Cell Reports
Report

Cell PRESS

Open ACCESS

Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline In Vivo

Michael Sachs,[1,2,3] Courtney Onodera,[2,4] Kathryn Blaschke,[1,2,3] Kevin T. Ebata,[2,3] Jun S. Song,[2,4,5] and Miguel Ramalho-Santos[2,3,*]

[1]Biomedical Sciences Graduate Program
[2]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research
[3]Department of Obstetrics, Gynecology and Reproductive Science
[4]Institute for Human Genetics
[5]Department of Epidemiology and Biostatistics
35 Medical Center Way, University of California, San Francisco, CA 94143, USA
*Correspondence: mrsantos@diabetes.ucsf.edu
http://dx.doi.org/10.1016/j.celrep.2013.04.032

SUMMARY

Developmental regulatory genes have both activating (H3K4me3) and repressive (H3K27me3) histone modifications in embryonic stem cells (ESCs). This bivalent configuration is thought to maintain lineage commitment programs in a poised state. However, establishing physiological relevance has been complicated by the high number of cells required for chromatin immunoprecipitation (ChIP). We developed a low-cell-number chromatin immunoprecipitation (low-cell ChIP) protocol to investigate the chromatin of mouse primordial germ cells (PGCs). Genome-wide analysis of embryonic day 11.5 (E11.5) PGCs revealed H3K4me3/H3K27me3 bivalent domains highly enriched at developmental regulatory genes in a manner remarkably similar to ESCs. Developmental regulators remain bivalent and transcriptionally silent through the initiation of sexual differentiation at E13.5. We also identified >2,500 "orphan" bivalent domains that are distal to known genes and expressed in a tissue-specific manner but silent in PGCs. Our results demonstrate the existence of bivalent domains in the germline and raise the possibility that the somatic program is continuously maintained as bivalent, potentially imparting transgenerational epigenetic inheritance.

INTRODUCTION

Pluripotency is dependent on the maintenance of a proper epigenetic landscape (Gaspar-Maia et al., 2011; Orkin and Hochedlinger, 2011). Bivalent domains, which are defined by the paradoxical coexistence of a permissive histone mark (H3K4me3) and a repressive mark (H3K27me3), are thought to play an important role in pluripotency by keeping developmental genes in a silenced state poised for activation upon differentiation (Azuara et al., 2006; Bernstein et al., 2006). In support of this model, bivalent domains have been shown to be present preferentially at developmental regulatory genes in undifferentiated embryonic stem cells (ESCs) and several adult tissues in vivo including sperm, testis, the cerebellum, and the hematopoietic compartment (Mikkelsen et al., 2007; Cui et al., 2009, 2012; Hammoud et al., 2009).The moderate derepression of developmental regulators in eed mutant ESCs, which have reduced levels of H3K27me3, further supports this model (Azuara et al., 2006; Boyer et al., 2006). However, it has recently been proposed that the concomitant presence of H3K4me3 and H3K27me3 at developmental genes in ESCs is an in vitro artifact resulting from suboptimal culture conditions (Hong et al., 2011; Marks et al., 2012). In addition, direct evidence for bivalency in the developing mammalian embryo is limited and conflicting due to technical difficulties associated with the low amounts of material available (Alder et al., 2010; Dahl et al., 2010; Rugg-Gunn et al., 2010). The universal nature of bivalency has been further questioned due to conflicting reports from nonmammalian species, with bivalent domains demonstrated to exist in zebrafish (Vastenhouw et al., 2010) but not detected in Xenopus or Drosophila embryos (Akkers et al., 2009; Schuettengruber et al., 2009). We therefore lack a clear understanding of whether bivalency exists in embryonic cells in vivo and how it relates to pluripotency.

Of particular interest to pluripotency are primordial germ cells (PGCs), the embryonic precursors of the germline. PGCs have a transcriptional profile similar to ESCs (Qin et al., 2012), including the silencing of developmental regulators, and can give rise to pluripotent stem cells when cultured in vitro. However, unlike ESCs, PGCs are unipotent and unable to contribute to chimeras. Additionally, PGCs are thought to undergo a period of "epigenetic erasure" or reprogramming via global removal of both DNA methylation and histone modifications that may prevent transmission of epimutations (Greer and Shi, 2012; Seisenberger et al., 2013). Due to the low number of PGCs present during development, nearly all evidence for this model is from immunofluorescence (IF)-based methodologies that lack gene level resolution and have shown conflicting results (Seki et al., 2007;

CrossMark

Hajkova et al., 2008; Kagiwada et al., 2013). To characterize the chromatin state of PGCs, we developed a low-cell-number chromatin immunoprecipitation (low-cell ChIP) protocol for the analysis of histone marks using less than 10,000 cells per IP without the need for carrier chromatin or preamplification. Using this technique, we performed ChIP sequencing (ChIP-seq) and ChIP-quantitative PCR (ChIP-qPCR) to show that bivalent domains are present at developmental regulatory genes at multiple stages of PGC development.

RESULTS

Low-Cell ChIP Protocol

We developed a low-cell ChIP protocol by using a magnetic-bead-based capture with key improvements over existing approaches (O'Neill et al., 2006; Acevedo et al., 2007; Dahl and Collas, 2008; Adli and Bernstein, 2011; Shankaranarayanan et al., 2011), including reduced handling of material for crosslinking, improvements to sonication reproducibility, and no requirement for carrier DNA or DNA preamplification (see Experimental Procedures). We validated the protocol using a transgenic mouse ESC line expressing an N-terminal myc-tagged H3.3 histone, due to technical simplicity and the availability of genome-wide data (Goldberg et al., 2010). ChIP was performed for myc-H3.3 and a control immunoglobulin G (IgG) with decreasing amounts of chromatin starting from 2.5 million to 12,500 cells per IP (Figure S1A). qPCR was used to interrogate the enrichment status of 15 different loci representing six categories of genomic regions. Importantly, 11 of the 12 loci expected to be bound by H3.3 (Goldberg et al., 2010) were as least 10-fold more enriched for myc-H3.3 compared to IgG, demonstrating the specificity and low background of the assay. Additionally, the relative enrichment between loci remains stable independent of starting cell number, demonstrating the robustness and sensitivity of the assay (Figure S1B). We envision that this protocol will be broadly useful to characterize the chromatin state of rare stem and precursor cell populations in vivo.

Genome-wide Analysis of H3K4me3 and H3K27me3 in PGCs

To investigate the presence of bivalent domains in PGCs, we used a transgenic *Oct4:GFP* mouse line to isolate PGCs (GFP+) and the surrounding somatic cells (GFP−) by fluorescence-activated cell sorting (FACS) at different stages of development as previously reported (Yoshimizu et al., 1999). Using our low-cell ChIP protocol, we performed ChIP-seq for H3K4me3 and H3K27me3 in two biological replicates of E11.5 PGCs, obtaining a total of approximately 386 million reads and between 15 and 32 million uniquely mappable reads after filtering per ChIP library (Table S1). The two biological replicates in our ChIP-seq libraries show very strong correlation, with Pearson correlation coefficients of 0.978 for H3K4me3 and 0.996 for H3K27me3 (Figures 1A and S2A; Table S1), indicative of the high reproducibility of the protocol. Furthermore, all IPs are enriched for signal over input as assessed using CHANCE (Table S1) (Diaz et al., 2012).

We identified H3K4me3-only regions, H3K27me3-only regions, and bivalent domains in both E11.5 PGCs (this study) and ESCs (Mikkelsen et al., 2007) using ChromHMM (Ernst

and Kellis, 2012). The overall distribution of these histone marks is similar between PGCs and ESCs, except for an increase in H3K27me3-marked DNA in PGCs (Figure S2B), which is in agreement with IF data (Seki et al., 2005, 2007; Kagiwada et al., 2013). Genes marked only with H3K27me3 in PGCs (Table S5) are involved in cell motility (Figure S2C), which may be a consequence of PGCs having just completed their migration to the gonads at E11.5.

Bivalent Genes in PGCs Are Enriched for Developmental Regulators

We next focused on bivalent regions. Consistent with their association with gene promoters, bivalent domains largely occur at transcriptional start sites (TSSs) in both E11.5 PGCs and ESCs (Figure S2D). E11.5 PGCs and ESCs have a similar distribution and a highly significant overlap of genes marked as bivalent (p < 2.2 × 10⁻¹⁶, Figures 1B, 1C, and S2E). Similarly to what has been described for ESCs (Bernstein et al., 2006), the 2,715 genes marked as bivalent in PGCs and ESCs are strongly enriched for regulators of early development (Figure 1D). We note that while this manuscript was under review, Ng et al. reported a ChIP-seq data set for several histone marks in PGCs, including H3K4me3 and H3K27me3 (Ng et al., 2013). However, the authors did not analyze bivalency in detail and the low signal for H3K4me3 in E11.5 PGCs in that study precludes a direct comparison with our work (Figure S2A). Therefore, we sought to validate our data using ChIP-qPCR in independent biological samples of PGCs. ChIP-qPCR at 12 genes representing eight developmental pathways confirmed that all are bivalently marked in PGCs, with bivalency defined as enrichment above a background of 10% (Table S4) for both H3K4me3 and H3K27me3 and a maximum 2-fold difference in enrichment between the two marks (Figure 1E). These bivalent regulators include transcription factors that are important for the specification of the three somatic germ layers, such as *Pax6, Nkx2.2, MyoD, brachyury (T), Gata6* and *FoxA1, and Hox* genes such as *HoxA3* and *HoxB9*. This is notable because ESCs are pluripotent and expected to maintain these genes in a poised transcriptional state for activation upon differentiation, but PGCs are unipotent and will not express any of these developmental regulators until after fertilization, in the next generation. In fact, the *Hox* cluster has been shown to be silenced as part of the early specification of PGC precursors at E7.25 (Saitou et al., 2002). A comparison of the expression pattern of genes bivalent in PGCs and ESCs demonstrates that these genes are transcriptionally repressed in ESCs and PGCs, but subsets are active in select lineage-restricted cell types such as mouse embryonic fibroblasts (MEFs), bone marrow, or brain tissue (Figure 1F). Taken together, the results of our genome-wide analysis reveal that the mouse embryonic germline in vivo is remarkably similar to cultured ESCs with regards to the presence of bivalent chromatin at silenced developmental regulatory genes.

PGCs Are Enriched for Bivalency at Developmental Genes Compared to the Surrounding Soma

We next compared the chromatin landscape of PGCs to the surrounding soma. Using ChIP-qPCR, we examined the 12 previously tested developmental regulators plus *Pou5f1/Oct4*
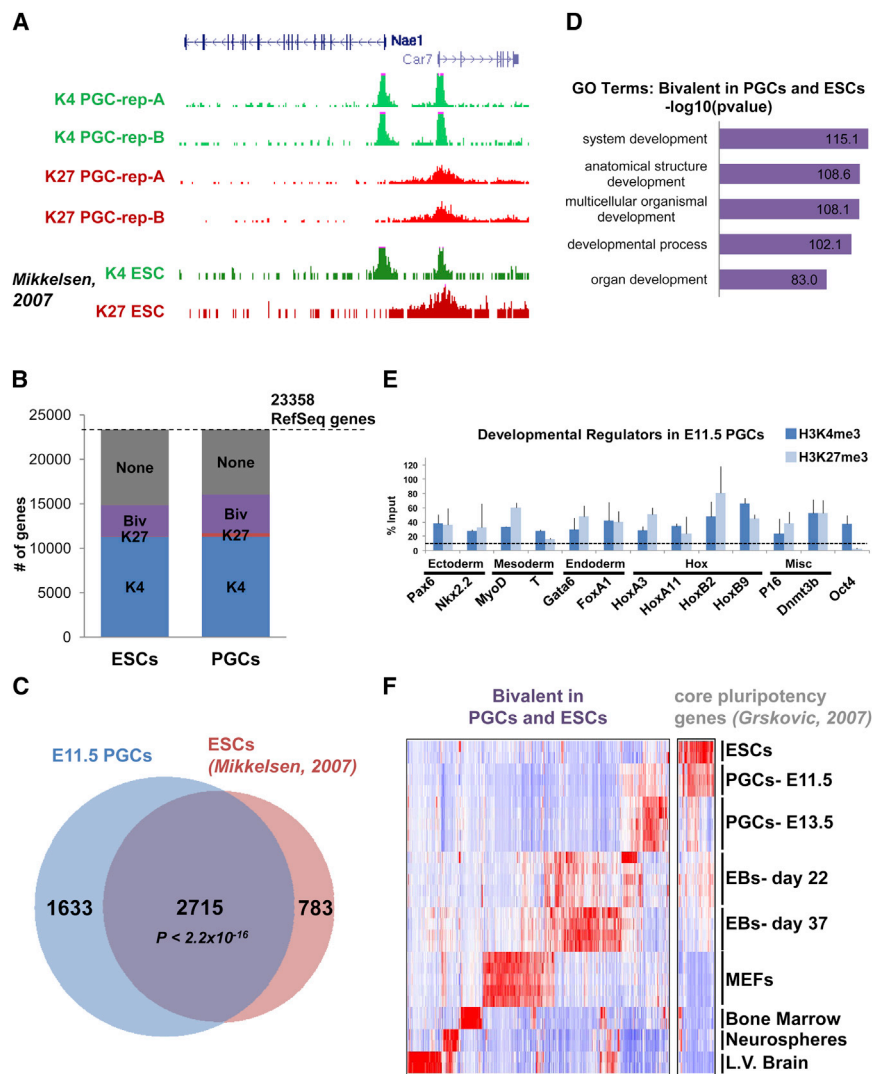
Figure 1. Bivalent Genes in E11.5 PGCs Are Enriched for Developmental Regulators

(A) Sample UCSC Genome Browser view of ChIP-seq signal for two biological replicates of E11.5 PGCs and for ESCs.

(B) E11.5 PGCs and ESCs have a similar number of genes marked by only H3K4me3, only H3K27me3, or both.

(C) Overlap of genes marked by H3K4me3 and H3K27me3 in E11.5 PGCs and ESCs, $p < 2.2 \times 10^{-16}$ using Fisher's exact test.

(D) Top five biological process gene ontology terms as determined using DAVID for genes marked as bivalent in both PGCs and ESCs.

(E) Validation of ChIP-seq using ChIP-qPCR enrichment for H3K4me3 and H3K27me3 in E11.5 PGCs for a subset of bivalent genes representing regulators of the three germ layers. Data are mean percentage of input ±SD of enrichment. The dotted line represents level of background enrichment.

(F) Heatmap showing expression of the subset of genes marked as bivalent in both PGCs and ESCs in indicated cell types. EBs, embryonic bodies; MEFs, mouse embryonic fibroblasts; L.V. Brain, lateral ventricles of adult brain. Blue indicates downregulation; red indicates upregulation.

See also Figures S1 and S2 and Tables S1, S4, and S5.

in ESCs and E11.5 soma (Figures 2A and 2B). Twelve of these developmental regulators are bivalent in both E11.5 PGCs and ESCs. However, in the soma most of these genes have stronger enrichment for either H3K4me3 or H3K27me3 and only four genes are bivalent (Figures 2B and 2C). These data indicate that developmental regulators of all somatic lineages exist in a bivalent state in ESCs and PGCs but not in somatic cells in vivo (Figure 2D), suggesting that this distribution of bivalent domains is a distinguishing characteristic of pluripotency-associated cells both in vitro and in vivo.
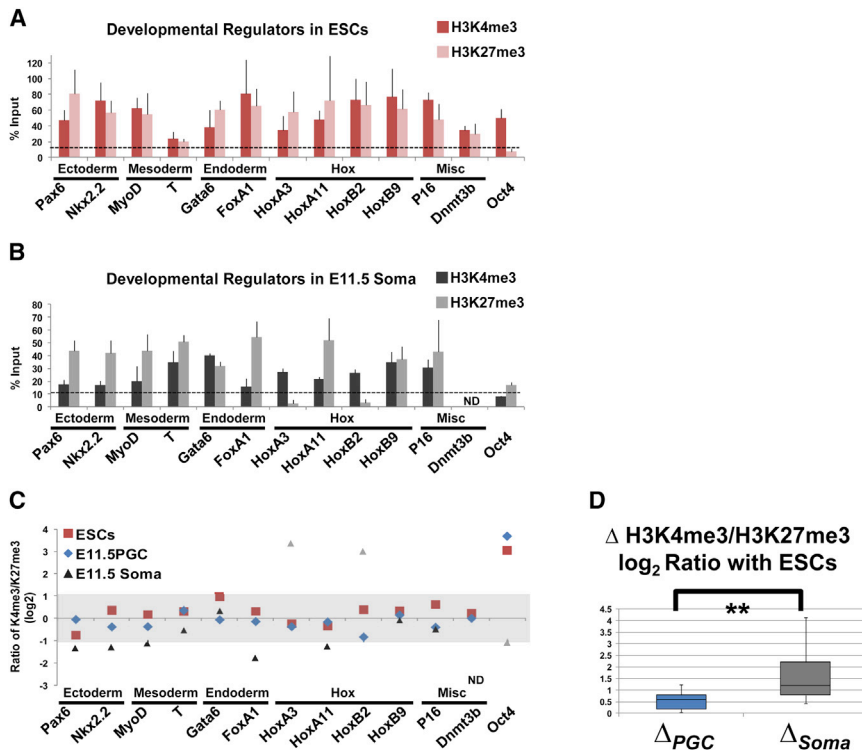
Developmental Regulators Remain Bivalent in PGCs during the Initiation of Sexual Differentiation

E11.5 PGCs are sexually indifferent and capable of giving rise to pluripotent stem cells when cultured in vitro. After E11.5, PGCs initiate the process of sexual differentiation and express sex-specific genes by E13.5. Furthermore, E13.5 PGCs can no longer give rise to pluripotent stem cells (Labosky et al., 1994). We therefore sought to investigate whether the process of sexual

differentiation and concomitant loss of the ability to give rise to pluripotent stem cells affects the bivalent chromatin state of PGCs. Using low-cell ChIP-qPCR, we examined eight representative developmental genes through a time course of PGC development from E11.5 to E13.5. In agreement with previously published data (Qin et al., 2012), somatic developmental genes, including *HoxA11*, *HoxB9*, *Pax6*, *Nkx2.2*, *T*, *MyoD*, *Gata6*, and *FoxA1*, are expressed at very low to undetectable levels from E11.5 through E13.5 (Figure 3). Interestingly, these developmental regulators remain bivalent in PGCs throughout this time period (Figure 3). In contrast, germline-specific genes, such as *Dazl*, *Dppa3*, and *Vasa*, are enriched for only H3K4me3, coinciding with their transcriptional activation (Figure S3). Taken together, these data indicate that somatic developmental regulators remain bivalent in the mouse embryonic germline from E11.5 through initiation of sexual differentiation at E13.5.

Orphan Bivalent Domains Are Silent in PGCs and Transcribed in a Tissue-Specific Manner

Surprisingly, we identified a large number of "orphan" bivalent domains that cannot be associated with any annotated promoter or gene in the RefSeq database and are distal to annotated genes by a median distance of 13 kb (Figure 4A). From a total of 9,132 bivalent domains in E11.5 PGCs, 2,886 are orphan, and many are also bivalent in ESCs (Table S2). Bivalent

Figure 2. PGCs Are Enriched for Bivalent Developmental Regulators Compared to Surrounding Somatic Cells at E11.5

(A) ChIP enrichment for H3K4me3 and H3K27me3 in ESCs at the indicated genes. Data are mean percentage of input ±SD of enrichment. The dotted line represents level of background enrichment.

(B) ChIP enrichment for H3K4me3 and H3K27me3 in E11.5 soma at the indicated genes as described in (A). Data are mean percentage of input ±SD of enrichment.

(C) The ratio of H3K4me3 to H3K27me3 $\log_2$ transformed for the indicated gene for ESCs, E11.5 PGCs, and E11.5 soma. The gray area indicates a ratio less than 2 and greater than 0.5. Genes where enrichment for either H3K4me3 or H3K27me3 is not above background are lighter in color.

(D) The $\log_2$ of the H3K4me3/H3K27me3 ratio was calculated as in (C) ($\log_2 R_{cell\ type}$) for all genes examined independent of enrichment over background. The differences between cell type ratios were calculated as follows: $\Delta_{PGC} = |\log_2 R_{PGC} - \log_2 R_{ESC}|$, and $\Delta_{soma} = |\log_2 R_{soma} - \log_2 R_{ESC}|$ for each gene. Box plots represent the distribution of the differences. Whiskers show the highest and lowest value. Statistical significance was assessed using the Wilcoxon matched-pairs signed rank test (one-sided), **p = 0.001. ND, not determined.

See also Figure S1 and Table S4.

domains are known to have a strong association with unmethylated CpG islands (CGIs) (Ku et al., 2008). We found that 1,413 (49.0%, p value < 1 × 10^{-5}) of the orphan bivalent domains in PGCs map to experimentally defined unmethylated CGIs (Illingworth et al., 2010). We next investigated the expression status of the regions containing orphan bivalent domains. If the orphan bivalent domains were similar to bivalent genes (Figure 1F), they would be silent in PGCs and expressed in somatic tissues. Indeed, only 23 (0.80%) orphan bivalent domains show signs of transcription in an RNA-seq data set from E11.5 PGCs (Seisenberger et al., 2012). Analysis of CAGE tag clusters showed that 927 of the 1,413 (65.6%, p value = 5.78 × 10^{-168}) orphan bivalent domains with CGIs are expressed in at least 1 of 22 tissue types (Figure 4B; Table S3) (Kawaji et al., 2009). Similarly, 461 of the 1,473 (31.3%, p value = 3.61 × 10^{-35}) orphan bivalent domains without CGIs are expressed in at least 1 of 22 tissue types (Figure 4B; Table S3). Furthermore, both classes of orphan bivalent domains (with or without CGI) display highly tissue-specific expression (Figure 4C), more so than bivalent developmental regulators (p value = 1.1 × 10^{-141}, 2.3 × 10^{-72} respectively), with a preferential activation in brain regions (Figure 4E). Interestingly, we found bivalent domains in PGCs at 300 noncoding RefSeq genes and 179 putative noncoding RNAs inferred from chromatin state (Guttman et al., 2009) and expression data (Carninci et al., 2005) (Figures 4D and S2F). These data suggest that, similar to protein-coding developmental regulators, some noncoding RNAs are bivalent in the germline and poised for activation upon lineage commitment.

DISCUSSION

In this work, we describe an optimized ChIP protocol suitable for low numbers of cells and its use to examine the genome-wide and gene-specific distribution of H3K4me3 and H3K27me3 in mouse PGCs. We report that developmental regulatory genes remain bivalent and transcriptionally silent in vivo in PGCs, but not in adjacent somatic cells, throughout E11.5–E13.5, in a manner highly similar to cultured ESCs. In addition to developmental genes, we identify some ~3,000 orphan bivalent domains that are enriched for CGIs and are expressed in a tissue-specific manner. Some of the bivalent domains identified here correspond to noncoding RNAs that we speculate may have developmental functions, similar to bivalent genes, although this remains to be determined. The findings presented here represent strong evidence for the existence of bivalent domains in the embryonic germline and raise a number of important questions.

Do bivalent promoters escape the epigenetic reprogramming reported to occur in PGCs from E8.5–E13.5? Our data support the suggestion that H3K27me3 may act as a potential mechanism to compensate for the loss of DNA methylation or H3K9me2/3 in midgestation PGCs (Sasaki and Matsui, 2008). It is also possible that the loss of certain histone marks observed by IF may occur primarily at abundant repetitive sequences of the genome and mostly spare unique genes, such as developmental regulators. The application of the low-cell ChIP protocol reported here should help answer this question for other histone marks and time points.
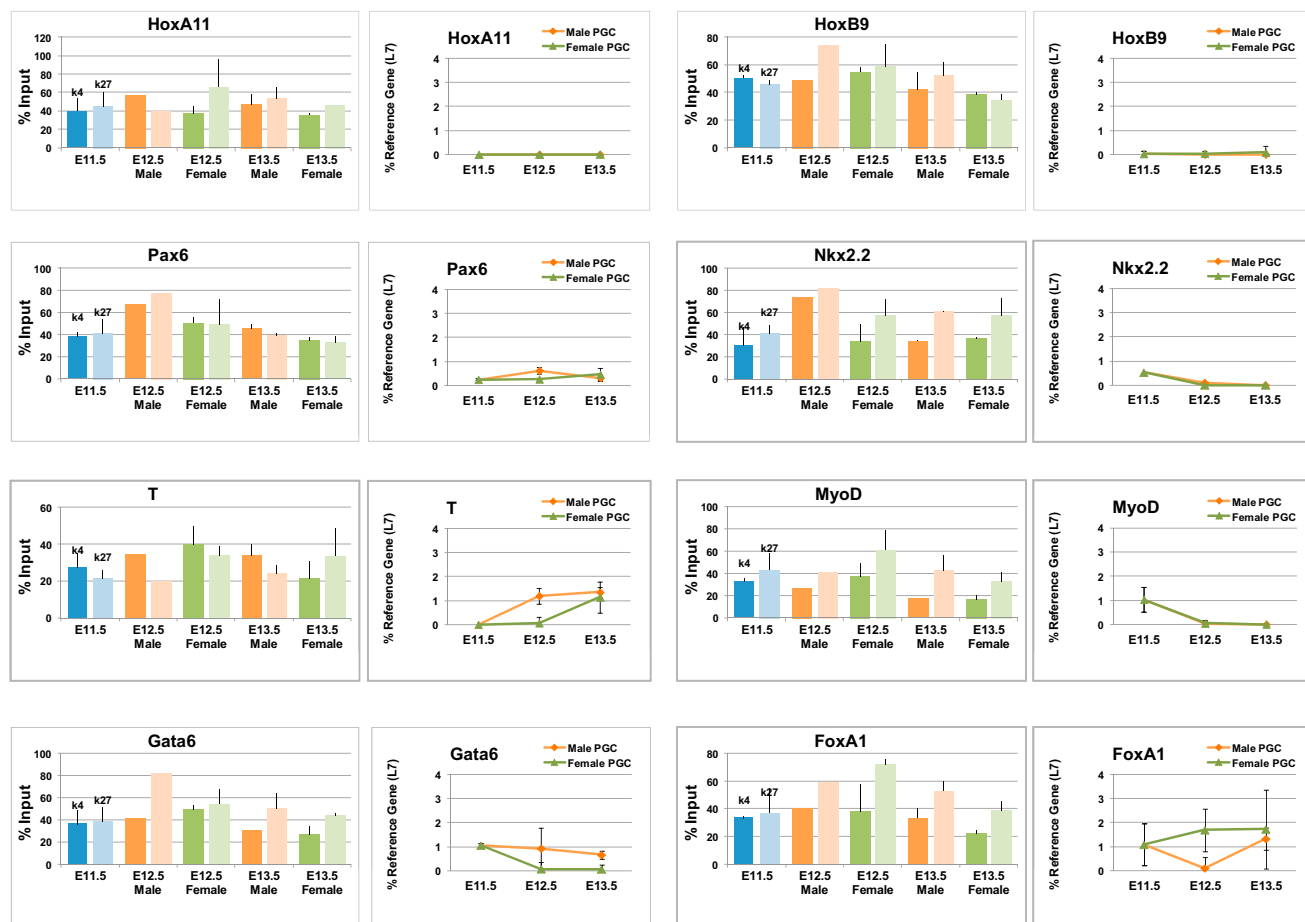
Figure 3. Developmental Regulators Remain Bivalent and Transcriptionally Repressed during Sexual Differentiation in the Germline
Developmental genes remain bivalent and transcriptionally silent from E11.5 to E13.5 in both the male and female germline. ChIP enrichment for H3K4me3 and H3K27me3 in E11.5–13.5 PGCs at the indicated genes is represented as the mean percent of input ±SD for at least two biological replicates. Transcription was measured with qRT-PCR for the indicated genes and represented as a percent of the housekeeping gene *L7* ±SD. At all stages of PGC development analyzed, somatic developmental genes have a similar level of enrichment for H3K4me3 and H3K27me3 and very low expression. See also Figures S1 and S3.

Does the germline continuously maintain the somatic program in a bivalent state until activation in the next generation? It is unclear why such a large set of important developmental regulators would remain bivalent in the developing germline, but not in somatic cells. The hypothesis that bivalency maintains developmental genes in a transcriptionally poised state is intuitive in pluripotent cells that activate these genes upon differentiation, such as ESCs (Azuara et al., 2006; Bernstein et al., 2006) or the epiblast (Rugg-Gunn et al., 2010). The presence of bivalent domains in PGCs may contribute to preventing expression of the somatic program in the germline. Our observations suggest that developmental regulators may be kept in a repressed but accessible state in the germline for activation postfertilization, in the next generation. The transmission of bivalency through the germline could provide a substratum for epigenetic inheritance. Surprisingly, recent work indicates that the sperm genome maintains a residual level of nucleosomes, and that these are enriched for H3K4me3/H3K27me3 bivalent marks at developmental regulators (Hammoud et al., 2009). It will be inter-

esting to determine whether bivalency is detected at other stages of germline development and in oocytes. Functional studies of regulators of bivalency in PGCs should provide important insights into these questions.

EXPERIMENTAL PROCEDURES

Isolation of PGCs

All experiments with mice followed guidelines of the UCSF institutional animal care and use committee. Male B6 mice homozygous for a transgenic *Oct4Δ* PE:*GFP* reporter were crossed with Swiss-Webster females. The gonadal regions from multiple embryos were isolated and pooled prior to enzymatic dissociation for FACS.

Low-Cell ChIP-qPCR

Cells were pooled into batches of ~50,000 cells and crosslinked in 1% formaldehyde (Sigma F8775-25ml) in PBS for 5 min at room temperature and quenched with 125 mM glycine. For each IP, 11 μl of protein A Dynabeads (Life Technologies 1001D) were preincubated with 2.4 μg of either anti H3K4me3 (Diagenode pAb003-050), H3K27me3 (Millipore 07-449), or IgG (Abcam ab46540) in ChIP lysis buffer. Crosslinked cells were lysed in Lo-Bind
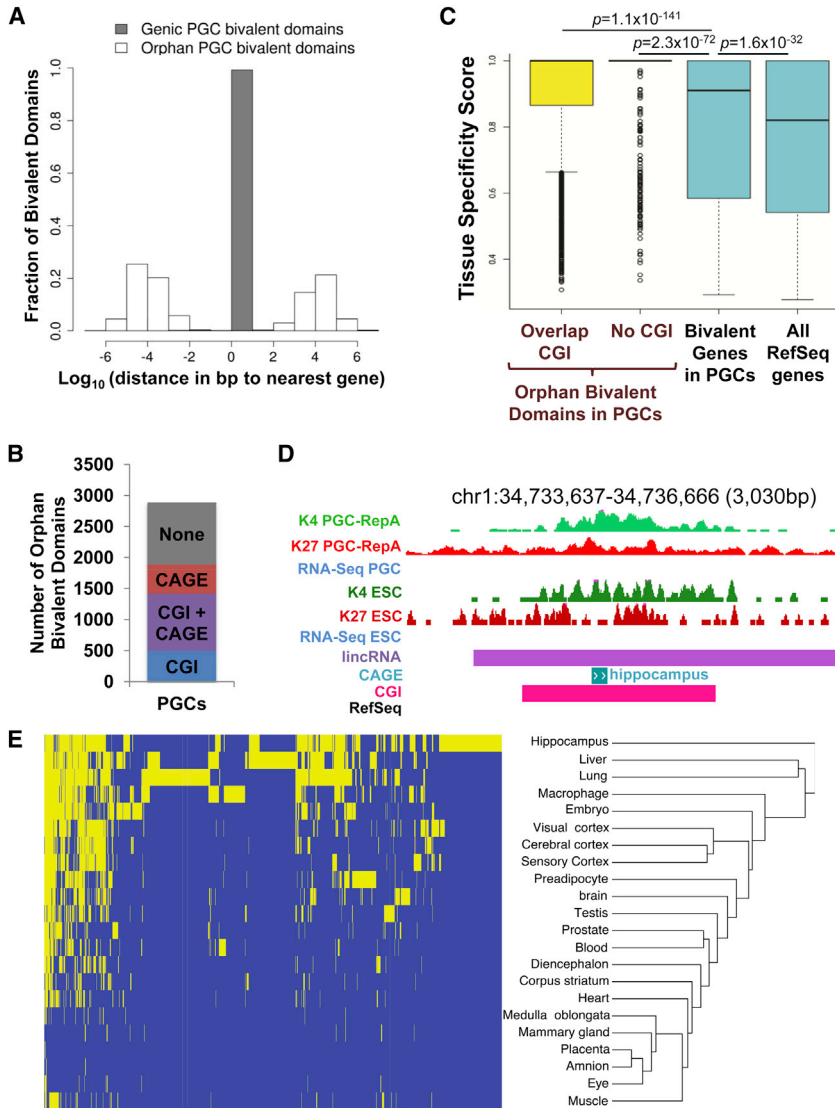
Figure 4. Orphan Bivalent Domains Overlap CpG Islands and Show Tissue-Specific Expression

(A) Distances of genic (assigned to gene, dark gray bars) and orphan PGC bivalent domains (white bars) to nearest gene, with transparent overlay. Distances are presented as the log base 10 of the genomic length in base pairs from bivalent domains to either their nearest gene, with bivalent domains upstream of their nearest gene assigned negative distances after logarithm. Bivalent domains occurring within a gene were assigned a genomic length of 1. TSSs were defined as the starting positions of RefSeq genes.

(B) Number of orphan bivalent domains in E11.5 PGCs with characteristics of transcription start site activity: occurring in unmethylated CpG islands (CGI) identified using CAP-seq (Illingworth et al., 2010), with a CAGE expression tag in 1 or more of 22 tissue types (Kawaji et al., 2009), or both.

(C) Distributions of the tissue-specificity scores of CAGE tag cluster expression in orphan bivalent domains with CGI in PGCs, orphan bivalent domains without CGI in PGCs, bivalent RefSeq genes, and all RefSeq genes. Whiskers indicate either 1.5 times the interquartile range (IQR) from the box or the extreme data points if those points are within 1.5 IQR. The p values are from the Wilcoxon rank sum test.

(D) UCSC Genome Browser view of an orphan bivalent domain in PGCs and ESCs (Mikkelsen et al., 2007) that overlaps a CGI (Illingworth et al., 2010) and a putative noncoding RNA (Guttman et al., 2009). RNA-seq data indicate this region is transcriptionally silent in PGCs (Seisenberger et al., 2012) and ESCs (Marks et al., 2012), while CAGE tag data indicate expression in the hippocampus.

(E) Hierarchical clustering of tissues based on the binary (on or off) expression status of CAGE tag clusters in regions marked as orphan bivalent domains in PGC. Yellow means expression and blue mean no expression.

See also Tables S2 and S3.

microfuge tubes (Eppendorf 022431021) followed by a 40 min sonication in a BioRuptor sonicator (Diagenode UCD-200) set to high power, 7 min ON, 15 min OFF, changing ice/water slurry every 10 min.

Lysate was diluted in lysis buffer and cleared of debris. Cleared lysate from ~50,000 cells was divided into four equal aliquots, and one aliquot used for input. The remaining three aliquots (equivalent to ~12,500 cells each) were used for IP of H3K4me3, H3K27me3, and IgG. Lysates were incubated with preformed bead/antibody overnight at 4°C with mixing. The chromatin/bead/antibody complexes were washed sequentially with lysis buffer three times, DOC buffer once, and TE buffer once, followed by a transfer to a fresh PCR tube. Chromatin was eluted using elution buffer (1% SDS, 0.1 M NaHCO3). IP and input samples were treated with RNaseA followed by Proteinase K treatment. Crosslinking was reversed by incubating overnight at 65°C while shaking. DNA purification was done using a QiaQuick PCR Purification Kit (QIAGEN 28104). E11.5 ChIP data are representative of three to seven biological replicates except for Dnmt3b, which was technically replicated. Additional PGC data represent a minimum of two biological replicates except for male E12.5, which was technically replicated. All primers are listed in Table S4. A full protocol is located in the Extended Experimental Procedures.

Low-Cell ChIP-Seq

ChIP material was obtained and processed as described above with the following modifications. Cells were crosslinked in 0.25% formaldehyde (Thermo Scientific 28906) in PBS for 10 min at room temperature prior to quenching with 125 mM glycine. Crosslinked material was sonicated using a Covaris sonicator for 12 min at duty 5%, intensity 3, and bursts 200. Two aliquots of E11.5 PGCs, consisting of 104,000 and 97,000 cells, were each divided equally to perform ChIP for H3K4me3, H3K27me3, and an input control. IP-DNA and inputs were purified using a MinElute PCR Purification Kit (QIAGEN 28004) and libraries generated using the ThruPLEX-FDPrep Kit (Rubicon Genomics R40048) with 20 cycles of amplification for IP-DNA and 15 cycles for input DNA.

ChIP-Seq Analysis

Reads from each ChIP-seq library were filtered to retain only unique sequences (Table S1). Reads were aligned to the mm9/NCBI build 37 mouse genome using bowtie and reads mapping only once were retained. Data for ESCs were handled in the same manner, obtained from GEO (accession number GSE12241) (Mikkelsen et al., 2007). To call and compare bivalent domains between the PGC and ESC data sets, a 13 state segmentation of the genome was generated using ChromHMM (v.1.06) (Ernst and Kellis, 2012). The four

samples (PGC H3K4me3, PGC H3K27me3, ESC H3K4me3, ESC H3K27me3) were treated as distinct marks from a single cell type to allow a direct and unbiased comparison of the genome segmentation in PGCs and ESCs. K-means clustering was used to group the state emission parameters for each sample into two groups ("on" or "off"), and these were used to classify each of the 13 ChromHMM segmentation states as "bivalent," "H3K4me3 only," "H3K27me3 only," or "none" for each cell type. Data from the replicate PGC libraries were pooled together for this analysis. H3 ChIP-seq data from the Mikkelsen data set was used as control for the ESC samples.

All gene set analysis is based on the mm9 RefSeq gene set. To ensure a conservative segregation of bivalent domains associated with regulation of annotated genes from orphan bivalent domains, genes were considered bivalent if an H3K4me3/H3K27me3 enriched region fell in the range 0.2 kb upstream of the TSS through the transcription end site (TES) of any isoform of the gene. H3K4me3/H3K27me3 enriched regions occurring outside the range from 0.2 kb upstream of the TSS through the TES of any isoform of a gene were called orphans. Genes were considered H3K4me3-only if no isoforms were bivalent and if any isoform had H3K4me3-only regions falling in the window from −0.2 kb upstream to 0.2 kb downstream of the TSS. Similarly, genes were considered to be H3K27me3-only if they were neither bivalent nor H3K4me3-only and if any isoform had regions of H3K27me3-only occurring in the ±0.2 kb window. Analysis of enrichment for Gene Ontology terms was done using DAVID (Huang et al., 2009).

CAGE Analysis

CAGE tag cluster locations in mm9 and tags per million (tpm) values for 22 tissue types were obtained from FANTOM 4 (Kawaji et al., 2009). To assess the statistical significance of the accumulation of CAGE tags, the probability p_i of observing a CAGE tag within 200 bp of each orphan bivalent domain was first estimated based on CAGE tag representation in the region X_i containing the bivalent domain extended 10 kb in each direction. Let Y_i be the indicator random variable that indicates whether the observed bivalent domain in X_i had a detectable CAGE tag cluster within 200 bp. Assuming that Y_i are independent Bernoulli random variables with probability p_i, the sum of Y_i is approximately normal with mean $\Sigma_i p_i$ and variance $\Sigma_i p_i (1-p_i)$, via the Lyapunov central limit theorem (see Extended Experimental Procedures).

Tissue-Specificity Calculation

For each CAGE tag cluster, tissue specificity score was computed based on the Jensen-Shannon divergence between the relative abundance of tpm values across the tissue types and the extreme distribution of being expressed in only one tissue type where the tag cluster has the greatest expression value (Cabili et al., 2011).

ACCESSION NUMBERS

The Gene Expression Omnibus accession number for the ChIP-seq data reported in this paper is GSE46396.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.04.032.

LICENSING INFORMATION

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ACKNOWLEDGMENTS

REFERENCES

Acevedo, L.G., Iniguez, A.L., Holster, H.L., Zhang, X., Green, R., and Farnham, P.J. (2007). Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques 43, 791–797.

Adli, M., and Bernstein, B.E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. Nat. Protoc. 6, 1656–1668.

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., François, K.-J., Stunnenberg, H.G., and Veenstra, G.J.C. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev. Cell 17, 425–434.

Alder, O., Lavial, F., Helness, A., Brookes, E., Pinho, S., Chandrashekran, A., Arnaud, P., Pombo, A., O'Neill, L., and Azuara, V. (2010). Ring1B and Suv39h1 delineate distinct chromatin states at bivalent genes during early mouse lineage commitment. Development 137, 2483–2492.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. 8, 532–538.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349–353.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science 309, 1559–1563.

Cui, K., Zang, C., Roh, T.-Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell 4, 80–93.

Cui, P., Liu, W., Zhao, Y., Lin, Q., Zhang, D., Ding, F., Xin, C., Zhang, Z., Song, S., Sun, F., et al. (2012). Comparative analyses of H3K4 and H3K27 trimethylations between the mouse cerebrum and testis. Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute 10, 82–93.

Dahl, J.A., and Collas, P. (2008). A rapid micro chromatin immunoprecipitation assay (microChIP). Nat. Protoc. 3, 1032–1045.

Dahl, J.A., Reiner, A.H., Klungland, A., Wakayama, T., and Collas, P. (2010). Histone H3 lysine 27 methylation asymmetry on developmentally-regulated

promoters distinguish the first two lineages in mouse preimplantation embryos. PLoS ONE 5, e9150.

Diaz, A., Nellore, A., and Song, J.S. (2012). CHANCE: comprehensive software for quality control and validation of ChIP-seq data. Genome Biol. 13, R98.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. 12, 36–47.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140, 678–691.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. 13, 343–357.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Moran, N., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature 452, 877–881.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature 460, 473–478.

Hong, S.-H., Rampalli, S., Lee, J.B., McNicol, J., Collins, T., Draper, J.S., and Bhatia, M. (2011). Cell fate potential of human pluripotent stem cells is encoded by histone modifications. Cell Stem Cell 9, 24–36.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet. 6, 6.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J. 32, 340–353.

Kawaji, H., Severin, J., Lizio, M., Waterhouse, A., Katayama, S., Irvine, K.M., Hume, D.A., Forrest, A.R., Suzuki, H., Carninci, P., et al. (2009). The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. Genome Biol. 10, R40.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 4, e1000242.

Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development 120, 3197–3204.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G.

(2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell 149, 590–604.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560.

Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.-H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. Dev. Cell 24, 324–333.

O'Neill, L.P., VerMilyea, M.D., and Turner, B.M. (2006). Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nat. Genet. 38, 835–841.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell 145, 835–850.

Qin, H., Blaschke, K., Wei, G., Ohi, Y., Blouin, L., Qi, Z., Yu, J., Yeh, R.-F., Hebrok, M., and Ramalho-Santos, M. (2012). Transcriptional analysis of pluripotency reveals the Hippo pathway as a barrier to reprogramming. Hum. Mol. Genet. 21, 2054–2067.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., Rossant, J., and Handling, E. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc. Natl. Acad. Sci. USA 107, 10783–10790.

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.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat. Rev. Genet. 9, 129–140.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. 7, e13.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell 48, 849–862.

Seisenberger, S., Peat, J.R., and Reik, W. (2013). Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. Curr. Opin. Cell Biol. 25, 1–8.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005). Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev. Biol. 278, 440–458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007). Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development 134, 2627–2638.

Shankaranarayanan, P., Mendoza-Parra, M.A., Walia, M., Wang, L., Li, N., Trindade, L.M., and Gronemeyer, H. (2011). Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. Nat. Methods 8, 565–567.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature 464, 922–926.

Yoshimizu, T., Sugiyama, N., De Felice, M., Yeom, Y.I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Schöler, H.R., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Dev. Growth Differ. 41, 675–684.