. Reads were aligned to sea urchin *wnt8* and *Xenopus laevis wnt8* to serve as negative and positive controls, respectively, for the alignment assay.

Xenopus and human oocyte transcript comparison

Yan et al. (2013) investigated the transcriptome of fully mature human oocytes using RNA sequencing followed by alignment to two human references (see Yan et al., 2013 for specific details). Using these publicly available data (deposited in Gene Expression Omnibus under accession number GSE60138), we considered genes to be expressed in the human oocyte if at least two of their three samples had a FPKM value ≥ 1 when aligned to either reference genome. After applying these criteria, we generated a list of human oocyte transcripts. Next, we compared the list of human oocyte transcripts to the set of 198 transcripts enriched at the vegetal pole of stage VI *Xenopus laevis* oocytes previously reported (Owens et al., 2017) to determine which transcripts are conserved in humans.

Acknowledgements

We thank Dr Mike Klymkowsky for sharing preliminary data showing *sox7* in germ plasm. We thank Dr Ken Cho for sharing preliminary Sox7 ChIP-seq data on

stage 8 *Xenopus tropicalis* embryos. We thank Dr Nelson Lau for the Piwi antibody and Dr Karen Newman for expert technical assistance.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: A.M.B., D.A.O., M.L.K.; Methodology: A.M.B., D.A.O., L.W., M.L.K.; Software: A.M.B., L.W.; Validation: A.M.B., D.A.O., M.L.K.; Formal analysis: A.M.B., D.A.O., L.W.; Investigation: A.M.B., D.A.O.; Resources: M.L.K.; Data curation: A.M.B., L.W.; Writing - original draft: A.M.B.; Writing - review & editing: A.M.B., D.A.O., L.W., M.L.K.; Visualization: A.M.B., D.A.O., L.W., M.L.K.; Supervision: M.L.K.; Project administration: A.M.B., M.L.K.; Funding acquisition: M.L.K.

Funding

This work was supported by the National Institutes of Health (HD072340 and GM102397 to M.L.K.). L.W. was supported by a National Science Foundation grant (IOS 1257967). Deposited in PMC for release after 12 months.

Data availability

All RNA-seq data are available at Gene Expression Omnibus with accession number GSE102047.

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.155978.supplemental>

References

- Agüero, T., Kassmer, S., Alberio, R., Johnson, A. and King, M. L. (2017). Mechanisms of vertebrate germ cell determination. *Adv. Exp. Med. Biol.* **953**, 383-440.
- Bensaude, O., Bonnet, F., Cassé, C., Dubois, M.-F., Nguyen, V. T. and Palancade, B. (1999). Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD). *Biochem. Cell Biol.* **77**, 249-255.
- Berekelya, L. A., Mikryukov, A. A., Luchinskaya, N. N., Ponomarev, M. B., Woodland, H. R. and Belyavsky, A. V. (2007). The protein encoded by the germ plasm RNA *Germes* associates with dynein light chains and functions in *Xenopus* germline development. *Differentiation* **75**, 546-558.
- Bregman, D. B., Du, L., van der Zee, S. and Warren, S. L. (1995). Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. *J. Cell Biol.* **129**, 287-298.
- Butler, A. M., Agüero, T., Newman, K. M. and King, M. L. (2017). Primordial germ cell isolation from *Xenopus laevis* embryos. *Methods Mol. Biol.* **1463**, 115-124.
- Charney, R. M., Forouzmand, E., Cho, J. S., Cheung, J., Paraiso, K. D., Yasuoka, Y., Takahashi, S., Taira, M., Blitz, I. L., Xie, X. et al. (2017). Foxh1 occupies cis-regulatory modules prior to dynamic transcription factor interactions controlling the mesendoderm gene program. *Dev. Cell* **40**, 595-607.e594.
- Chawengsaksophak, K., Svingen, T., Ng, E. T., Epp, T., Spiller, C. M., Clark, C., Cooper, H. and Koopman, P. (2012). Loss of *Wnt5a* disrupts primordial germ cell migration and male sexual development in mice. *Biol. Reprod.* **86**, 1-12.
- Chen, L., Yuan, D., Zhao, R., Li, H. and Zhu, J. (2010). Suppression of TSPAN1 by RNA interference inhibits proliferation and invasion of colon cancer cells in vitro. *Tumori* **96**, 744-750.
- Chuang, C.-Y., Lin, K.-I., Hsiao, M., Stone, L., Chen, H.-F., Huang, Y.-H., Lin, S.-P., Ho, H.-N. and Kuo, H.-C. (2012). Meiotic competent human germ cell-like cells derived from human embryonic stem cells induced by BMP4/WNT3A signaling and OCT4/EpCAM (epithelial cell adhesion molecule) selection. *J. Biol. Chem.* **287**, 14389-14401.
- Colozza, G. and De Robertis, E. M. (2014). Maternal syntabulin is required for dorsal axis formation and is a germ plasm component in *Xenopus*. *Differentiation* **88**, 17-26.
- Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K. et al. (2014). Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397-406.
- Forristall, C., Pondel, M., Chen, L. and King, M. L. (1995). Patterns of localization and cytoskeletal association of two vegetally localized RNAs, Vg1 and Xcat-2. *Development* **121**, 201-208.
- Frankenberg, S. and Renfree, M. B. (2013). On the origin of POU5F1. *BMC Biol.* **11**, 56.
- Futaki, S., Hayashi, Y., Emoto, T., Weber, C. N. and Sekiguchi, K. (2004). Sox7 plays crucial roles in parietal endoderm differentiation in F9 embryonal carcinoma cells through regulating Gata-4 and Gata-6 expression. *Mol. Cell. Biol.* **24**, 10492-10503.
- Ge, X., Grotjahn, D., Welch, E., Lyman-Gingerich, J., Holguin, C., Dimitrova, E., Abrams, E. W., Gupta, T., Marlow, F. L., Yabe, T. et al. (2014). Hecate/Grip2a

- acts to reorganize the cytoskeleton in the symmetry-breaking event of embryonic axis induction. *PLoS Genet.* **10**, e1004422.
- Getsios, S., Huen, A. C. and Green, K. J.** (2004). Working out the strength and flexibility of desmosomes. *Nat. Rev. Mol. Cell Biol.* **5**, 271-281.
- Guo, R., Gu, J., Zhang, Z., Wang, Y. and Gu, C.** (2017). MiR-451 promotes cell proliferation and metastasis in pancreatic cancer through targeting CAB39. *Biomed. Res. Int.* **2017**, 2381482.
- Haimovich, G., Choder, M., Singer, R. H. and Trcek, T.** (2013). The fate of the messenger is pre-determined: a new model for regulation of gene expression. *Biochim. Biophys. Acta* **1829**, 643-653.
- Hayashi, K., de Sousa Lopes, S. M. C. and Surani, M. A.** (2007). Germ cell specification in mice. *Science* **316**, 394-396.
- Heasman, J., Wylie, C. C., Hausen, P. and Smith, J. C.** (1984). Fates and states of determination of single vegetal pole blastomeres of *X. laevis*. *Cell* **37**, 185-194.
- Heasman, J., Wessely, O., Langland, R., Craig, E. J. and Kessler, D. S.** (2001). Vegetal localization of maternal mRNAs is disrupted by VegT depletion. *Dev. Biol.* **240**, 377-386.
- Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L. et al.** (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* **328**, 633-636.
- Hinkley, C. S., Martin, J. F., Leibham, D. and Perry, M.** (1992). Sequential expression of multiple POU proteins during amphibian early development. *Mol. Cell. Biol.* **12**, 638-649.
- Horvay, K., Claussen, M., Katzer, M., Landgrebe, J. and Pieler, T.** (2006). *Xenopus* Dead end mRNA is a localized maternal determinant that serves a conserved function in germ cell development. *Dev. Biol.* **291**, 1-11.
- Houston, D. W. and King, M. L.** (2000). A critical role for *Xdazl*, a germ plasm-localized RNA, in the differentiation of primordial germ cells in *Xenopus*. *Development* **127**, 447-456.
- Hudson, C., Clements, D., Friday, R. V., Stott, D. and Woodland, H. R.** (1997). *Xsox17alpha* and *-beta* mediate endoderm formation in *Xenopus*. *Cell* **91**, 397-405.
- Irie, N., Weinberger, L., Tang, W. W. C., Kobayashi, T., Viukov, S., Manor, Y. S., Dietmann, S., Hanna, J. H. and Surani, M. A.** (2015). SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253-268.
- Jiang, X., Zhang, J. and Huang, Y.** (2015). Tetraspanins in cell migration. *Cell Adh. Migr.* **9**, 406-415.
- Kanatsu-Shinohara, M., Takashima, S., Ishii, K. and Shinohara, T.** (2011). Dynamic changes in EPCAM expression during spermatogonial stem cell differentiation in the mouse testis. *PLoS ONE* **6**, e23663.
- Kimura, T., Nakamura, T., Murayama, K., Umehara, H., Yamano, N., Watanabe, S., Taketo, M. M. and Nakano, T.** (2006). The stabilization of beta-catenin leads to impaired primordial germ cell development via aberrant cell cycle progression. *Dev. Biol.* **300**, 545-553.
- King, M. L., Messitt, T. J. and Mowry, K. L.** (2005). Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. *Biol. Cell* **97**, 19-33.
- Kloc, M. and Etkin, L. D.** (1995). Two distinct pathways for the localization of RNAs at the vegetal cortex in *Xenopus* oocytes. *Development* **121**, 287-297.
- Ko, H. L., Lam, T. H., Ng, H., Toh, J., Wang, L. W. and Ren, E. C.** (2017). Identification of Slug and SOX7 as transcriptional repressors binding to the hepatitis B virus core promoter. *J. Hepatol.* doi:10.1016/j.jhep.2017.08.033
- Lacerda, S. M. S. N., Costa, G. M. J. and de França, L. R.** (2014). Biology and identity of fish spermatogonial stem cell. *Gen. Comp. Endocrinol.* **207**, 56-65.
- Lai, F. and King, M. L.** (2013). Repressive translational control in germ cells. *Mol. Reprod. Dev.* **80**, 665-676.
- Lai, F., Singh, A. and King, M. L.** (2012). *Xenopus* Nanos1 is required to prevent endoderm gene expression and apoptosis in primordial germ cells. *Development* **139**, 1476-1486.
- Laird, D. J., Altshuler-Keylin, S., Kissner, M. D., Zhou, X. and Anderson, K. V.** (2011). Ror2 enhances polarity and directional migration of primordial germ cells. *PLoS Genet.* **7**, e1002428.
- Lau, N. C., Ohsumi, T., Borowsky, M., Kingston, R. E. and Blower, M. D.** (2009). Systematic and single cell analysis of *Xenopus* Piwi-interacting RNAs and Xiwi. *EMBO J.* **28**, 2945-2958.
- Mishra, M., Akatsu, H. and Heese, K.** (2011). The novel protein MANI modulates neurogenesis and neurite-cone growth. *J. Cell. Mol. Med.* **15**, 1713-1725.
- Mishra, M., Lee, S., Lin, M. K., Yamashita, T. and Heese, K.** (2012). Characterizing the neurite outgrowth inhibitory effect of Mani. *FEBS Lett.* **586**, 3018-3023.
- Morichika, K., Sugimoto, M., Yasuda, K. and Kinoshita, T.** (2014). Possible regulation of Oct60 transcription by a positive feedback loop in *Xenopus* oocytes. *Zygote* **22**, 266-274.
- Nieuwkoop, P. D. and Faber, J.** (1956). Normal Table of *Xenopus laevis* (Daudin). Amsterdam: North Holland Publishing.
- Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanynch, T., Nussenzweig, M., Tarakhovskiy, A. et al.** (2005). *Blimp1* is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213.
- Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T. and Saitou, M.** (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571-584.
- Owens, D. A., Butler, A. M., Aguero, T. H., Newman, K. M., Van Booven, D. and King, M. L.** (2017). High-throughput analysis reveals novel maternal germline RNAs crucial for primordial germ cell preservation and proper migration. *Development* **144**, 292-304.
- Pearson, J. C., Lemons, D. and McGinnis, W.** (2005). Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6**, 893-904.
- Philippidou, P., Walsh, C. M., Aubin, J., Jeannotte, L. and Dasen, J. S.** (2012). Sustained Hox5 gene activity is required for respiratory motor neuron development. *Nat. Neurosci.* **15**, 1636-1644.
- Randle, S. J. and Laman, H.** (2016). F-box protein interactions with the hallmark pathways in cancer. *Semin. Cancer Biol.* **36**, 3-17.
- Riadi, G., Ossandón, F., Larraín, J. and Melo, F.** (2016). Towards the bridging of molecular genetics data across *Xenopus* species. *BMC Genomics* **17**, 161.
- Saitou, M. and Yamaji, M.** (2012). Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* **4**, a008375.
- Saitou, M., Barton, S. C. and Surani, M. A.** (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300.
- Séguin, C. A., Draper, J. S., Nagy, A. and Rossant, J.** (2008). Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell* **3**, 182-195.
- Seydoux, G. and Dunn, M. A.** (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. *Development* **124**, 2191-2201.
- Sive, H. L., Grainger, R. M. and Harland, R. M.** (2000). Early Development of *Xenopus laevis*: A Laboratory Manual. Cold Spring Harbor, NY, USA: CSHL Press.
- Stennard, F., Carnac, G. and Gurdon, J. B.** (1996). The *Xenopus* T-box gene, *Antipodean*, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation. *Development* **122**, 4179-4188.
- Stennard, F., Zorn, A. M., Ryan, K., Garrett, N. and Gurdon, J. B.** (1999). Differential expression of VegT and Antipodean protein isoforms in *Xenopus*. *Mech. Dev.* **86**, 87-98.
- Tada, H., Mochii, M., Orii, H. and Watanabe, K.** (2012). Ectopic formation of primordial germ cells by transplantation of the germ plasm: direct evidence for germ cell determinant in *Xenopus*. *Dev. Biol.* **371**, 86-93.
- Taguchi, A., Tak, M., Motoishi, M., Or, H., II, Moch, M., II and Watanabe, K.** (2012). Analysis of localization and reorganization of germ plasm in *Xenopus* transgenic line with fluorescence-labeled mitochondria. *Dev. Growth Differ.* **54**, 767-776.
- Tang, W. W. C., Kobayashi, T., Irie, N., Dietmann, S. and Surani, M. A.** (2016). Specification and epigenetic programming of the human germ line. *Nat. Rev. Genet.* **17**, 585-600.
- Tarbashevich, K., Koebnick, K. and Pieler, T.** (2007). XGRIP2.1 is encoded by a vegetally localizing, maternal mRNA and functions in germ cell development and anteroposterior PGC positioning in *Xenopus laevis*. *Dev. Biol.* **311**, 554-565.
- Venkatarama, T., Lai, F., Luo, X., Zhou, Y., Newman, K. and King, M. L.** (2010). Repression of zygotic gene expression in the *Xenopus* germline. *Development* **137**, 651-660.
- Wang, C., Guo, Y., Wang, J. and Min, Z.** (2014). The suppressive role of SOX7 in hepatocarcinogenesis. *PLoS ONE* **9**, e97433.
- Xanthos, J. B., Kofron, M., Wylie, C. and Heasman, J.** (2001). Maternal VegT is the initiator of a molecular network specifying endoderm in *Xenopus laevis*. *Development* **128**, 167-180.
- Yamaguchi, T., Taguchi, A., Watanabe, K. and Orii, H.** (2013). Germes is involved in translocation of germ plasm during development of *Xenopus* primordial germ cells. *Int. J. Dev. Biol.* **57**, 439-443.
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J. et al.** (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131-1139.
- Zhang, J. and King, M. L.** (1996). *Xenopus* VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. *Development* **122**, 4119-4129.
- Zhang, J., Houston, D. W., King, M. L., Payne, C., Wylie, C. and Heasman, J.** (1998). The role of maternal VegT in establishing the primary germ layers in *Xenopus* embryos. *Cell* **94**, 515-524.
- Zhang, C., Basta, T., Fawcett, S. R. and Klymkowsky, M. W.** (2005). SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in *Xenopus*. *Dev. Biol.* **278**, 526-541.
- Zheng, C., Jin, F. Q. and Chalfie, M.** (2015). Hox proteins act as transcriptional guarantors to ensure terminal differentiation. *Cell Rep.* **13**, 1343-1352.
- Zhao, K., Wang, L., Li, T., Zhu, M., Zhang, C., Chen, L., Zhao, P., Zhou, H., Yu, S. and Yang, X.** (2017). The role of miR-451 in the switching between proliferation and migration in malignant glioma cells: AMPK signaling, mTOR modulation and Rac1 activation required. *Int. J. Oncol.* **50**, 1989-1999.