

Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling

5 Yukihiro Yabuta^{1*}, Kazuki Kurimoto^{1*}, Yasuhide Ohinata¹, Yoshiyuki Seki¹ and
Mitinori Saitou^{1,2,3}

¹Laboratory for Mammalian Germ Cell Biology, Center for Developmental Biology, RIKEN Kobe Institute, 2-2-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan.

10 ²Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, 4-1-8 Hon-cho, Kawaguchi, Saitama 332-0012, Japan.

³Laboratory of Molecular Cell Biology and Development, Graduate School of Biostudies, Kyoto University, Oiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan.

15

*These authors contributed equally to this work.

Short title

20 Gene expression dynamics for germline formation

Summary sentence

Quantitative single-cell gene expression profiling during the 36-hour period from embryonic day 6.75 to 8.25 of germ cell development identifies temporal coordination of critical events associated with germ cell specification in mice.

25

Grant support

This study was supported in part by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and by a PRESTO project grant from the Japan Science and Technology Agency.

30

Correspondence should be addressed to:

Mitinori Saitou, M.D., Ph.D.

35 Laboratory for Mammalian Germ Cell Biology, Center for Developmental Biology, RIKEN Kobe Institute, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047 Japan.

Tel: +81-78-306-3376

Fax: +81-78-306-3377

40 E-mail: saitou@cdb.riken.jp

Abstract

Germ cell fate in mice is induced in proximal epiblast cells at embryonic day (E) 6.5 by signaling molecules. *Prdm1* (also known as *Blimp1*)-positive lineage-restricted precursors of primordial germ cells (PGCs) initiate the formation of a cluster to differentiate into *Dppa3* (also known as *stella*)-positive PGCs from around E7.0 onwards in the extraembryonic mesoderm. Around E7.5, they begin migrating toward the definitive endoderm, with concomitant extensive epigenetic reprogramming. To gain a more precise insight into the mechanism of PGC specification and its subsequent development, we exploited quantitative single-cell gene expression profiling and explored gene expression dynamics during the 36 hours of PGC differentiation from embryonic day (E) 6.75 to 8.25, in comparison with those in somatic neighbors. This analysis revealed that the transitions from *Prdm1*-positive PGC precursors to *Dppa3*-positive PGCs and to more advanced migrating PGCs involve a highly dynamic, stage-dependent transcriptional orchestration that begins with the regaining of the pluripotency-associated gene network, followed by stepwise activation of PGC-specific genes, differential repression of the somatic mesodermal program, as well as the potential modulations of signal transduction capacities, and unique control of epigenetic regulators. The information presented here as to a cascade of events involving PGC development should serve as a basis for detailed functional analyses of the gene products associated with this process, as well as for properly reconstituting PGCs and their descendants in culture.

Introduction

65 Segregation of the germline from somatic lineages is one of the most essential events in development, since this process ensures the acquisition, modification and reservation of the ‘totipotent’ genome for subsequent generations. There are essentially two modes, often referred to as ‘preformation’ and ‘epigenesis’, respectively, to guarantee this process across the metazoans [1]. In preformation, a maternally inherited localized determinant, germ plasm, dictates germline specification so that early blastomeres that
70 inherit these RNA-protein granules adopt germ cell fate. In contrast, in epigenesis, germ cell fate is specified by signaling molecules at a relatively late stage of development in a potentially equivalent population of cells. Many model organisms, including *C. elegans*, *D. melanogaster*, *X. laevis* and *D. rerio*, specify germ cell fate by preformation, whereas mammals determine germ cell lineage by epigenesis [1, 2].

75

Classically, in the mouse, primordial germ cells (PGCs), the first population of the germ cell lineage appearing in the embryo, have been identified as an alkaline phosphatase-positive cluster of approximately 40 cells in the extraembryonic mesoderm at around embryonic day (E) 7.0 [3-5]. Subsequent works have revealed that signaling
80 molecules of the Bmp family from extraembryonic tissues and their signal transducers are important for the establishment of germ cell lineage [6-13]. However, until relatively recently, the intrinsic molecular mechanisms underlying germ cell specification and the properties of the resultant founder population of PGCs have been poorly defined.

85

Using single-cell gene expression analysis, Saitou et al. identified *Ifitm3* (also known as *fragils*) and *Dppa3*, which are highly and specifically expressed in the founder PGCs, respectively. Those authors proposed a molecular model for germ cell specification, a key event of which is the repression of *Homeobox (Hox)* genes in the founder PGCs at
90 E7.5 [14]. More recently, *Prdm1*, a potent transcriptional repressor of a histone methyltransferase subfamily, was found to mark the origin of the germ cell lineage in the most proximal layer of the epiblast as early as E6.25~E6.5 prior to the onset of gastrulation [15, 16]. These *Prdm1*-positive cells initiate to form a cluster within the extraembryonic mesoderm at around E6.75~E7.0 to differentiate into PGCs with

95 characteristic alkaline phosphatase activity, the expression of *Dppa3* and the repression
of *Hox* genes. Importantly, genetic lineage-tracing experiments showed that
Prdm1-expressing cells contribute only to *Dppa3*-positive PGCs, indicating that
Prdm1-positive cells in early gastrulating embryos are most likely lineage-restricted to
PGCs [15, 16]. Subsequently, *Dppa3*-positive PGCs start the migration toward the
100 definitive hindgut endoderm, with concomitant genome-wide epigenetic reprogramming
that includes substantial demethylation of DNA and histone3 lysine9 dimethylation
(H3K9me2) at around E8.0, and up-regulation of H3K27me3 at around E8.75 [17].

Despite these essential advances, there remains much to be learned about the precise
105 mechanisms of germ cell specification and the resultant characteristics of established
PGCs. The clarification of both of these still poses a formidable challenge, mainly
because of technical difficulties associated with the analysis of a small number of cells
that almost constantly change their properties cell-by-cell. For a precise understanding
and subsequent reconstruction of germ cell specification in vitro, an essential first step
110 is to develop a technique that enables quantitative analysis of this in vivo process.

In the present study, we exploited a modified quantitative single-cell cDNA
amplification method followed by real-time quantitative PCR (Q-PCR) analysis to
better understand the gene expression dynamics involved in germ cell specification.
115 We examined the expression of multiple genes with distinct functional categories at four
different time points during the 36 hours of germ cell differentiation from E6.75 to 8.25
and compared it with the expression in somatic neighbors. Our results reveal that
germ cell specification involves a dynamic, stage-dependent transcriptional control that
begins with the regaining of the pluripotency-associated gene network, followed by
120 stepwise acquisition of PGC-specific genes, differential repression of the somatic
mesodermal program, and a specific modulation of signal transduction capacities and
epigenetic regulators. These findings presumably reflect a complex network of
signaling molecules necessary for the proper setting of this potentially immortal cell
lineage and should serve as a basis for detailed functional analyses of the gene products
125 associated with this process, as well as for properly reconstituting PGCs and their
descendants in culture.

Materials and Methods

Isolation of total RNA from embryonic stem (ES) cells and cDNA synthesis

130 E14tg2a ES cells were cultured as described [18]. Total RNA from the ES cells was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany). For the synthesis of control cDNAs, 1 µg of total RNA was reverse-transcribed using Superscript III (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Single-cell-level cDNA amplification

135 Amplification of cDNAs from single-cell RNA or single-cell-level RNA was performed as described with modifications [19-21]. The ES cell total RNA (~1000 ng/µl) was serially diluted into concentrations of 2.5 ng/µl, 250 pg/µl, and 25 pg/µl. Then, 0.4 µl (10 pg, corresponding to the typical amount of mammalian single-cell-level total RNA) of the final dilution (25 pg/µl) was added into 4.75 µl of single-cell lysis buffer (1 x PCR buffer II [Applied Biosystems, Foster City, CA], 1.5 mM MgCl₂ [Applied Biosystems], 0.5% NP40, 5 mM DTT, 0.3 U/µl Prime RNase Inhibitor [Eppendorf, Hamburg, Germany], 0.3 U/µl RNAGuard RNase Inhibitor [Amersham Biosciences, Piscataway, NJ], 0.95 ng of the cDNA synthesis primer and 0.05 mM each of dATP, dCTP, dGTP and dTTP [Amersham Biosciences]). The cDNA synthesis primer
145 sequence is 5'-TATAGAATTCGCGGCCGCTCGCGA(T)₂₄-3'. All the primers described in this manuscript were purchased from Operon Biotechnology (Huntsville, AL) or Hokkaido System Science (Sapporo, Japan). The reaction mixture was heated at 70 °C for 90 sec and was immediately put on ice for 1 min. After brief centrifugation, a 0.3 µl volume of RT mixture (133.3 U/µl SuperScript III [Invitrogen],
150 3.33 U/µl RNAGuard RNase Inhibitor [Invitrogen], and 1.1-1.3 µg/µl T4 gene 32 protein [Roche, Basel, Switzerland]) was added to each reaction tube. The reaction mixture was incubated at 50 °C for 5 min and heat-inactivated at 70 °C for 10 min. The tubes were immediately put on ice for 1 min, and after 15 sec centrifugation, a poly-A tailing reaction was performed by adding 5 µl of TdT reaction buffer (1 x PCR buffer II, 1.5 mM MgCl₂, 3 mM dATP, 0.1 U/µl RNaseH [Invitrogen], and 0.75 U/µl Terminal Deoxynucleotidyl Transferase [Invitrogen]) and incubating at 37 °C for 15
155 min. The reaction was inactivated by incubation at 70 °C for 10 min. 50 µl of the

PCR mixture (1 × ExTaq buffer, 0.25 mM each of dATP, dCTP, dGTP, and dTTP, 1.2 μg
cDNA synthesis primer, and 0.05 U/μl ExTaq Hot Start Version [Takara]) containing
160 EX Taq buffer (Mg²⁺ included) was added, and mineral oil was overlaid. The second
strand was synthesized on a thermal cycler (Applied Biosystems) according to the
following protocol: 94 °C 3 min, 50 °C 2 min, and 72 °C 3 min. The PCR reaction was
immediately followed with the protocol: 94 °C 45 sec, 67 °C 1 min, and 72 °C 3 min
with a 6 sec extension per cycle for 24 cycles. The amplified cDNA was purified
165 using a QIAquick PCR purification kit (Qiagen) and dissolved in 50 μl of buffer EB (10
mM TrisCl [pH 8.5]).

Q-PCR and estimation of the gene copy number per cell

Q-PCR was performed using the 7900 Real Time PCR System (Applied Biosystems)
170 and SYBR Green PCR Master Mix (Applied Biosystems) according to the
manufacturer's instructions. The sequences of the primers for the genes examined are
listed in Supplemental Table S1. The doubling efficiencies of all the primers were
measured using plasmids (pGEM-T [Promega, Madison, WI]) bearing corresponding
specific inserts as templates. All the amplified samples were diluted appropriately so
175 that the threshold cycle (Ct) values of the genes examined are within the range of 11 to
30, within which Ct values of the standard plasmids were accurately correlated with
their concentrations. All the experiments were carried out twice, and the average of
the Ct values was then converted into approximate copy numbers as follows:

$$e_{xi} = S_x \cdot c_{xi} + I_x$$

180 where e_{xi} indicates the log₁₀ of transcripts per sample i , S_x and I_x are respectively the
slope and intercept estimated from the plot of copy numbers against Ct values of each
plasmid standard (Supplemental Table S1), and c_{xi} is the Ct value from Q-PCR.

Single-cell isolation and global cDNA synthesis

185 All the animals were treated with appropriate care approved by the RIKEN ethics
guidelines. Embryos of the CD-1 background at E6.75, E7.25, E7.75 and E8.25 were
collected in Dulbecco's modified Eagle's Medium (DMEM) (Gibco, Gaithersburg, MD)
with 0.5% BSA. Embryonic fragments from posterior extraembryonic mesoderm
(E6.75), the base of the allantois (E7.25-E7.75), and hindgut endoderm (E8.25) were

190 dissected out by a glass needle and incubated with 0.05% trypsin and 0.5 mM EDTA for
7 min, followed by dissociation into single cells by trituration with a mouth pipette.
Dissociated single cells were randomly picked up and placed directly in a tube with
4.75 μ l of the single-cell lysis buffer as described above. The entire process was
performed as quickly as possible in order to minimize the effect of trypsin/EDTA
195 treatment on gene expression.

The single-cell cDNA synthesis was performed exactly as described above for
single-cell-level cDNA amplification. The amplified samples were then screened to
identify the germline cells using gene-specific PCR for *Prdm1*, *Ifitm3* and *Pou5f1* (also
200 known as *Oct4*) (E6.75), or *Prdm1* and *Pou5f1* (E7.25-E8.25). Six cells from each
stage were selected and further amplified by PCR for the mass Q-PCR analysis. The
amplification was done in buffer containing 1 x Ex Taq buffer, 0.2 mM dNTP, 1 ng/ μ l
cDNA synthesis primer, and 0.1 unit/ μ l Ex Taq polymerase with 1/20 vol of template
with the following protocol: 95 $^{\circ}$ C 1 min, 67 $^{\circ}$ C 2 min, and 72 $^{\circ}$ C 6 min for 5 cycles.
205 The amplified cDNA was purified using a QIAquick PCR purification kit (Qiagen) and
dissolved in 50 μ l of buffer EB (10 mM TrisCl [pH 8.5]).

Normalization of gene expression measurements

The \log_{10} -expression levels (e_{xi}) of genes in single germline or somatic cells measured
210 by Q-PCR were normalized to minimize the technical variations in each experiment, as
follows:

$$E_{xi} = e_{xi} + T_i$$

215 where T_i is the normalization parameter for sample i , and e_{xi} and E_{xi} are the
 \log_{10} -expression levels of gene x before and after normalization, respectively. For
normalization standards, we used three housekeeping genes, *Gapdh*, *Arbp* and *Ppia*,
whose expression levels were assumed to vary little among cell types. The
normalization parameter T_i was calculated so that the variance of the normalized
220 expression levels of the standard genes was minimized using the least-squares method:

$$T_i = -\sum_{x=1}^M \left\{ e_{xi} - \left(\sum_{j=1}^N e_{xj} \right) / N \right\} / M$$

where N is the number of samples (42; 6 independently amplified samples at each of
 225 four and three developmental stages of the germline and somatic cells, respectively),
 and M is the number of genes used for normalization (three).

Since E_{xi} indicates the \log_{10} -expression levels after 24 cycles and/or 5 additional cycles
 of PCR amplifications, E_{xi} was then transformed into the \log_{10} of copy numbers per
 230 sample C_{xi} using one of the two following formul:

$$C_{xi} = E_{xi} - D_{24} \quad (\text{samples amplified by 24 cycles}),$$

$$C_{xi} = E_{xi} - D_{24} - D_5 \quad (\text{samples amplified by 24 plus 5 additional cycles}),$$

235 where D_{24} and D_5 are the \log_{10} -amplification coefficient during 24 or additional 5 cycles
 of PCR, respectively. D_{24} was calculated by comparison of expression levels of three
 housekeeping genes, *Gapdh*, *Arbp* and *Ppia*, in nonamplified cDNA sample (1 μ g of ES
 cell total RNA), and those in samples amplified by 24 cycles from a single-cell-level
 RNA (10 pg). In our condition, D_{24} was calculated to be 4.97 ± 0.13 . D_5 was
 240 calculated by comparison of expression levels of the same three genes before and after 5
 cycles of PCR amplification of ES cell samples amplified by 24 cycles and it was
 1.40 ± 0.06 .

Evaluation of the statistical significance of the single-cell gene expression differences

245 The significance of the difference between the estimated gene expression levels at a
 certain stage of the germ cell and somatic cell development was evaluated by analysis
 of variance (ANOVA) for the seven groups analyzed (E6.75, E7.25, E7.75, and E8.25
 germline cells, and E6.75, E7.25, and E7.75 somatic cells). For the genes that showed
 a statistically significant difference in expression levels ($P < 0.05$ by ANOVA) at least
 250 in one group, the difference in the expression levels between the germ and somatic cells
 at each of the corresponding stages was further examined by Student's t -test.

Heat map

255 The Microsoft Excel Visual Basic macro was utilized to make a heat map that has a gradient of blue, green, yellow, and red, according to the logarithm of the expression levels. The genes were grouped according to the results of ANOVA ($P < 0.05$) and Student's t -test ($P < 0.05$) (Supplemental Table S2).

Hierarchical clustering analysis

260 The unsupervised hierarchical clustering of the Q-PCR data obtained from all 42 single cells examined was performed using GenePattern software (<http://www.broad.mit.edu/cancer/software/genepattern/>) with the default setting. The analysis was conducted on the 57 genes significantly detected in at least one cell type (i.e., categories A-D of Figure 4; see Results).

265

Results

Simple semi-quantitative single-cell cDNA amplification method

We developed a simple semi-quantitative single-cell cDNA amplification method based on original PCR-based protocols described previously (see Materials and Methods) [19-21]. We isolated total RNA from ES cells (2.5×10^6) and diluted it to the single-cell level (~10 pg) for use as templates for amplification. Using Q-PCR, we evaluated the relative abundance of the transcripts in amplified samples. As shown in Fig. 1A and B, in our amplification protocol five genes expressed at widely different levels in ES cells (*Gapdh*, *Pou5f1*, *Sox2*, *Nanog*, *Dppa5*(also known as *Esg1*)) were amplified exponentially and relatively proportionally at least until the 24th cycle of PCR amplification, indicating a powerful and proportional amplification of a small amount of synthesized cDNA in our amplification procedure with fewer than 24 cycles. We then evaluated the representation and reproducibility of amplification from the single-cell level cDNA at the 24th cycle by comparing the relative abundance of as many as 21 genes (*Ppia*, *Arbp*, *Gapdh*, *Dppa5*, *Zfp42*(also known as *Rex1*), *Pou5f1*, *Sox2*, *Ifitm3*, *Eed*, *Nanog*, *Dnmt3b*, *Yy1*, *Dnmt1*, *Esp11*, *Ezh2*, *Ehmt2*(also known as *G9a*), *Eras*, *Ehmt1*(also known as *Glp1*), *Nodal*, *Dppa3* and *Tial1*) in amplified cDNAs from nine independent experiments and in nonamplified cDNAs synthesized directly from undiluted (1 μ g) total RNA. As shown in Fig. 1C and D, most of the 21 genes, especially the ones expressed in more than 100 copies per cell, were represented correctly in the amplified products with high reproducibility. The majority of the genes were plotted between four-fold difference lines compared with the nonamplified control in each experiment (for all expression-level ranges, $R^2=0.884$, with 62 and 84% of the genes plotted between 2.0- and 4.0-fold difference lines, respectively, and for the genes more than 100 copies per cell, $R^2=0.883$, with 81 and 96% of the genes plotted between 2.0- and 4.0-fold difference lines, respectively). Although it must be noted that some genes showed relatively poor amplification, which might be due to their specific three-prime sequences, these results collectively indicate that our amplification procedure detects most, if not all, of the genes expressed at the single-cell level representatively and reproducibly.

Population analysis of Prdm1-positive cells during the specification of germ cell fate

Using the simple quantitative single-cell cDNA amplification method, we set out to precisely explore the expression dynamics of key genes associated with the specification of germ cell fate in mice. We dissected out the regions where the PGC precursors (defined as *Prdm1*-positive and *Dppa3*-negative cells) or PGCs (defined as *Prdm1*-positive and *Dppa3*-positive cells) were most likely localized as small as possible from embryos of CD1 background at E6.75 (early/mid-streak (E/MS) stage), E7.25 (not/early bud (0/EB) stage), E7.75 (early head fold (EHF) stage), and E8.25 (8~10 somite stage) (Fig. 2A). For embryos from E6.75 to E7.75, we removed the visceral endoderm from the fragments, and for embryos at E7.25 and E7.75, we removed the extending allantois as much as possible. We then dissociated them into single cells, which we picked up randomly to synthesize single-cell cDNAs. The number of single cells isolated at each stage is summarized in Fig. 2B. All the amplified samples were screened by gene-specific primers for *Prdm1* [15], *Ifitm3* [14, 22, 23], and *Dppa3* [14, 24] for E6.75 embryos and for *Prdm1* and *Dppa3* for the embryos from other stages so as to identify the PGC precursors (E6.75) and the PGCs (E7.25, E7.75, and E8.25), respectively. To examine the precise character of *Prdm1*-positive cells during the course of PGC differentiation, all the *Prdm1*-positive cells from all stages were also examined for the expression of *Pou5f1* [25] and *Sox2* [26], which are master regulators for pluripotency [27], and *Hoxb1*, a key gene repressed in the founder PGCs [14], and the approximate copy number of each gene expressed in single cells was estimated (see Materials and Methods, Figs. 3 and 4 and data not shown). Almost all *Prdm1*-positive cells we analyzed were *Pou5f1*-positive from E6.75 to E7.75 (86 of 89, 97%), whereas eight of thirty *Prdm1*-positive cells were negative for *Pou5f1* at E8.25. The identities of these *Prdm1*-positive *Pou5f1*-negative cells at E8.25 remain undetermined.

Figure 2B-D summarizes the transition of the proportion of the sub-populations of *Prdm1*-positive and *Pou5f1*-positive cells at the four developmental time points. These data indicate potentially important events for understanding the germ cell specification process. First, *Hoxb1* mRNA is present initially in at least some PGC precursors. At E6.75, as much as 35% (9 of 26) of the *Prdm1*-positive population was

Hoxb1-positive (sum of blue and sky blue circles in Fig. 2C), and at E7.25, 30% (12 of
330 40) of the *Prdm1*-positive population was *Hoxb1*-positive. The ratio of
Hoxb1-positive cells dropped after E7.75 in the populations we analyzed, but
approximately 10% of the *Prdm1*-positive cells still expressed *Hoxb1*. Second, almost
all cells with *Dppa3* expression showed no *Hoxb1* expression (57 of 61, 93%),
irrespective of the original embryonic stages (Fig. 2B and 2C and data not shown).
335 Third, the expression of *Sox2* was detected only in a subpopulation of *Prdm1*-positive
cells at E6.75; only 27% of the *Prdm1*-positive cells at this stage expressed *Sox2* in
more than 10 copies per cell (7 of 26, data not shown) (54% in all *Sox2* expression level
ranges). The proportion of *Sox2*-positive cells increased gradually during the course
of PGC specification, and most of the cells we analyzed had become *Sox2*-positive at
340 E8.25 (91%; 20 of 22) (Fig. 2B and 2D). Interestingly, similar to the case with the
mutually exclusive expression of *Dppa3* and *Hoxb1*, almost all *Sox2* strongly positive
cells were *Hoxb1*-negative (data not shown). Fourth, *Sox2* expression in the
Prdm1-positive populations preceded *Dppa3* expression in the same populations (Fig.
2D), demonstrating that *Sox2* shows germline-specific expression (see below and Figs.
345 3 and 4) in between *Prdm1* and *Dppa3*. Collectively, these data suggest a possibility
that the *Prdm1*-positive lineage-restricted PGC precursors [15], at least some of which
are initially positive for *Hoxb1* and may have gene expression properties similar to
those of neighboring somatic cells, up-regulate *Sox2* followed by *Dppa3* to acquire the
PGC property (see Discussion). The established PGCs do not show *Hoxb1* expression.

350

Gene expression dynamics associated with PGC specification

To investigate more extensively how the PGCs acquire their characteristic properties,
we went on to analyze the expression of various functional categories of genes during
the specification of PGC fate. Six amplified cells at each of four stages were selected
355 for the analyses; two *Prdm1*- positive and *Hoxb1*-positive and four *Prdm1*-positive and
Hoxb1- negative cells at E6.75, and six *Prdm1*-positive and *Dppa3*-positive cells for all
the other stages. At the same time, gene expression in neighboring somatic cells
(*Prdm1*-negative and *Hoxb1*-positive cells) that likely share a common origin with
PGCs at E6.75, E7.25, and E7.75 were analyzed for comparison. We did not include
360 somatic cells at E8.25 because many of the PGCs at E8.25 migrate in the hindgut [17],

and neighboring somatic cells at this stage are no longer a similar population as they are at E6.75-E7.75. The 42 selected cells (6 PGCs at each of four stages, 6 somatic cells at each of three stages) were further amplified quantitatively by five more PCR cycles (see Materials and Methods) and the expression levels of multiple genes including three housekeeping genes, *Gapdh*, *Ppia*, and *Arbp* were analyzed by Q-PCR (Fig. 3).

Housekeeping genes

First we examined the expression of housekeeping genes. The mRNAs for *Gapdh*, *Arbp*, and *Ppia* were estimated to be present on the order of approximately 10^3 to 10^4 molecules per cell relatively constantly in all the single-cell cDNAs examined. This agreed with the previous report on the transcript number estimation of housekeeping genes within a single cell [28], although the cells used were different, demonstrating further the quantitative amplification with our method. We therefore used the expression levels of these three genes to normalize all the gene expression data in this manuscript (see Materials and Methods).

Key molecules associated with germline specification

We next quantified the expression of four PGC markers, *Ifitm3*, *Dppa3*, *Prdm1*, and *Akp2* (also known as Tissue non-specific alkaline phosphatase) [29]. Consistent with the previous observations [14, 22], although *Ifitm3* was detected in somatic cells to some extent, it was expressed much more strongly (from 10^2 to more than 10^3 copies per cell) in the PGC precursors and the PGCs. *Dppa3* expression was detected only in a small population of cells at E6.75 (7.7%, 2 of 26) (Fig.2B), which we consider the first few cells that acquire PGC characteristics, but was sharply and strongly up-regulated at E7.25 (around 10^3 copies per cell at E8.25), whereas no expression was observed in *Prdm1*-negative somatic neighbors. *Prdm1* was expressed in the PGC precursors (E6.75) and the PGCs at the range of 10^1 - 10^2 copies per cell by this method. *Akp2*, the most classical marker for PGCs, was detected not only in the PGC precursors and the PGCs but also in somatic neighbors, with PGCs having the highest expression (~ 5×10^2 copies per cell). These results are highly consistent with previous in situ hybridization and transgenic analyses [14, 15, 22, 29], supporting the idea that our simple single-cell cDNA amplification method successfully quantifies gene expression

levels in single cells and reinforces the earliest and specific onset of *Prdm1* expression in the germ cell lineage.

395

We then went on to determine the onset of the expression of several genes that are known to have critical roles in PGC development. *Nanos3*, a mouse homolog of the evolutionarily conserved RNA binding protein Nanos [30-32], has been known to be important for PGC development, since the loss of its function leads to the disappearance of PGCs by E12.5 [32]. Since the PGCs in the absence of *Nanos3* are apparently affected as early as E8.5, *Nanos3* is considered to express from earlier stages, but its expression has been confirmed only after E9.5 [32]. We therefore examined its expression and found that *Nanos3* expression is indeed initiated specifically in PGCs at E7.25 (10^1 - 10^2 copies per cell). This finding is consistent with the phenotype of the *Nanos3* mutant and raises *Nanos3* as an early germ-cell-specific transcript as *Dppa3*. We next examined the expression of Dead end (*Dnd*) [33], the mouse orthologue of the Zebrafish RNA binding protein Dead end [34]. *Dnd* has been recently identified as a gene responsible for *Ter* mutation, in which PGC development is severely affected as early as E8.0 [35], and is suggested to have a role in RNA metabolism in the germ cell lineage [33]. However, as is the case for *Nanos3*, the precise manners of *Dnd* expression in early PGCs and somatic neighbors have not been well analyzed. We found that at E6.75, *Dnd* is expressed both in the PGC precursors and somatic neighbors at some frequency but is sharply up-regulated only in the PGCs after E7.25. The expression level of *Dnd* increased to about $\sim 2 \times 10^2$ copies per cell at E8.25 in PGCs. We then examined the expression of *Tiall*, which encodes another RNA binding protein and whose loss of function leads to early (\sim E9.5) PGC death with unknown mechanisms [36], and found that *Tiall* was expressed at similar levels in both the single-cell cDNAs of PGCs and somatic neighbors, suggesting that *Tiall* may be up-regulated in PGCs subsequently. These findings demonstrate the power of our single-cell cDNA quantitative gene expression analyses for the accurate description of gene expression profiles.

420

The Kit tyrosine kinase receptor, which is known to express in PGCs as early as E7.5 [37], is involved in germ cell development [38, 39], together with its ligand, *Kitl* [39].

425 We therefore monitored the expression levels of these genes. Consistent with and
more clearly than the previous *in situ* hybridization data, *Kit* expression was specifically
up-regulated in PGCs at E7.25, and its expression was maintained throughout the
observed stages. *Kit* in neighboring somatic cells, on the other hand, was not detected
between E6.75-E7.75. In contrast, *Kitl* was temporarily expressed highly in PGCs at
430 E7.25, then gradually decreased within 24 hours. It is of note that PGCs at E7.25 that
form a tight cluster express both *Kit* and *Kitl* at a high level (both for $\sim 10^2$ copies per
cell). *Cxcl12* is a gene that encodes a chemokine that regulates PGCs to colonize
gonads [40]. We found that this gene does not show a significant difference in its
expression between PGCs and their somatic neighbors during the period we examined.

435

Since *Prdm1* encodes a protein that is a member of the PR domain-containing protein
family, we examined the expression of all the other family members (*Prdm1-Prdm16*).
The results showed that the expression levels of *Prdm4*, *Prdm5*, and *Setd8* were similar
between PGCs and somatic cells at the stages examined. Interestingly, however, we
440 found that *Prdm14* was up-regulated exclusively in the PGCs, in a manner more similar
to those of *Dppa3*, *Nanos3*, and *Kit*, suggesting the involvement of *Prdm14* in PGC
specification.

Pluripotency-associated genes

445 Next, the expression of genes related to pluripotency—*Pou5f1*, *Sox2*, *Nanog*, *Dppa5*,
and *Eras*—were monitored. *Pou5f1*, *Sox2*, and *Nanog* are reported to be required for
the pluripotency of inner cell mass cells of blastocysts as well as for the derivation and
maintenance of ES cells [25, 26, 41, 42]. We found that *Pou5f1* is expressed in the
Prdm1-positive PGC precursors and the PGCs at all of the observed stages at similar
450 levels (~ 100 copies per cell), while neighboring somatic cells gradually down-regulated
the expression. Despite consistent high expression levels of *Pou5f1* in the PGC
precursors and PGCs, *Sox2* expression was low or absent at E6.75 (see also Fig. 2B and
D, as well as the above section on population analyses), but was progressively
up-regulated in the following stages. Importantly, *Sox2* expression was hardly
455 detected in neighboring somatic cells, indicating that the POU5F1-SOX2 complex
exists only in PGC precursors after E6.75 in the posterior extraembryonic mesodermal

regions we analyzed. *Nanog* was found to be expressed continuously in the PGC precursors, but was down-regulated sharply in somatic neighbors after E6.75. In contrast, the expression of *Dppa5*, a gene highly enriched in early embryos, ES cells, and PGCs in the genital ridges [43], was detected only at E6.75 at low levels in both the PGC precursors and somatic cells, and was not up-regulated until at least E8.25 in the PGCs. We have also examined the expression of *Eras*, a gene involved in the tumorigenic activity of the ES cells [44] and found that this gene was not expressed in any of the cells we examined. These results indicate that PGC specification acquires the expression of key pluripotency-associated transcription factors but does not require up-regulation of all the genes associated with ES cells. By further monitoring the expression of other *Sox* genes, we found that *Sox3* and *Sox17* are transiently and specifically up-regulated in PGCs at around E7.25.

470 *Hox* genes and mesodermal markers

Previous reports suggested that *Hox* cluster genes were repressed during the specification of the PGC fate, whereas neighboring somatic cells up-regulate the expression of these genes to acquire the somatic mesodermal fates [14, 15]. Therefore, the expression of several genes related to mesodermal differentiation was monitored in the single-cell samples. As shown in Figure 2 and described in the above section, the time course analysis indicated that somatic neighbors express *Hoxb1* already at E6.75 (10^1 - 10^2 copies per cell) and continue to express it throughout the stages examined, whereas PGCs down-regulated it. Somatic neighbors started to express *Hoxa1* at E7.25 and continued to express it. On the other hand, although some PGCs showed temporarily low expression levels of it at around E7.25, they essentially shut off its expression at subsequent stages. Regarding the expression of *Evx1*, *T*, and *Fgf8*, which are typical markers for posterior mesoderm [45-48], PGC precursors at E6.75 showed expression levels as high as those of somatic neighbors, but PGCs from E7.25 onwards gradually down-regulated them, suggesting that some of the initially acquired mesodermal properties of the PGC precursors are eventually extinguished in the germline. It is also of note that the expression of *Fgf8* was more tightly associated with PGCs from E6.75 to E7.75. *Wnt5b*, another gene that marks the posterior mesoderm, including the allantois and primitive streak [49], showed rapid

down-regulation at E7.25 in PGCs (from ~20 copies per cell at E6.75 to ~0 copy at
490 E7.25), whereas it continued to be expressed in somatic neighbors relatively more
constantly. *Wnt5a* [50] was constantly expressed at low levels (~10 copies per cell) in
both the germ and somatic lineages during the period examined. We then examined
the expression of *Snai1*, a key gene involved in the epithelial-mesenchymal transition
[51-53]. As expected, we detected that somatic mesodermal cells express this gene
495 relatively highly, especially at E7.25. In contrast, PGCs strongly repress the
expression of *Snai1* as early as E7.25. Additionally, we found that the expression of
Myc, a prototypical oncogene involved in cell proliferation and a target for
Prdm1-mediated repression in plasma cell differentiation [54], was sharply repressed
only in the germ cell lineage after E7.25.

500

Epigenetic regulators

Next, we focused on the expression of epigenetic regulators. As we recently reported,
PGCs undergo extensive reprogramming of their epigenome soon after their
specification, which includes the erasure of DNA methylation and H3K9me2 as well as
505 the up-regulation of H3K27me3 [17]. Therefore, we first examined the expression of
DNA methyltransferases (DNMTs) [55, 56]. We found that *Dnmt1*, a maintenance
methyltransferase essential for early embryogenesis [55], was constantly expressed at
the mRNA levels in both PGCs and somatic neighbors at the stages we examined (~10²
copies per cell). Remarkably, however, we found that the expression of a *de novo*
510 DNA methyltransferase, *Dnmt3b*, was strongly suppressed in PGCs at E7.25, which is
in sharp contrast with its relatively constant expression in somatic neighbors. The
expression of *Dnmt3a* was constantly low in both the germ and somatic cell lineages
during the period we examined.

515 We next monitored the expression of histone methyltransferases. *Ehmt2* is a histone
methyltransferase predominantly involved in the di-methylation of H3K9, and its
activity is essential for early embryonic development [57]. EHMT1, a protein highly
similar to EHMT2, forms a complex with EHMT2, and both of these components are
essential for the H3K9 dimethylation activity of the complex [58]. We found that
520 *Ehmt2* was expressed relatively constantly in both the germ and somatic lineages during

the period we examined, whereas *Ehmt1* was apparently down-regulated significantly in the PGCs after E7.25. *Ezh*, *Eed* and *Suz12* are members of the polycomb group (PcG) complex, which is involved in trimethylation of H3K27 [59-62]. These three genes were continuously expressed at similar levels in both the germ and somatic cell lineages during E6.75-E8.25. YY1 is also a member of the PcG complex [63, 64], and its transcript was continuously expressed at all stages in both PGCs and somatic cells.

Signaling molecules

Next, we scrutinized the expression of genes involved in cellular signal transductions [65, 66]. Although the expression levels of this category of genes measured by this method were generally low and it was somewhat difficult to identify significant differences between germ and somatic lineages, we nonetheless found some statistically significant differences that may reflect the different properties of these spatially neighboring populations. It has been known from gene knockout studies that molecules involved in Bmp signal transduction (*Bmp4*, *Bmp8b*, *Bmp2*, *Smad1*, *Smad5*, *Acvr1*(also known as *Alk2*)) [6-13] are important for the specification of germ cell fate. However, there has been very little information on the intrinsic expression of these molecules in early germline cells. We first examined the expression of type I (*Acvr1* and *Bmpr1a*(also known as *Alk3*)) and type II (*Acvr2b* and *Bmpr2*) Bmp receptors. We detected *Acvr1* at a low level in both the germ and somatic lineages from E6.75 onwards. In contrast, we found that *Bmpr1a* is expressed at a relatively high level (~10² copies per cell) in the PGC precursors at E6.75 but was repressed gradually to a low level in the PGCs, whereas it was expressed at a similar level in somatic neighbors throughout the examined stages. *Acvr2b* was expressed more constantly in both germ and somatic cells, whereas we could not detect the expression of *Bmpr2*.

We next examined the expression of *Smads*, intracellular mediators of BMP signaling [65]. Our results showed that in the germ cell lineage, *Smad1* was significantly expressed at E6.75 (~10² copies per cell) but was progressively down-regulated, which is consistent with a previous report using the *Smad1*-LacZ allele [10]. *Smad5* expression was only sporadically observed in the germ cell lineage from E6.75 to E7.75, whereas somatic neighbors expressed this gene more consistently throughout the period

we examined. The expression of *Smad4*, a co-Smad that forms heterodimers with other Smad proteins to work as transcription factors in the nucleus, was detected more consistently in the germ cell lineage and somatic neighbors, although it seemed slightly
555 down-regulated in the germ cell lineage at E7.25 and E7.75. These results are consistent with the expression of Bmp receptor molecules. In contrast, we found that PGCs at E7.25 and E7.75 significantly and specifically up-regulate the expression of *Smad3*, which mediates TGFbeta signaling through heterodimerization with *Smad4*.
560 Collectively, these results suggest that PGCs may decrease sensitivity to Bmp signaling and instead increase sensitivity to TGFbeta signaling at around E7.25 to E7.75. Additionally, we found that the germline cells express nodal specifically at E6.75 and E7.75, albeit at low levels.

565 We next examined whether or not the JAK/STAT signaling pathway [66] may play a role in germ cell specification. Toward this end, we analyzed the expression of intracellular signal transducers of *Jak1*, *Jak2*, *Tyk2*, *Stat2*, *Stat3*, *Stat5 α* , *Stat5 β* and *Irf8* as well as receptors for interferon, including *Ifnar1* and *Ifnar2*. This analysis revealed that, although *Ifnar1* was expressed consistently in both germ and somatic lineages,
570 none of the signal transducers examined were detected at significant levels in both the germ and somatic mesodermal lineages, suggesting that the JAK/STAT signaling pathway may not be important in germ cell specification nor in the formation of extraembryonic mesoderm.

575 *Telomere-related genes and RNAi machinery*

Telomere maintenance is one of the key functions by which germ cells inherit a complete genome into the next generations [67]. We therefore examined the expression of genes associated with telomere maintenance, including *Tert*, *Terf1*, *Terf2*, *Pot1*, *Rif1* and *Wrn*, and found that there appears to be no distinguishable features
580 between early germline cells and somatic neighbors regarding the expression of these genes (Fig. 4). In addition, we examined the expression of *Dicer1* and *Eif2c2* (also known as *Ago2*), key components of RNAi machinery [68]. We found that *Dicer1* was expressed more constantly in the germline and that *Eif2c2* was hardly detectable with our method in both the germ and somatic lineages (Fig. 4).

Summary of gene expression profiles

Figure 4 summarizes the gene expression profiles in the form of a heat map, based on the statistical analysis of gene expressions. We have classified the genes analyzed into five categories according to their expression patterns and levels. Those are the genes with A) very high expression levels in both the germ and somatic cell lineages (housekeeping genes), B) exclusive or higher expression in the germ cell lineage at at least one stage of development, C) down-regulation in the germ cell lineage but up-regulation or more constant expression in the somatic cells, D) relatively constant expression in both the germ and somatic lineages and E) no significant detectable expression. Group B) consists of genes such as *Ifitm3*, *Prdm1*, *Dppa3*, *Sox2*, *Prdm14*, *Nanos3*, *Kit* and *Dnd*. *Prdm1* and *Sox2* were two of the genes that are expressed specifically in PGCs throughout E6.75-8.25. *Dppa3*, *Kit*, *Nanos3* and *Prdm14* were more consistently up-regulated after E7.25, and *Sox3* and *Sox17* were transiently up-regulated at around E7.25. Group C) includes genes such as *Hoxb1*, *Hoxa1*, *Dnmt3b*, *Snai1* and *Myc*. These genes showed complete transcriptional repression in PGCs at E7.25-7.75, whereas they were constantly expressed in the neighboring somatic cells. Group D) includes genes such as *Ezh2*, *Eed*, *Suz12*, *Ehmt2*, *Prdm4*, *Setd8*, *Dnmt1* and *Smad4*. Group E) contains genes that showed almost no detectable expression at observed stages, or possibly genes with expression levels lower than in group D).

We performed unsupervised hierarchical clustering of the Q-PCR data obtained from all 42 single cells using the 57 genes significantly detected in at least one cell type (i.e., categories A-D in Figure 4). The results showed that 1) the PGC precursors at E6.75 are more similar to the clusters of somatic neighbors, 2) PGCs at E7.25 constitute an independent cluster, to which three PGCs at E7.75 showed similar properties, and 3) PGCs at E8.25 constitute another independent cluster, to which three PGCs at E7.75 exhibited high similarity. These data suggest that a key property of PGCs is initially acquired between E6.75 and E7.25, and that PGCs progressively alter their gene expression from E7.25 to E7.75, and to E8.25, indicating that our present study identifies dynamic gene expression transitions during the early development of the germ cell lineage.

Discussion

Using a simple quantitative single-cell cDNA amplification method followed by Q-PCR, we have analyzed gene expression dynamics during the 36 hours of germ cell differentiation from E6.75 to 8.25. The precise mode of the specification of germ cell fate is an active area of research. It has not yet been strictly determined how the *Prdm1*-positive cells increase their number or how long the recruitment of the germ cell lineage from their precursors continues (see below). Nonetheless our present study, involving gene expression analysis of multiple *Prdm1*-positive cells at four different time points, not only provided a precise view of the dynamic process with which PGCs eventually acquire their characteristic profile, which thus far has been reported only in part [14, 15], but also presented novel profiles, including that of signal transduction capacity and that of epigenetic regulation in the PGCs. Since the method we have employed is very simple and straightforward, it should be generally applicable to any biological situations where accurate single-cell analysis might be required.

Our analyses revealed that in the CD1 background, at least a certain proportion of the *Prdm1*-positive cells at E6.75 and at E7.25 (approximately 35% and 30%, respectively) showed expression of *Hoxb1*. Even at E7.75 and E8.25, approximately 10% of the *Prdm1*- and *Pou5f1*-positive populations we analyzed in this study expressed this gene. On the other hand, a previous report involving genetic lineage-tracing experiments showed that *Prdm1*-expressing cells contribute only to *Dppa3*-positive PGCs when examined in between the LB and the 3-somite stages, indicating that *Prdm1*-positive cells are most likely lineage-restricted to PGC fate [15, 16]. It would therefore be reasonable to assume that *Prdm1*-positive and *Hoxb1*-positive cells develop at least until E7.25 to become *Prdm1*- and *Dppa3*-positive PGCs with the repression of *Hoxb1*. This idea is supported by the finding that other key mesodermal markers, including *T*, *Fgf8*, *Wnt5b*, and *Snai1*, are also initially expressed at E6.75 in the majority of the *Prdm1*-positive cells we analyzed, but these genes are repressed along with the progression of PGC specification apparently in an orderly, gene-dependent manner (Fig. 3, 4 and data not shown). Since PGC specification in mice takes place in the region potentially influenced by numerous mesoderm-inducing activities, initial up-regulation of mesodermal markers in PGC precursors would not be a total surprise.

650 On the other hand, the fate of *Prdm1*-positive and *Hoxb1*-positive cells at E7.75 and
E8.25 is unknown and deserves careful further investigation. The expression of *Prdm1*
detected by in situ hybridization, and that in *Prdm1*-mEGFP transgenic mice, continues
in the most proximal posterior epiblast cells and in some nascent mesodermal cells
apparently originating from these epiblast cells at least until E8.25. Some of these
655 nascent mesodermal cells show clear expression of DPPA3 (data not shown).
Therefore, in one possibility, *Prdm1*-positive and *Hoxb1*-positive cells at E7.75 and
E8.25 still go on to form *Dppa3*-positive and *Hoxb1*-negative PGCs, as is the case for
earlier epiblast cells. Alternatively, these cells fail to contribute anymore to the germ
cell lineage and entirely adopt somatic mesodermal fates. It would be important to
660 determine until what stage all the *Prdm1*-positive epiblast cells contribute exclusively to
the germ cell lineage and how long the most proximal *Prdm1*-positive epiblast cells can
produce *Dppa3*-positive PGCs.

Our analysis also identified an exclusive re-activation of *Sox2* expression in the
665 *Prdm1*-positive population. Most of the cells (66 out of 76, ~87%) that acquired *Sox2*
expression were negative for *Hoxb1*, and almost all the cells (57 out of 61, ~93%) that
were positive for *Dppa3* showed complete repression of *Hoxb1* expression (Fig. 2).
Furthermore, the ratio of *Dppa3*-positive cells among the *Sox2*-positive cells increases
during the PGC specification period (14% at E6.75, 83% at E7.25, 94% at E7.75 and
670 95% at E8.25). Considering that somatic neighbors are negative for *Sox2* (Figs. 3, 4),
the up-regulation of *Sox2* that preceded *Dppa3* expression and the repression of *Hoxb1*
may be a key step for the specification of PGC fate. Since SOX2, which forms a
heterodimeric complex with POU5F1, has been shown to be a key player for
pluripotency in ES cells [26], the exclusive re-acquisition of *Sox2* expression in the
675 germ cell lineage may herald the regaining of the potential pluripotency in this lineage.
We also found that another key player in pluripotency, *Nanog*, is specifically maintained
only in the PGCs among extraembryonic mesodermal cells after E7.25. A previous
report showed that expression of *Nanog* is first down-regulated in the epiblast after
implantation [41], which is consistent with our recent single-cell microarray analysis
680 [69]. More recent reports [70, 71] showed that in early post-implantation embryos,

NANOG expression persists in the epiblast cells at E6.5 and E7.5, suggesting that NANOG expression may be up-regulated in the epiblast cells at these later stages. Interestingly, in these reports NANOG protein in PGCs was detected only after E7.5. However, we observed continuous expression of *Nanog* mRNA in the *Prdm1*-positive
685 cells from E6.75. This apparent discrepancy might be due to some post-transcriptional regulation of *Nanog* mRNA in these cells. Collectively, these data provide the first clear indication that only germ cell lineage acquires the ability to regain the expression of key transcription factors for pluripotency, at least at the mRNA level, after E6.75.

690 The study also identified that *Nanos3* and *Dnd*, two potential RNA binding proteins essential for PGC development [32, 33], *Kit*, a receptor tyrosine kinase essential for PGC survival, and *Prdm14*, a PR-domain-containing gene with unknown function, show specific up-regulation in PGCs at least by E7.25. Therefore, we now know that
695 *Prdm1*, *Prdm14*, *Sox2*, *Dppa3*, *Nanog*, *Kit*, *Nanos3*, and *Dnd* are among the genes that specifically associate with the PGC specification process. Also, from our time-course analysis, *Sox2* expression is considered to occur prior to *Dppa3* expression. Precise determination of the onset of the expression of these genes in the *Prdm1*-positive cells and how the expression of each gene is affected in the mutants of these genes would clarify a molecular pathway and ideally a network leading to the formation of functional
700 germ cell lineage.

It is also of note that some of the general epigenetic regulators are repressed in the PGCs. The expression of both *Dnmt3b* and *Ehmt1* is repressed after E7.25, which may be important for the observed genome-wide demethylation of DNA and H3K9me2 in
705 PGCs at around E8.0 [17]. In contrast, the expression of all the known components of H3K27 trimethylation is kept constant in the germ cell lineage, which would be essential for the exclusive up-regulation of H3K27me3 in PGCs at around E8.75. These findings suggest that specific epigenetic reprogramming in the germ cell lineage is preceded by, and a consequence of, a transcriptional regulation of epigenetic
710 regulators in this lineage. It would be reasonable to assume that factors involved in germ cell specification systematically set up this event to occur.

Our results also revealed that germ cell specification may accompany a specific control of signal transduction machineries at the transcriptional level. We found that PGCs
715 apparently reduce the competence of responding to BMP signals at E7.75, by repressing the expression of the receptor *Bmpr1a* and the intracellular signal transducers of *Smad1* and *Smad5* at the transcription level. Instead, they specifically up-regulate *Smad3* expression, which transmits TGFbeta signaling. A recent study indicated that Bmp4 signaling from the extraembryonic ectoderm is mediated through the visceral endoderm
720 and suggested that some unidentified factor from the endoderm induces the germ cell fate [13]. Investigations of the intrinsic expression of signal transducing molecules in *Prdm1*-positive cells during the PGC specification period may lead to the identification of a pathway for the induction of the germ cell lineage.

725 In summary, our single-cell quantitative gene expression profiling during the PGC specification process identified the precise dynamics of the activation and repression of a specific set of genes associated with PGC development. It will be important to validate the functional significance of the transcriptional changes of these genes, since there still be only a few information available for the functional relevance of these
730 transcriptional changes [15, 32, 33, 38, 39]. This study would be a key step toward the full elucidation of the PGC specification process using the single-cell microarray technology we recently developed [69], and may provide information essential for the proper reconstitution of PGCs and their descendants in culture.

735 *Acknowledgment*

We thank all the members of our laboratory for their discussion of this study.

References

- 740 1. Extavour CG, Akam M. Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 2003; 130: 5869-5884.
2. Eddy EM. Germ plasm and the differentiation of the germ cell line. *Int Rev Cytol* 1975; 43: 229-280.
3. Chiquoine AD. The identification, origin and migration of the primordial germ cells in the mouse embryo. *Anat. Rec.* 1954; 118: 135-146.
- 745 4. Ginsburg M, Snow MH, McLaren A. Primordial germ cells in the mouse embryo during gastrulation. *Development* 1990; 110: 521-528.
5. Lawson KA, Hage WJ. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp* 1994; 182: 68-84.
- 750 6. Lawson KA, Dunn NR, Roelen BA, Zeinstra LM, Davis AM, Wright CV, Korving JP, Hogan BL. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 1999; 13: 424-436.
7. Ying Y, Liu XM, Marble A, Lawson KA, Zhao GQ. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol* 2000; 14: 1053-1063.
- 755 8. Tremblay KD, Dunn NR, Robertson EJ. Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. *Development* 2001; 128: 3609-3621.
9. Ying Y, Zhao GQ. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev Biol* 2001; 232: 484-492.
- 760 10. Hayashi K, Kobayashi T, Umino T, Goitsuka R, Matsui Y, Kitamura D. SMAD1 signaling is critical for initial commitment of germ cell lineage from mouse epiblast. *Mech Dev* 2002; 118: 99-109.
- 765 11. Chang H, Matzuk MM. Smad5 is required for mouse primordial germ cell development. *Mech Dev* 2001; 104: 61-67.
12. Chu GC, Dunn NR, Anderson DC, Oxburgh L, Robertson EJ. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. *Development* 2004; 131: 3501-3512.
- 770 13. de Sousa Lopes SM, Roelen BA, Monteiro RM, Emmens R, Lin HY, Li E, Lawson KA, Mummery CL. BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 2004; 18: 1838-1849.
14. Saitou M, Barton SC, Surani MA. A molecular programme for the specification of germ cell fate in mice. *Nature* 2002; 418: 293-300.
- 775 15. Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 2005; 436: 207-213.
- 780 16. Saitou M, Payer B, O'Carroll D, Ohinata Y, Surani MA. Blimp1 and the emergence of the germ line during development in the mouse. *Cell Cycle* 2005; 4: 1736-1740.
17. Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* 2005;
- 785

- 278: 440-458.
18. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* 2005; 123: 917-929.
 - 790 19. Brady G, Iscove NN. Construction of cDNA libraries from single cells. *Methods Enzymol* 1993; 225: 611-623.
 20. Dulac C, Axel R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 1995; 83: 195-206.
 - 795 21. Saito H, Kubota M, Roberts RW, Chi Q, Matsunami H. RTP family members induce functional expression of mammalian odorant receptors. *Cell* 2004; 119: 679-691.
 22. Lange U, Saitou M, Western P, Barton S, Surani M. The *Fragilis* interferon-inducible gene family of transmembrane proteins is associated with germ cell specification in mice. *BMC Dev Biol* 2003; 3: 1.
 - 800 23. Tanaka SS, Matsui Y. Developmentally regulated expression of *mil-1* and *mil-2*, mouse interferon-induced transmembrane protein like genes, during formation and differentiation of primordial germ cells. *Mech Dev* 2002; 119 Suppl 1: S261-267.
 - 805 24. Sato M, Kimura T, Kurokawa K, Fujita Y, Abe K, Masuhara M, Yasunaga T, Ryo A, Yamamoto M, Nakano T. Identification of *PGC7*, a new gene expressed specifically in preimplantation embryos and germ cells. *Mech Dev* 2002; 113: 91-94.
 - 810 25. Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor *Oct4*. *Cell* 1998; 95: 379-391.
 26. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on *SOX2* function. *Genes Dev* 2003; 17: 126-140.
 - 815 27. Boiani M, Scholer HR. Regulatory networks in embryo-derived pluripotent stem cells. *Nat Rev Mol Cell Biol* 2005; 6: 872-884.
 28. Frigessi A, van de Wiel MA, Holden M, Svendsrud DH, Glad IK, Lyng H. Genome-wide estimation of transcript concentrations from spotted cDNA microarray data. *Nucleic Acids Res* 2005; 33: e143.
 - 820 29. MacGregor GR, Zambrowicz BP, Soriano P. Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. *Development* 1995; 121: 1487-1496.
 30. Wang C, Lehmann R. *Nanos* is the localized posterior determinant in *Drosophila*. *Cell* 1991; 66: 637-647.
 - 825 31. Subramaniam K, Seydoux G. *nos-1* and *nos-2*, two genes related to *Drosophila nanos*, regulate primordial germ cell development and survival in *Caenorhabditis elegans*. *Development* 1999; 126: 4861-4871.
 32. Tsuda M, Sasaoka Y, Kiso M, Abe K, Haraguchi S, Kobayashi S, Saga Y. Conserved role of *nanos* proteins in germ cell development. *Science* 2003; 301: 1239-1241.
 - 830 33. Youngren KK, Coveney D, Peng X, Bhattacharya C, Schmidt LS, Nickerson ML, Lamb BT, Deng JM, Behringer RR, Capel B, Rubin EM, Nadeau JH, Matin A.

- The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature* 2005; 435: 360-364.
- 835 34. Weidinger G, Stebler J, Slanchev K, Dumstrei K, Wise C, Lovell-Badge R, Thisse C, Thisse B, Raz E. dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. *Curr Biol* 2003; 13: 1429-1434.
- 840 35. Noguchi T, Noguchi M. A recessive mutation (ter) causing germ cell deficiency and a high incidence of congenital testicular teratomas in 129/Sv-ter mice. *J Natl Cancer Inst* 1985; 75: 385-392.
36. Beck AR, Miller IJ, Anderson P, Streuli M. RNA-binding protein TIAR is essential for primordial germ cell development. *Proc Natl Acad Sci U S A* 1998; 95: 2331-2336.
- 845 37. Manova K, Bachvarova RF. Expression of c-kit encoded at the W locus of mice in developing embryonic germ cells and presumptive melanoblasts. *Dev Biol* 1991; 146: 312-324.
38. Bernstein A, Chabot B, Dubreuil P, Reith A, Nocka K, Majumder S, Ray P, Besmer P. The mouse W/c-kit locus. *Ciba Found Symp* 1990; 148: 158-166; discussion 166-172.
- 850 39. Besmer P, Manova K, Duttlinger R, Huang EJ, Packer A, Gyssler C, Bachvarova RF. The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis. *Dev Suppl* 1993: 125-137.
- 855 40. Ara T, Nakamura Y, Egawa T, Sugiyama T, Abe K, Kishimoto T, Matsui Y, Nagasawa T. Impaired colonization of the gonads by primordial germ cells in mice lacking a chemokine, stromal cell-derived factor-1 (SDF-1). *Proc Natl Acad Sci U S A* 2003; 100: 5319-5323.
41. Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A. Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003; 113: 643-655.
- 860 42. Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* 2003; 113: 631-642.
- 865 43. Western P, Maldonado-Saldivia J, van den Bergen J, Hajkova P, Saitou M, Barton S, Surani MA. 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.
- 870 44. Takahashi K, Mitsui K, Yamanaka S. Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. *Nature* 2003; 423: 541-545.
45. Bastian H, Gruss P. A murine even-skipped homologue, Evx 1, is expressed during early embryogenesis and neurogenesis in a biphasic manner. *Embo J* 1990; 9: 1839-1852.
- 875 46. Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* 1990; 343: 617-622.
47. Herrmann BG. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* 1991; 113: 913-917.
48. Crossley PH, Martin GR. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in

- 880 the developing embryo. *Development* 1995; 121: 439-451.
49. Takada S, Stark KL, Shea MJ, Vassileva G, McMahon JA, McMahon AP. Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev* 1994; 8: 174-189.
50. Yamaguchi TP, Takada S, Yoshikawa Y, Wu N, McMahon AP. T (Brachyury) is a
885 direct target of Wnt3a during paraxial mesoderm specification. *Genes Dev* 1999; 13: 3185-3190.
51. Smith DE, Franco del Amo F, Gridley T. Isolation of *Sna*, a mouse gene homologous to the *Drosophila* genes *snail* and *escargot*: its expression pattern suggests multiple roles during postimplantation development. *Development*
890 1992; 116: 1033-1039.
52. Nieto MA. The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Biol* 2002; 3: 155-166.
53. Barrallo-Gimeno A, Nieto MA. The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development* 2005; 132: 3151-3161.
895
54. Shaffer AL, Lin KI, Kuo TC, Yu X, Hurt EM, Rosenwald A, Giltnane JM, Yang L, Zhao H, Calame K, Staudt LM. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* 2002; 17: 51-62.
- 900 55. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; 69: 915-926.
56. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases *Dnmt3a* and *Dnmt3b* are essential for de novo methylation and mammalian development. *Cell* 1999; 99: 247-257.
- 905 57. Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H, Shinkai Y. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* 2002; 16: 1779-1791.
58. Tachibana M, Ueda J, Fukuda M, Takeda N, Ohta T, Iwanari H, Sakihama T,
910 Kodama T, Hamakubo T, Shinkai Y. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev* 2005; 19: 815-826.
59. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. *Drosophila* enhancer of *Zeste*/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 2002; 111: 185-196.
915
60. Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* 2002; 111: 197-208.
61. Cao R, Zhang Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27
920 in histone H3. *Curr Opin Genet Dev* 2004; 14: 155-164.
62. Cao R, Zhang Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* 2004; 15: 57-67.
63. Satijn DP, Hamer KM, den Blaauwen J, Otte AP. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in *Xenopus* embryos. *Mol Cell Biol* 2001; 21: 1360-1369.
925

64. Atchison L, Ghias A, Wilkinson F, Bonini N, Atchison ML. Transcription factor YY1 functions as a PcG protein in vivo. *Embo J* 2003; 22: 1347-1358.
- 930 65. Massague J, Seoane J, Wotton D. Smad transcription factors. *Genes Dev* 2005; 19: 2783-2810.
66. O'Shea JJ, Gadina M, Schreiber RD. Cytokine signaling in 2002: new surprises in the Jak/Stat pathway. *Cell* 2002; 109 Suppl: S121-131.
67. Cech TR. Beginning to understand the end of the chromosome. *Cell* 2004; 116: 273-279.
- 935 68. Bernstein E, Allis CD. RNA meets chromatin. *Genes Dev* 2005; 19: 1635-1655.
69. Kurimoto K, Yabuta Y, Ohinata Y, Ono Y, Uno KD, Yamada RG, Ueda HR, Saitou M. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res* 2006; 34: e42.
- 940 70. Yamaguchi S, Kimura H, Tada M, Nakatsuji N, Tada T. Nanog expression in mouse germ cell development. *Gene Expr Patterns* 2005; 5: 639-646.
71. Hatano SY, Tada M, Kimura H, Yamaguchi S, Kono T, Nakano T, Suemori H, Nakatsuji N, Tada T. Pluripotential competence of cells associated with Nanog activity. *Mech Dev* 2005; 122: 67-79.

945 **Figure Legends**

Figure 1

Performance of the single-cell cDNA amplification method. **A)** Examination of the amplification profiles of *Gapdh* (orange triangles), *Dppa5* (purple triangles), *Sox2* (green diamonds), *Pou5f1* (blue circles) and *Nanog* (red diamonds) from the single-cell-level amount of ES cell total RNA (10 pg) with our method. Estimated transcript copy numbers in the amplified samples (see Materials and Methods) with respect to the amplification cycle numbers were plotted for five genes expressed at different levels in ES cells. **B)** Relative abundance of the transcripts compared to that of *Gapdh* in the amplified samples with different cycle numbers. Note that representation of the genes compared was gradually distorted after the 24th cycle. The color code to represent the genes is the same with that in **A)**. **C)** Transcript representations in the samples amplified by 24 cycles compared with those in unamplified original total RNA. Relative log₁₀-expression levels in amplified samples from nine individual experiments were plotted against those in nonamplified cDNA synthesized from 1 µg of total RNA. The 95% and 99% confidence intervals were plotted as dashed and dotted lines, respectively. Note that most (62%) of the amplified genes are within the two-fold difference lines (green area) from the nonamplified control. Genes examined are from higher expression levels, *Ppia*, *Arbp*, *Gapdh*, *Dppa5*, *Zfp42*, *Pou5f1*, *Sox2*, *Ifitm3*, *Eed*, *Nanog*, *Dnmt3b*, *Yy1*, *Dnmt1*, *Espl1*, *Ezh2*, *Ehmt2*, *Eras*, *Ehmt1*, *Nodal*, *Dppa3* and *Tial1*. **D)** Statistical analysis of the single-cell-level cDNA amplification. The frequencies of the probes within the indicated fold differences from the nonamplified control are shown for both the probes of all expression-level ranges and probes expressed in more than 100 copies per cell. The frequencies and the R^2 values of the detected probes are also shown.

970

Figure 2

Analysis of the *Prdm1*-positive subpopulations during PGC specification. **A)** Images of representative embryos from four different stages used to synthesize single-cell cDNAs. Boxed areas indicate the locations of the dissected fragments where PGCs and their somatic neighbors reside. **B)** A summary of the single-cell analysis and the proportion of the *Prdm1*-positive subpopulations, categorized by the expression of

975

Dppa3(stella) and *Hoxb1* or *Dppa3* and *Sox2*, at the four different stages. **C)** Graphical representation of the transition of the proportion of the *Prdm1*-positive subpopulations classified by the expression of *Dppa3* and *Hoxb1* shown in **B)**. The transition of the proportion of the *Dppa3* (+) and *Hoxb1* (-) cells, *Dppa3* (+) and *Hoxb1* (+) cells, *Dppa3* (-) and *Hoxb1* (+) cells, and *Dppa3* (-) and *Hoxb1*(-) cells are represented by lines with green, sky blue, blue, and orange circles, respectively. **D)** Graphical representation of the transition of the proportion of the *Prdm1*-positive subpopulations classified by the expression of *Dppa3* and *Sox2* shown in **B)**. The transition of the proportion of the *Sox2* (+) cells and *Dppa3* (+) cells are shown with lines with red and blue circles, respectively. The inset represents the ratio of *Dppa3* (+) cells among *Sox2* (+) cells and the ratio of *Sox2* (+) cells among *Dppa3* (+) cells with lines having purple and green diamonds, respectively.

990 **Figure 3**

Gene expression dynamics during the specification of PGCs. The expression levels (approximate copy number per single cell on the log₁₀ scale) of the genes examined in the germline cells (green) and their somatic neighbors (orange) were plotted, along with the average values and standard deviations, with respect to the developmental stages (four stages for the germline cells and three for somatic cells). The significance of the difference in gene expression levels was examined by ANOVA and Student's *t*-test (see Materials and Methods). Genes that showed statistically significant differences by ANOVA ($P < 0.05$) are shown by an asterisk to the left of each gene.

1000 **Figure 4**

A heat-map representation of the gene expression profiles during PGC specification and in the somatic neighbors. The copy number data were converted into a heat map that represents expression levels by colors as shown at the bottom of the figure. Genes were classified into five groups by the patterns of expressions; **A)** housekeeping genes, **B)** genes up-regulated in the germ cell lineage, **C)** genes down-regulated in the germ cell lineage, **D)** genes with no significant differences in expression between the germline and somatic cells, and **E)** genes expressed at low levels in both germline and somatic cells (average $< 10^{0.3}$) (see also Supplemental Table S2). **F)** Unsupervised

hierarchical clustering of all the single cells examined. PGCs at E6.75, E7.25, E7.75,
1010 and E8.25 are indicated with pale blue, pale green, green, and blue, respectively.
Somatic neighbors at E6.75, E7.25, and E7.75 are indicated with magenta, orange, and
red, respectively.

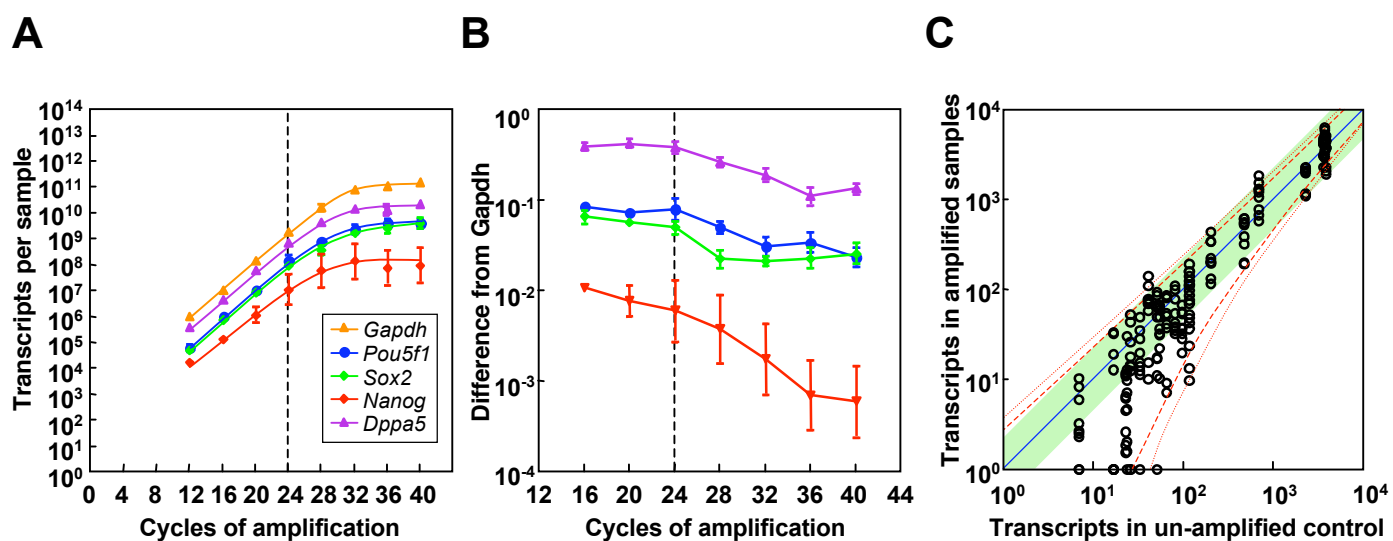
Supplemental Table S1

1015 A list of primer sequences used to analyze the expression levels of genes examined in
this study. Amplicon sizes, amplification slopes and intercept values are also provided
(see Materials and Methods).

Supplemental Table S2

1020 Statistical results of the gene expression differences analyzed by ANOVA and Student's
t-test (see Materials and Methods). Bolded values represent statistically significant
values ($P < 0.05$). Green color represents up-regulation in the germline and red color
represents genes up-regulation in the somatic neighbors.

Figure 1, Yabuta et al.



D

	< 2	< 4	< 10	Detected	R ²
Total range	62% (118/189)	84% (159/189)	90% (170/189)	90% (171/189)	0.884
> 10 ² copies/cell	81% (66/81)	96% (78/81)	99% (80/81)	100% (81/81)	0.883

