Germ plasm localisation dynamics mark distinct phases of transcriptional and post-transcriptional regulation control in primordial germ cells

Fabio M. D'Orazio[1,2], Piotr Balwierz[1,2], Yixuan Guo[3], Benjamín Hernández-Rodríguez[4], Aleksandra Jasiulewicz[1], Juan M. Vaquerizas[4], Bradley Cairns[3], Boris Lenhard[2] and Ferenc Müller[1]

[1]Institute of Cancer and Genomics Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK.
[2]MRC London Institute of Medical Sciences and Faculty of Medicine, Imperial College, London, UK
[3]Department of Oncological Sciences and Huntsman Cancer Institute, Howard Hughes Medical Institute, University of Utah School of Medicine, Salt Lake City, UT, 84112, USA
[4]Max Planck Institute for Molecular Biomedicine, Roentgenstrasse 20, 48149 Muenster, Germany.

Correspondence to: f.mueller@bham.ac.uk, b.lenhard@imperial.ac.uk

SUMMARY

In many animal models, primordial germ cell (PGC) development depends on maternally-deposited germ plasm to avoid somatic cell fate. Here, we show that PGCs respond to regulatory information from the germ plasm in two distinct phases and mechanisms in zebrafish. We show that PGCs commence zygotic genome activation together with the rest of the embryo with no demonstrable differences in transcriptional and chromatin accessibility levels. Thus, cytoplasmic germ plasm determinants only affect post-transcriptional stabilisation of RNAs to diverge transcriptome from somatic cells, which, unexpectedly, also activate germ cell-specific genes. Perinuclear relocalisation of germ plasm is coupled to dramatic divergence in chromatin opening and transcriptome from somatic cells characterised by PGC-specific chromatin topology. Furthermore, we reveal Tdrd7, regulator of germ plasm localisation, as crucial determinant of germ fate acquisition.

INTRODUCTION

In sexually-reproducing organisms, the germ line ensures that parental genetic information is passed on from one generation to the next. The commitment of the embryonic germ line follows unique steps (Strome and Updike, 2015): in metazoans, the germ fate can be either oocyte-inherited (predetermined) (Eddy, 1975; Williamson and Lehmann, 1996) or zygotically-triggered (induced) (Lawson et al., 1999; Ying et al., 2001). In mammals, germ cells are generated during gastrulation in response to extracellular signals from the surrounding embryonic cells (Ying et al., 2001). On the other hand, most non-mammalian model organisms, such as *C. elegans, D. melanogaster, X. laevis and D. rerio,* require maternal transmission of germ cell-specific factors, which aggregate in a self-containing structure known as germ plasm and distribute into primordial germ cells (Eddy, 1975; Seydoux and Braun, 2006). The germ plasm has been shown to be sufficient and necessary to trigger the germ fate (Gross-Thebing et al., 2017; Tada et al., 2012). The function of the germ plasm is to avoid somatic lineage differentiation of the host cells by at least two mechanisms. Firstly, germ plasm components post-transcriptionally regulate maternal RNA stability and translation (Charlesworth et al., 2006; Iguchi et al., 2006; Nakamura et al., 2004; Siddall et al., 2006; Wilhelm et al., 2003) while they are cleared in the rest of the embryo (Giraldez et al., 2006; Mishima et al., 2006). Also, germ plasm-specific factors, such as Nanos, Dazl and Dead-end (Dnd) function in RNA processing pathways and are indispensable for PGC development (Köprunner et al., 2001; Oulhen et al., 2017; Suzuki et al., 2010). Dnd, induces selective translation of mRNAs by liberating them from micro-RNA (miRNA) inhibition (Kedde et al., 2007), while the RNA-binding-Protein (RBP) DAZL was found to inhibit the translation of several mRNA involved in pluripotency, somatic differentiation and apoptosis in murine PGCs (Chen et al., 2014). Secondly, germ plasm components have been associated with block or delay of Zygotic Genome Activation (ZGA) of the hosting cell and proposed as key factors for avoiding somatic cell fate. In *C. elegans* and *D. melanogaster,* the germ plasm proteins PIE-1 and Pgc delay ZGA in the PGCs, allowing the disengagement between the germ and the somatic lines (Batchelder et al., 1999; Mello et al., 1996; Strome and Updike, 2015).

In contrast to the extensive, genome-wide DNA demethylation observed in migrating mammalian PGCs (Bender et al., 2004; Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015) reprogramming of DNA methylation has not been seen in zebrafish (Ortega-

Recalde et al., 2019; Skvortsova et al., 2019). On the other hand, epigenetic regulators maternally transmitted via germ granules have been implicated in germ fate acquisition (Gaydos et al., 2012; Rechtsteiner et al., 2010; Strome et al., 2014), suggesting that alternative mechanisms of germ plasm-mediated transcriptional regulation may exist.

In this study, we aimed to characterise the potential roles of the germ plasm during PGC formation. We hypothesized that the distinct localisation patterns of the germ plasm before and during PGC migration may represent distinguishable cytoplasmic and nuclear functions in PGC specification. We profiled transcriptome, open chromatin and DNA methylome of developing PGCs at high temporal resolution and discovered two distinct phases of PGC specification during zebrafish embryogenesis. Consequently, we suggest that the first function of the germ plasm is solely cytoplasmic and does not influence transcription or chromatin landscape of the pre-migrating PGCs. However, the second phase of germ cell formation requires chromatin reorganisation, resulting in extensive transcriptional changes that temporally coincide with the association between germ plasm and nuclear membrane. By systematic identification of open chromatin and subsequent prediction of cis-regulatory elements, we show that putative enhancers in proximity of developmental genes remain compacted during PGC specification, resulting in the lack of somatic gene expression. Moreover, predicted PGC-specific regulatory elements appear to be distinctly proximal to the Transcription Start Site (TSS), in sharp contrast to the somatic-specific elements, which spread more distally. The observed chromatin organisation of regulatory elements may underlie unique DNA topology contributing to repression of somatic lineage differentiation. Finally, by inhibiting the translation of Tudor Domain 7a (Tdrd7a), which leads to disruption of germ plasm perinuclear localisation, we demonstrate the requirement of this germ cell determinant in defining germ cell-specific open chromatin and transcriptional landscape.

RESULTS

Characterisation of primordial germ cell transcriptome before and during migration

In order to investigate the roles of germ granules, we aimed to characterise the early phases of germ line development via extensive profiling of epigenetic and transcriptional features. For this reason, we focused on the first day of zebrafish embryogenesis, when the PGCs show the most dynamic and active behaviour. The Tg(Buc-GFP) line of *D. rerio* with fluorescently-marked germ plasm (Riemer et al., 2015) was utilised to isolate PGCs and non-fluorescent somatic cells by FACS (Supplementary Figure 1A). Total transcriptome, open chromatin and DNA methylation were analysed at various stages of zebrafish PGC development (Figure 1A). We first assessed transcriptome features associated with developmental stages and cell type and identified major changes coinciding with key events of development. As shown in Figure 1B and C, hierarchical clustering and Principal Component Analyses (PCA) demonstrated minimal transcriptome variance at and immediately after zygotic genome activation (high and dome stages) with small observable difference between replicates of germ plasm-containing and somatic cells (Figure 1C, Supplementary Figure 1B and Supplementary Table 1). However, gradual divergence between the somatic and PGC transcriptomes was coincidental with migration of PGCs and perinuclear localisation of the germ plasm (at 10 somites stage), leading to a marked separation of steady-state transcript content between PGCs and somatic cells by prim-5 stage.

In order to determine the transcriptional contribution to PGC development over time, we classified differentially expressed genes into four main groups of temporal expression patterns using k-means clustering (Figure 1D and Supplementary Figure 1C). Within each cluster, genes with distinct biological functions could be identified, as highlighted by group-specific enrichment for gene ontology terms (Supplementary Figure 1D). Genes belonging to Cluster 1 were upregulated in the somatic cells at 10 somites and prim-5 (late) stages over every other sample. As expected, these were associated with developmental processes, differentiation and protein translation. Cluster 2 included those genes upregulated in late PGCs. Interestingly, we noted several pathways involved in chromatin reorganisation, in particular related to DNA packaging. At this developmental stage, germ cells are found in the future genital ridge and show a very low proliferative activity. Accordingly, genes downregulated in the PGCs at late stages versus

all other stages were mainly involved in cellular division (Cluster 4), while the Cluster 3 grouped genes for germ fate and thus validate the successful sorting of PGCs.

Taken together, these results demonstrate successful isolation and genome-wide comparative characterisation of PGCs at various developmental stages, which show stage- and cell type-specific transcriptomic differences. Overall, early germ cells have a similar transcriptional profile with the rest of the embryo, which gradually diverges as lineage specification proceeds. At the prim5 stage, analysis of PGC transcriptomes suggests reduced cell proliferation and epigenetic reorganisation compared to somatic cells (Cluster 2 and 4, Supplementary Figure 1D).

Zebrafish primordial germ cells do not delay zygotic genome activation

Global transcriptomic analysis suggested that PGCs and somatic cells are broadly similar during early zebrafish development. Therefore, although other germ plasm-dependent organisms delay transcriptional activation in the PGCs, the mild transcriptional difference between PGCs and somatic cells at ZGA suggested that transcriptional activation of the germ line may occur alongside the rest of the embryo. We therefore asked whether gene expression in PGCs is repressed during the time when ZGA commences in all the blastomeres. We utilised a recently developed in-vivo transcription imaging tool (MOVIE) and 4D imaged the accumulation and localisation of *microRNA-430* (*miR-430*) primary transcripts, which are the earliest known expressed genes highly active during ZGA in zebrafish (Giraldez et al., 2006; Hadzhiev et al., 2019). Upon injection of fluorescently-labelled morpholinos, we monitored *miR-430* expression in embryos in which germ plasm was tracked by GFP. Interestingly, *miR-430* expression was detectable in somatic cells as well as germ plasm-containing cells already before the main wave of genome activation (Figure 2A, Supplementary Figure 2A and 2B), suggesting that germ plasm does not substantially delay/inhibit early transcription. Also, *miR-430* expression faded around the onset of epiboly in both PGCs and somatic cells, suggesting similar temporal transcriptional regulation between the PGCs and somatic cells.

In order to better understand the relation between ZGA and germ plasm, we studied the composition of the PGC transcriptome by identifying differentially expressed genes across ZGA via RNA-seq (Methods). The number of significantly upregulated genes increased more than an order of magnitude when the germ plasm-carrying cells

transition from 256-cell to high stage (135 genes when padj < 0.1 and 124 when padj < 0.05) or from high to dome stages (1926 genes when padj < 0.1 and 1629 when padj < 0.05) (Figure 2B, 2E and Supplementary Table 2). Interestingly, of the 135 upregulated genes after ZGA, 60 were predicted to be zygotically transcribed without maternal contribution (Figure 2C). Zygotic genes were predicted as those with expression lower than 2 tpm at 256-cell stage (no zygotic transcription), which show an increase in expression levels at the subsequent analysed developmental stage. To validate our analysis, we compared our list of predicted zygotic genes with an independent RNA-seq dataset (White et al., 2017). The high degree of overlap between the predicted zygotic and maternal genes within the two datasets supported our arbitrary discrimination between maternally-deposited and zygotically-expressed transcripts (Supplementary Figure 2C). Notably, out of 60 zygotic genes found upregulated in the PGCs at high stage, 56 matched predicted zygotic genes from analysis of the independent RNA-seq experiment (data not shown). This observation suggests that transcription activation does occur at the time when the rest of the blastomeres are also becoming transcriptionally active. Interestingly, among the significantly upregulated genes in the PGCs at high stage, germ cell markers and germ plasm-localised transcripts such as *ddx4*, *dnd1*, *tdrd7a*, *gra* and *dazl* were found upregulated from the previous stage (Supplementary Figure 2D). This was unexpected as germ plasm markers are of maternal origin and are known not to be transcribed until after gastrulation (Blaser et al., 2005; Knaut et al., 2000; Weidinger et al., 1999).

Then, the occurrence of ZGA in the germ plasm-carrying cells was further confirmed after performing a regression-based clustering of genes with similar expression profiles for three stages spanning the first wave of ZGA (256-cell, high and dome) in PGCs and somatic cells. Interestingly, we found two clusters of genes upregulated in both PGCs and somatic cells and one in which genes were upregulated in PGCs exclusively (Figure 2D).

Finally, to further verify the observation of occurrence of active transcription between 256-cell and high stage in PGCs, we measured absolute RNA levels over time. After normalisation for an internal RNA control (Jiang et al., 2011) (Supplementary Figure 2F), we found a significant increase in transcript levels in both PGCs and somatic cells at high stage, confirming that active transcription is present in both cell types before the high stage (Figure 2F).

Based on our results, we concluded that the germ plasm does not cause general transcriptional repression in zebrafish, and zebrafish PGCs do not delay ZGA as it has been seen in *C. elegans*, *D. melanogaster* or *X. laevis*.

Selective retention of zygotically-produced transcripts explains transcriptome differences between germ plasm-carrying cells and somatic cells at ZGA

The observations that germ plasm-carrying cells do not delay transcriptional activation in comparison to the rest of the embryo -yet they appear to carry de-novo generated germ cell-specific transcripts- prompted us to ask whether differential transcription occurs between germ plasm-carrying cells and somatic cells at ZGA. Therefore, we sought to study the nature of ZGA in the pre-migratory PGCs in comparison to the somatic cells. To verify whether the PGCs undergo selective transcriptional activation, we performed differential gene expression analysis of isolated germ plasm-carrying cells and somatic cells at each stage spanning ZGA period. This analysis revealed that before ZGA, already a total of 22 genes were differentially expressed between the two cell types (padj < 0.1) (Figure 3A), confirming that maternal mRNAs are selectively retained in the PGCs as shown previously (Eno et al., 2018; Gazdag et al., 2009; Gerovska and Araúzo-Bravo, 2016; Gorokhova et al., 2007; Levine et al., 2000; Rothschild et al., 2013). However, it was noteworthy that 11 out of 22 identified transcripts have not yet been associated with germ cell or PGC functions and are candidates for novel maternal germ plasm-transcripts. The remaining 11 transcripts instead were either known germ plasm markers (*gra, tdrd7, rg514a, ca15b, dnd1* and *dazl*) or being previously associated with germ cell development/survival (*hook2, tgfa, zswim5, b4galt6, camk2g1*) (Supplementary Table 3).

During genome activation (high stage), we identified 142 genes significantly differently regulated between PGCs and somatic cells (Figure 3A). Interestingly, when the number of upregulated genes from one stage to the next was taken into account, we noted a similar trend of gene upregulation in both PGCs and somatic cells, indicating that both the cell types experienced equal transcriptional activation signals (Supplementary Figure 3A). By confirming that ZGA in PGCs and somatic cells is similar, we observed that the relative abundance of transcripts upregulated from 256-cell to high stage in only one cell type was still comparable between germ cells and somatic cells at high stage (Supplementary Figure 3B). Also, the fold change of increased gene expression between

256-cell and high stage showed high correlation between germ plasm-carrying cells and somatic cells (Supplementary Figure 3C). Taken together, these observations suggested that both the cell types activate their genomes similarly.

To address whether retention of the germ plasm in PGCs impacts directly on PGC-specific transcriptional activation and to trace any transcriptional source of the small variation in steady state RNA levels during genome activation, we monitored transcriptional activity using chromatin accessibility state at cis-regulatory elements as a surrogate. We performed Assay for Transposase-Accessible Chromatin combined with sequencing (ATAC-seq) in PGCs and somatic cells. Open chromatin landscapes were analysed globally and compared between PGCs and somatic cells at high and dome stages (start of main ZGA wave). Interestingly, global comparison of open chromatin regions between PGCs and somatic cells at high stage indicated a high degree of correlation (Figure 3B). No distinguishable difference in chromatin accessibility for PGCs and somatic cells was observed on either promoters or distal elements of genes with differential expression between PGCs and somatic cells (Figure 3C). The germ line gene *ddx4*, whose expression was shown by the RNA-seq analysis to increase from 256-cell to high stage only in PGCs, appeared to possess similar degree of chromatin accessibility at its promoter and cis regulatory regions in both cell types (Figure 3D). Taken together, these observations prompted us to hypothesise that, while zygotic transcriptional activation is broadly similar between somatic and germ plasm-carrying cells, the detected differential gene expression was caused by post-transcriptional regulation.

While it was previously reported that maternal RNAs are selectively protected in the PGCs from miRNA-dependent degradation (Mishima et al., 2006), this mechanism has not yet been shown to occur on zygotically-active germ cell genes. As the top-scoring upregulated genes in the PGCs from 256-cell to high stage are known-germ plasm markers, we hypothesised that these are not only deposited by the mother in the germ plasm but are also produced zygotically throughout the early embryo including the somatic cells where they are selectively cleared. To test this hypothesis and to discriminate between maternally-provided and zygotic germ plasm transcripts, we took advantage of our RNA-seq data to perform differential Intron Retention (IR) analysis (Middleton et al., 2017). As intron splicing is co-transcriptional (Merkhofer et al., 2014), newly-transcribed RNAs are expected to show increased intron retention. We compared IR scores for the whole transcriptome before and after ZGA, and observed increase of

intron retention in both PGCs and somatic cells upon transcriptional activation (Supplementary Figure 3D). Then we focused on assessing IR in de-novo transcription of germ cell-specific transcripts in somatic cells. A significant increase in intron retention from the previous stage was observed when compared to random sampling of the dataset (Figure 3E). This result further supports the hypothesis that genes transcribed in the PGCs at ZGA were also activated in the somatic cells.

To further validate this observation, we performed quantitative PCR (qPCR) on nuclear and cytoplasmic cell fractions before and after ZGA after removal of PGCs via FACS (Supplementary Table 4). Interestingly, when transcription has started, we saw a significant increase in fold change expression for these genes in the nuclear fraction but not in the cytoplasmic fraction, suggesting that germ cell-related transcripts could be produced by the somatic cells (Supplementary Figure 3E). In order to demonstrate that somatic cells transcribe germ cell-specific RNAs, we studied the localisation of newly-transcribed pre-mRNA within the embryo. Within those transcripts that were selectively upregulated in the PGCs from 256-cell to high stage, *dazl* was one of the highest-scoring hits (Supplementary Figure 3F). We therefore designed RNA probes targeting intronic sequences to visualise unprocessed, newly-produced *dazl* pre-mRNA and carried out fluorescence in-situ hybridization at different stages with and without transcription block. Strikingly, *dazl* was seen to be actively transcribed in the somatic cells, as demonstrated by staining of nuclear foci in the whole embryo at high stage (Figure 3F).

Taken together, these results suggest that transcription of germ cell genes occurs throughout the embryo at ZGA. Upon integration with ATAC- and RNA-seq results, we propose that there is no transcriptional delay or differential transcription in germ plasm-carrying cells when zygotic genome is activated. Selective protection of zygotic transcripts together with maternal transcript by the germ plasm may thus contribute to early germ cell specification and onset of migration.

Primordial germ cells gain specific transcriptomic and epigenetic features during migration

We next asked how transcriptome and chromatin states reflect the distinct ontogeny of PGCs and somatic cells during migration and further development. About 4 hours after fertilisation (dome stage), PGCs initiate independent and guided movements, which will bring them into the future genital ridge site (Raz, 2003). The germ plasm undergoes

extensive morphological changes during germ cell migration (Figure 1A). Also, this period coincides with gastrulation and germ layer formation, with remarkable transformation of the whole embryo transcriptome upon lineage diversification (Raj et al., 2018). Our differential gene expression analysis between PGCs and somatic cells indicates increasing divergence of transcriptomes coinciding with germ fate acquisition in the migrating PGCs (as shown in Figure 1D). The number of differentially expressed genes between PGCs and somatic cells gradually increased over time, reaching almost one-third of the total transcriptome at prim-5 stage (Figure 4A). To trace the cell type as the source of transcriptome variation, we looked at the number of upregulated genes from one stage to the in PGCs and somatic cells selectively. We noted that the increase in differential gene expression between PGCs and somatic cells was accompanied by an increase in differential gene expression over time (Figure 4B, Supplementary Figure 4A and Supplementary Table 5), suggesting that both PGCs and somatic cells were undergoing cellular commitment. In-depth analysis of the migratory PGC transcriptome showed overexpression of several chromatin remodelers. The DNA methyltransferase 3 (*dnmt3bb.1*), responsible for de-novo methylation of CpG islands, was upregulated in PGCs when compared to somatic cells, while the Ten Eleven Translocase 2 (*tet2*) was downregulated. Concomitantly, upregulation of the arginine methyltransferase *prmt6*, lysine demethylases (*kdm7* and *kdm8*) and bromodomain-containing proteins (*brdt* and *brd9*) in the PGCs suggested for potential mechanisms of chromatin remodelling via histone modifications (data not shown).

The differential expression of chromatin regulation prompted us to ask whether germ fate acquisition involves observable epigenetic and chromatin changes as was described in murine PGCs. Therefore, we compared global chromatin accessibility in PGCs and somatic cells via ATAC-seq. After selection of reproducible peaks via irreproducible discovery rate (IDR) filtering (Zhang et al., 2017) (IDR < 0.05) (Supplementary Figure 4B and Supplementary Table 6), we compared global variability of chromatin accessibility among developmental stages and cell types (Figure 4C and Supplementary Figure 4C). In accordance with transcriptome results, early PGCs and somatic cells have similar open chromatin profiles and cluster by stage rather than by cell type. In contrast, marked separation of PGC and somatic cell chromatin accessibility profiles was observed after gastrulation and resulted in the identification of lineage specific sites of open chromatin (Supplementary Figure 4D).

Following differential chromatin accessibility analysis, we found 12591 peaks upregulated in the PGCs (logFC < -1, padj < 0.05) and 23771 peaks upregulated in the somatic cells (logFC > 1, padj < 0.05) at prim-5 stage (Figure 4D). We then used self-organising map (SOM) analysis on PGC and somatic ATAC-seq data at different stages to identify patterns of cis-regulatory element accessibility via an unsupervised approach (Figure 4E). We identified 9 clusters of open chromatin dynamics. Of these, SOM identified 9869 sites that were specific for late stage PGCs and 12654 sites that were less accessible in the PGCs compared to the late stage somatic cells. Interestingly, genes in proximity of somatic-specific ATAC peaks were associated with Gene Ontology (GO) terms for embryonic morphogenesis, tissue formation and embryonic development. These results highlighted that germ fate acquisition requires chromatin regulation.

To gain more insight into the epigenetic specification of PGCs under the control of the germ plasm, we profiled the DNA methylome of pre- and migratory PGCs by performing Reduced Representation Bisulfite Sequencing (RRBS) (Murphy et al., 2018). In contrast to mammals (Guo et al., 2015; Hill et al., 2018), we found no extensive DNA methylation reprogramming of PGCs during these stages (Figure 4G and Supplementary Figure 4F). Analysis of differentially methylated CpGs between PGCs and somatic cells identified 3825 significantly differentially methylated regions (Supplementary Table 7) (only 1.77% of all recovered CpGs from all samples), revealing an overall highly similar methylation programme, in accord with recent studies (Ortega-Recalde et al., 2019; Skvortsova et al., 2019). Based on this result, we concluded that DNA methylation dynamics and chromatin remodelling in zebrafish PGCs were uncoupled.

The finding that post-migratory PGCs show a specific chromatin accessibility landscape in contrast to the early germ plasm-carrying cells suggests that the onset of germ fate acquisition occurs during PGC migration and coincides with the subcellular re-localisation of the germ granules. In addition, differences between PGCs and somatic cells were more marked when chromatin accessibility was profiled in comparison with DNA methylation. This prompted us to investigate more closely the contribution of chromatin regulation to gene expression on different genetic elements.

Primordial germ cell-specific open chromatin profile is enriched for promoter-proximal putative enhancers and depleted for promoter-distal, developmental putative enhancers

After having established that migrating PGCs undergo chromatin rearrangements, we asked how their epigenetic features contribute to transcription and germ fate acquisition. First, we discriminated between core promoters and distal cis-regulatory elements based on ATAC-seq results in migrating PGCs and somatic cells (prim-5 stage). Gene promoters were defined as sequences within 500bp from annotated Transcription Start Sites (TSSs), while the excluded ATAC peaks (accessible chromatin) within 50kb from the closest TSS were defined as distal elements. Comparison between normalised ATAC signals in PGCs and somatic cells showed that the chromatin profiles were more dissimilar on non-promoter associated open chromatin regions (Figure 5A). In order to further dissect the chromatin accessibility state across the genome, we performed differential chromatin accessibility analysis (logFC > 1, padj < 0.05, fold enrichment > 4) and we focussed on the distribution of differentially regulated ATAC peaks over genic elements. While most of the somatic ATAC peaks occurred at promoter-distal regions, PGC-specific ATAC peaks tended to coincide with promoter (Supplementary Figure 5A). In the PGCs, 43% of the upregulated open regions were found within 1kb from the TSS, while 11% was associated with introns. On the other hand, 41% of the upregulated regions in the somatic cells were found within introns, and only 6% was associated with promoters. We aimed to define putative somatic enhancer regions by intersecting the upregulated ATAC peaks with active enhancer histone mark H3K27ac (Bogdanovic et al., 2012). Out of all identified somatic-specific ATAC peaks (logFC > 1 and padj < 0.05), almost 30% were associated with H3K27ac histone marks. In contrast, less than 10% of PGC-specific ATAC peaks (logFC < -1) matched location of somatic H3K27ac peaks (Supplementary Figure 5B). The functional relevance of cell type-specific chromatin accessibility was estimated by GO analysis of genes associated with differential open chromatin regions away from promoters. As expected, somatic cells were enriched for open chromatin regions in proximity of genes for developmental and differentiation pathways, while PGC-specific ATAC peaks were found in proximity of genes for germ fate, cellular transport and stem cell differentiation (Figure 5B and Supplementary Figure 5C). As GO terms for both gene expression and chromatin accessibility pointed at similar pathways involved in PGC specification, we sought to verify whether accessible DNA regions would be predictive of

transcriptional activity. To verify this, we compared the fold change of gene expression between PGCs and somatic cells for a subset of transcripts characterised by significant differential open chromatin between the two cell types. Interestingly, we noted a significant correlation between cell type-specific, promoter-associated ATAC peaks and transcriptional output in both PGCs and somatic cells (Figure 5C, left). On the other hand, the accessibility of distal elements was not predictable of transcription in the PGCs, while a higher correlation between transcription and chromatin accessibility was observed in the somatic cells (Figure 5C, right). This result suggests that transcriptional regulation in migratory PGCs is less dependent on distal elements. In support, we observed an inverse trend of correlation between open chromatin and gene expression in relation to distance from the TSS for PGCs and somatic cells (Supplementary Figure 5D). Moreover, the cumulative distribution of PGC/somatic-specific cis-regulatory elements in relation to the closest transcription start sites showed that PGC-specific cis-regulatory elements were found more proximal to TSSs in comparison to somatic-specific ones (Figure 5D). Taken together, these results indicate that regions of open chromatin predicted to drive gene expression from a distance are less frequent in PCGs.

Next we asked what epigenetic mechanisms may be involved in keeping putative distal enhancers closed in PGCs. Two mechanisms have been implicated in enhancer activation: DNA demethylation and histone priming (Bogdanović et al., 2016; Zhang et al., 2018). Although little variation was seen in global DNA methylation profiles of PGCs and somatic cells at prim-5 stage (Figure 4E), we explored methylation dynamics in the PGCs further by discriminating promoter and enhancer regions. When comparing methylation levels across stages and cell types, we observed that promoters preserve their methylation status. Out of 13 validated genes between PGCs and somatic cells, only *ddx4* showed differential DNA methylation on its promoter (Supplementary Figure 5E). In contrast, more dynamic methylation pattern was seen on putative enhancers. In particular, demethylation observed in late somatic cells was not observed in the PGCs, which instead resembled the early profiles characterised by higher levels of DNA methylation (Figure 5E and Supplementary Table 7). Interestingly, putative enhancers showed concomitant higher methylation and significant lower chromatin accessibility in the PGCs compared to the somatic cells (Figure 5F). Next, we tested whether the observed methylation changes were coupled with changes in modulators of DNA methylation. We found upregulation in the PGCs of four out of six *dnmt3bb.1* and *dnmt3bb.2* genes, while no

significant change was detected for *dnmt1* gene. On the other hand, *tet2* expression was higher in the somatic cells compared to the PGCs, suggesting that PGCs may lack mechanisms of hydroxymethylation-mediated DNA demethylation catalysis (Breiling and Lyko, 2015; Hill et al., 2018) (Figure 5F).

Based on these results, we propose that PGCs avoid somatic cellular commitment and somatic fate by inhibiting transcription of developmental genes upon block of chromatin opening of their enhancers. Moreover, PGCs appear to transcribe genes which, unlike developmental genes, are enriched in promoter-proximal regulatory elements and thus appear to be regulated by contrasting chromatin topology from that seen in somatic cells.

TDRD7a, a germ plasm-segregating protein, is required for maintaining germ line-specific chromatin and transcriptome signature

Our genome-wide analysis of chromatin accessibility demonstrated a germ line-specific chromatin reprogramming, which coincides with the transition of the subcellular localisation of the germ granules from cytoplasmic to perinuclear during the early phases of migration (5-6hpf) (Doitsidou et al., 2002; Strasser et al., 2008; Updike et al., 2011; Weidinger et al., 2003). After this stage, the germ plasm associates with the nuclear membrane, although the role of this interaction remains unknown. We hypothesised that the observed transcriptome divergence between PGCs and somatic cells could be driven by perinuclear germ plasm-mediated gene regulation mechanisms detectable on the level of chromatin changes. A known germ plasm marker is the protein Tudor domain 7 (Tdrd7), which is also one of the highest differentially expressed genes in the reported RNA-seq dataset at all developmental stages (data not shown). When translation of Tdrd7 is inhibited by morpholino interference, the germ plasm is incapable of fragmenting and forming small localised granules (Strasser et al., 2008). In order to test whether Tdrd7 and perinuclear germ granules are required for genome-wide chromatin rearrangement observed in PGCs, we investigated the chromatin accessibility and transcriptome profiles of embryos, in which germ plasm localisation was perturbed upon morpholino-mediated Tdrd7 knock-down (KD) (Supplementary Figure 6A and Supplementary Table 8).

Differential gene expression analysis of Tdrd7 KD and wild type PGCs at prim-5 stage indicated a dramatic effect by the loss of Tdrd7 (Figure 6A and Supplementary Table 9). Gene ontology analysis indicated that downregulated genes were enriched for genes involved in reproduction and germ cell development, while upregulated genes were

enriched in genes associated with developmental process, organogenesis and protein expression processes (Supplementary Figure 6B). These observations suggested that, in absence of Tdrd7 and correct germ plasm distribution, germ cell character was lost in favour of somatic lineage fate acquisition. Indeed, PCA analysis of variance among Tdrd7 wild type and KD embryo transcriptomes demonstrated that lack of Tdrd7 in PGCs shifted their transcriptome profile towards somatic-like transcriptome (Figure 6B). Detailed RNA-seq analysis highlighted that genes downregulated in PGCs upon Tdrd7 KD also include genes associated with pluripotency and zygotic genes for gametogenesis. While ubiquitously expressed housekeeping genes were unaffected (Figure 6C and Figure 6E). It was noteworthy, that maternally-inherited germ plasm transcripts showed only slightly reduced levels in the Tdrd7-lacking PGCs as compared to control PGCs (Supplementary Figure 6C and Supplementary Figure 6D). This could be caused inhibition of late PGC-specific transcription upon Tdrd7 KD, while maternal mRNAs were preserved. These observations indicate that disruption of Tdrd7 function with germ plasm mis-localisation leads to global deregulation of PGC transcription programme and mild notable effect on germ plasm RNAs.

Next, we asked whether the causes of differential gene expression in Tdrd7 KD PGCs could be traceable to the chromatin state. To this end, ATAC-seq was carried out on Tdrd7 KD PGCs at prim-5 stage and compared to control PGCs. Global analysis of open chromatin in Tdrd7 KD embryos revealed a marked effect by Tdrd7 loss in the PGCs. Open chromatin of Tdrd7 KD PGCs resembled more to that in the wild type somatic cells than to wild type and PGCs injected with a morpholino control, suggesting that the observed transcriptome phenotype was accompanied by chromatin changes (Figure 6D).

Analysis of putative cis-regulatory element in proximity of misregulated genes in Tdrd7 morphant embryos showed a general tendency towards compaction of PGC-specific genes regions, while development- and morphogenesis-associated genes gained open chromatin peaks in comparison to control PGC cells (Figure 6E and Supplementary Figure 6D). To get a more comprehensive overview of mechanisms governing chromatin regulation in the PGCs and to confirm the association between chromatin accessibility profile and apparent reprogramming of PGCs towards the somatic fates, we have carried out a global analysis of open chromatin behaviour upon Tdrd7 knock down and generated SOM classes of putative regulatory elements (Supplementary Figure 6E). Unsupervised sample clustering confirmed epigenetic differences between wild type and

Tdrd7 KD PGCs, indicating distinct shift of PGC chromatin states towards somatic fate, when translation of *tdrd7* was inhibited and the germ granules were mis-localised. Of note, PGC-specific genes, such as *dazl*, which are expressed zygotically as well as inherited maternally in the germ plasm, were shown to be associated with closure of promoter and candidate enhancers despite only mild reduction in their RNA levels (Supplementary Figure 6D).

These results demonstrate transcription regulatory roles for the germ plasm determinant Tdrd7 and suggest that perinuclear germ plasm-nucleus cross-talk is required for germ cell-specific gene activation and fate decision.

DISCUSSION

In this study, we have investigated the transcriptome, chromatin accessibility and DNA methylation dynamics during early development of PGCs in a germ plasm-dependent vertebrate. The use of the Tg(Buc-GFP) line, combined with ATAC- and RNA-seq, has allowed us to investigate the nuclear and cytoplasmic roles of the germ plasm at unprecedented resolution. Moreover, by following chromatin and transcriptional features in the germ line over time, we were able to describe two distinct roles for germ cell-specific cytoplasmic granules. We link these two roles with two distinct germ plasm subcellular distributions and distinguished an early and a late phase of germ fate acquisition in zebrafish (Figure 7). In contrast to many animal models, we demonstrated that germ plasm contributes to post-transcriptional regulation of both maternal and zygotic gene products, with no detectable effect on transcription during and after genome activation. Following commencement of PGC migration and re-localisation of the germ plasm, PGC-specific chromatin accessibility profiles evolve together with PGC-specific transcription, which we show are Tdrd7-dependent and likely mediated by re-localisation of the germ granules.

ZGA is not delayed in zebrafish germ plasm-carrying cells

Germ plasm factors block or delay zygotic genome activation by sequestering RNA polymerase II (RNA Poll II) or its co-factors at the time of transcriptional activation in many organisms (Batchelder et al., 1999; Mello et al., 1996). By delaying the ZGA, the PGCs avoid the commencement of developmental differentiation programmes characteristic to the developing somatic lineages. In mouse, in which PGCs develop by induction and without maternal germ plasm (Lawson et al., 1999; Ohinata et al., 2009; Saitou and Yamaji, 2012), newly-formed germ cells undergo transcriptional quiescence for a short period of time (Kurimoto et al., 2008). The mechanisms of transcription pausing are not fully understood, however they involve inhibition of the elongation factor P-TEFb by Pgc and PIE-1 proteins in *D. melanogaster and C. elegans* (Hanyu-Nakamura et al., 2008; Mello et al., 1996) and sequestration of the general transcription factor TAF4 in *C. elegans* (Guven-Ozkan et al., 2008). In sharp contrast, immunostaining of Serine 2-phosphorylated RNA Poll II, showed localised nuclear foci in zebrafish germ plasm-carrying cells as early as 256-cell stage (Knaut et al., 2000), suggesting that delay of genome activation in the germ line may not occur in this species and that alternative

mechanisms of germ fate acquisition may exist. In this study, we imaged transcription in-vivo for the first time in PGCs and together with transcriptome analysis, provide independent lines of evidence that the germ plasm does not delay the first wave of Zygotic Genome Activation (ZGA) in zebrafish. These findings indicate striking plasticity in establishing the germ cell fate among clades and raise the prospect that transcriptional delay is not essential for PGC formation. Interestingly, since massive transcriptional activation has been previously shown to cause double strand break (DSB) in the embryonic germ line (Butuči et al., 2015), by not delaying ZGA zebrafish PGCs may experience a milder transition from a non-transcribing to a transcribing state while being less exposed to DNA damage.

Why do PGCs need zygotic contribution to maternally deposited PGC mRNAs?

We provide several lines of evidence to show that PGCs and somatic cells transcribe similar set of genes during blastula stages and that there is no overt difference in their chromatin states. We analysed chromatin accessibility, de-novo transcript abundance, intron retention and intron-containing nascent RNA localisation, which all indicate no difference in transcription between the PGCs and the somatic cells. Strikingly, we report that PGC-specific RNAs are also zygotically transcribed in the somatic precursors, where their maternal counterparts are known to be promptly degraded by zygotic machineries (Giraldez et al., 2006; Mishima et al., 2006). Thus, it is unclear why somatic cells produce PGC-specific RNAs. This unexpected observation may be explained by chromatin accessibility organisation of undifferentiated pluripotent stem cells, such as the blastomeres of the zebrafish embryo at and immediately after ZGA. Pluripotent cells display less compact chromatin organisation than differentiating cells (Andrey and Mundlos, 2017) and are characterised by accessible chromatin with low, but detectable activity of a broad range of genes. In contrast, differentiating cells are characterised by gradual divergence of accessibility of lineage and cell type specific-enhancers (Ladstätter and Tachibana, 2019; Lu et al., 2016; Perino and Veenstra, 2016). Therefore, it is feasible that during ZGA, zebrafish blastomeres carry a pluripotent ES-like chromatin and corresponding transcription state characterised by broad capacity of gene expression. This primordial state of chromatin organisation, likely lacks or only commences the formation of chromatin architecture characteristic of differentiating cells during development (Kaaij et al., 2018). It is conceivable that, in this primordial chromatin

556 architecture, sophisticated, enhancer-dependent regulation has not yet been established
557 and gene expression occurs without lineage-specificity and as such it tolerates, for
558 example, germ-cell specific gene expression in the somatic progenitors. In this model,
559 emphasis is on post-transcriptional control of gene expression and germ plasm-
560 mediated, selective stabilisation of zygotic transcripts could be the primary source of
561 divergence of transcriptome between PGCs and somatic cells (Figure 7). In support of
562 this, several germ plasm factors are known to contribute to post-transcriptional and
563 translational regulation. For example, the RNA binding germ plasm component Dnd1
564 stabilises maternal mRNAs selectively by restricting access of miRNAs responsible for
565 clearing maternal mRNAs in the somatic cells (Giraldez et al., 2006; Kedde et al., 2007).
566 Likewise, the germ factor Nanos has been implicated in destabilisation of RNAs and
567 translational inhibition in the germ cells in combination with Pumilio and CCR4-NOT (Lee
568 et al., 2017; Suzuki et al., 2012).
569 An additional question emerging from our finding is why zebrafish PGCs need zygotic
570 contribution to PGC-specific mRNA activity at ZGA, while most of these PGC-specific RNAs
571 are already present either in the germ plasm or in the cytoplasm of germ plasm-
572 containing PGCs. While we cannot answer this question directly, it is likely that such early
573 activation of zygotic transcription is either redundant, similarly to the thousands of
574 mRNAs which are both present maternally and zygotically in the somatic progenitors
575 (Haberle et al., 2014; Harvey et al., 2013) or that the zygotic component is required to
576 gradually take over and compensate for loss and or dilution of maternal mRNA in dividing
577 PGCs.
578

579 **A germ plasm-mediated epigenetic reprogramming engages germ fate**

580 Germ plasm-carrying cells initiate independent and active migratory movements by
581 dome stage (Blaser et al., 2005; Bontems et al., 2009; Eno and Pelegri, 2016; Eno et al.,
582 2018; Raz, 2003; Yoon et al., 1997), suggesting that migration may, in part, be
583 transcriptionally regulated. However, given the minimal transcriptional differences
584 observed between PGCs and somatic cells, the main mode of regulation is expected to be
585 post-transcriptional. As shown before, early migratory movements are triggered by Dnd-
586 mediated loss of cell adhesion (Blaser et al., 2005) as well as the interaction between the
587 C-X-C chemokine receptor type 4b (Cxcr4b) on the PGC surface and the ligand Cxcl12a
588 (Doitsidou et al., 2002). Additionally, translational inhibition of *nanos* and *oskar* mediated

589 by the germ plasm factors Bruno and CUP is required during germ cell formation in *D.*
590 *melanogaster* (Nakamura et al., 2004; Wilhelm et al., 2003), while the RNA-binding
591 protein DAZL inhibits translation of several mRNAs involved in pluripotency, somatic
592 differentiation and apoptosis in mouse PGCs (Chen et al., 2014). It is thus not surprising
593 that the germ plasm-specific factor *dazl* is degraded after ZGA in the somatic precursors
594 to restrict its function onto the PGCs in zebrafish. While we have not explored the mRNA
595 clearance process involved in removing zygotically-transcribed mRNAs from the somatic
596 cells, this is likely achieved via the global regulator of maternal mRNA degradation, *miR-*
597 *430* (Giraldez et al., 2006). Based on our results, we speculate that the germ plasm is
598 exclusively required to protect and process both maternal and zygotic transcripts during
599 onset of migration in the germ plasm-carrying cells. This feature provides the germ
600 plasm-carrying cells with unique capacities, likely required for triggering migration and
601 providing foundation for downstream molecular cues, which will initiate the second
602 phase of PGC development.
603 We have demonstrated, that unique chromatin accessibility programming occurs in the
604 zebrafish PGCs and that this cis-regulatory element-associated topology contributes to
605 avoiding somatic differentiation. Also, we were able to associate germ plasm re-
606 localisation as an extranuclear effector, which acts on chromatin remodelling coinciding
607 with perinuclear localisation (Figure 7). In line with this observation, interference with
608 germ plasm localisation via its regulator Tdrd7 impacts on the PGC fate, which correctly
609 migrate, but experience somatic differentiation during and following migration.
610 Notably, our results with DNA methylation analysis demonstrate that even in the absence
611 of global DNA demethylation (Macleod et al., 1999; Bogdanović et al., 2016; Skvortsova
612 et al., 2019), local, differential methylation between PGCs and somatic cells is observed
613 at regulatory sites. However, in contrast to mammalian PGCs, where DNA methylation is
614 almost completely erased genome-wide (Guo et al., 2015; Hill et al., 2018), we observe an
615 inverse trend in zebrafish, where hypermethylation is found on somatic putative
616 enhancers, suggesting remarkably different mechanism for epigenetic programming of
617 the germ line among vertebrates. This was further corroborated with open chromatin
618 analysis genome-wide, which showed lack of opening somatic enhancers in PGCs.
619 Interestingly, although our study led to the identification of PGC-specific putative
620 regulatory elements, we found low correlation between opening of distal ATAC peaks and
621 transcription. Higher correlation between chromatin accessibility and transcription was

622 detected for TSS-proximal cis-acting elements in PGCs than in somatic cells. This finding,
623 in line with previous reports in humans (Guo et al., 2017), suggests that transcription
624 activation in PGCs is achieved by short-range interactions. Nevertheless, additional
625 studies focusing on the DNA interactions, Topological Associated Domains (TADs)
626 formation and spatial organisation of the chromatin in the developing germ line will
627 contribute to better understand the chromatin architecture-associated mechanisms of
628 acquisition and maintenance of totipotency.
629

630 **Importance of germ plasm components and subcellular organisation**
631 In accordance with previous studies, our results confirm that removal of individual germ
632 plasm components is sufficient to trigger somatic differentiation in the PGCs (Gross-
633 Thebing et al., 2017; Lee et al., 2017). Tdrd7-depleted PGCs preserve germ granules, germ
634 factors and correctly reach the genital ridge (Hosokawa et al., 2007; Strasser et al., 2008),
635 suggesting that pathways upstream migratory movements are unaffected. However,
636 despite this mild phenotype, we report a remarkable somatic-like chromatin accessibility
637 and consequent transcriptional reprogramming. Based on this result, we speculate that
638 lack of Tdrd7 and incorrect germ granule localisation are sufficient to diverge the fate of
639 the embryonic germ line and induce activation of somatic differentiation pathways.
640 Tdrd7 is known to interact with Piwi, a piRNAs processor (Huang et al., 2011). Piwi-
641 mediated piRNAs processing has been associated with epigenetic changes in both the
642 somatic and the germ line (Houwing et al., 2007), therefore it is tempting to speculate
643 about a link between piRNA pathways and germ fate maintenance. For example, piRNAs
644 are known to control transposon silencing via H3K9me3 and, in general, to regulate
645 chromatin state on piRNA-target regions (Huang et al., 2013; Sienski et al., 2012).
646 Interestingly, it has been recently reported that germ granules protect germ line
647 transcripts from piRNA-mediated silencing, regulating the pace of release from the
648 cytoplasm to the nucleus (Ouyang et al., 2019). To date, we cannot conclude on whether
649 the germ granules re-localisation is epistatic to the epigenetic germ fate acquisition or
650 vice-versa. However, the multiple lines of evidence associating Piwi and piRNAs to
651 epigenetic regulation strongly suggest a possible requirement of the germ plasm-nuclear
652 interaction in order to trigger initiation of epigenetic germ fate. In conclusion, we suggest
653 that perinuclear localisation of the germ plasm and Tdrd7 are involved in chromatin
654 reprogramming of gonadal PGCs during somitogenesis. Our discoveries have

655 fundamental implications in the understanding of pluripotent fate acquisition and the
656 functional relationship between subcellular aggregates with epigenetic and chromatin
657 reprogramming.
658

ACKNOWLEDGEMENT

This work was supported by the H2020 programme Zencode-ITN, Wellcome Trust Investigator award (106955/Z/15/Z) and BBSRC (BB/L010488/1) to F.M. and B.L. and HFSP programme grant to F.M. and B.C. and NEURAM FET H2020 project of the European Commission to FM. We thank Brian Dalley for sequencing expertise, K.T. Varley for helping us adapt Patch Bisulfite PCR, the HCI high-throughput genomics core, flow cytometry core, HCI fish facility, and the Centralized Zebrafish Animal Resource (CZAR) facility at the University of Utah. Financial support was received from the Howard Hughes Medical Institute and the Huntsman Cancer Institute core facilities (CA042014). We thank Erez Raz and Katsiaryna Tarbashevicz at the Center for Molecular Biology of Inflammation (ZMBE) in Muenster for advice on the isolation of PGCs and comments on the manuscript. We thank Roland Dosch at the University Medical Center in Göttingen for providing the Tg(Buc-GFP) line and the BMSU facility at the University of Birmingham for zebrafish maintenance. We thank the Tech Hub core at the University of Birmingham for the flow cytometry and NGS equipment and technical support provided.

AUTHORS CONTRIBUTIONS

F.M.D. and F.M. conceived the study and wrote the manuscript. All the authors critically revised the manuscript. All experiments were performed by F.M.D. except for the following: Y.G. generated RRBS libraries from isolated genomic DNA. A.J. produced light sheet images of morpholino-injected Buc-GFP embryos. ATAC-seq data were analysed by F.M.D. and P.B. RNA-seq analysis was performed by F.M.D. and B.H.R. RRB-seq analysis was designed by B.C. and Y.G and performed by Y.G. and B.C.

METHODS

Animal procedures

All animal work was performed under the Project Licence # b6b8b391, in accordance with the UK Home Office regulations and UK Animals (Scientific Procedures) Act 1986. Fish pairs were crossed in 1 litre breeding tanks and kept overnight separated. The next morning, the gate was removed and the eggs collected at intervals of 5 minutes to ensure synchrony. Fertilised eggs were dechorionated by 10mg/mL Pronase and serial washes in sterile E3 medium. After dechorionation, embryos were kept in agarose-coated petri dishes at 28.5 °C in a 14/10 hours of light/dark respectively. Wild type (AB), Tg(Buc-GFP) and Tg(kop:eGFP) lines were used during this work.

Microinjections of zebrafish embryos

Morpholinos were diluted in phenol red and about 0.3 pM were injected into the yolk of fertilised zebrafish embryos with a glass needle as described in Hadzhiev et al., 2019.

Transcription block

Transcription block was achieved through embryo incubation in 1 µM triptolide (Sigma T3652) in E3 medium from the single-cell stage to completion of the experiment.

PGC preparation for FACS

PGCs were isolated at different stages via FACS. Tg(Buc-GFP) heterozygous embryos were grown at the desired stage by incubation in E3 medium supplemented with 1mg/ml gentamicin at 28.5°C. Embryos were washed three times in sterile water before collection and about 200 of them were pulled in single microcentrifuge tubes. 500 µl of HBSS supplemented with 0.25% BSA and 10mM Hepes were added and dissociation occurred by pipetting for 2 minutes with a glass pipette. Excess of yolk was removed by two rounds of 3 minutes centrifugation at 350 x g at 4°C, while pelleted cells were resuspended in 1ml HBSS supplemented with 0.25% BSA and 10mM Hepes prior of filtering. Cell suspension was kept on ice for the entire isolation procedure.

Fluorescent in-situ hybridization

Dechorionated embryos were collected at the desired stage, washed in cold PBS and fixed in 4% PFA at 4°C for 1 hour. Fixed embryos were then dehydrated in increasing dilutions

of methanol (25, 50, 75, 100 %) and left from overnight to one month at -20°C. The embryos were rehydrated in decreasing dilutions of methanol (75, 50, 25%) and washed 5 times in PBST (0.1%) for 5 minutes with gentle agitation. In order to acclimatise the sample to the high temperature and the hybridization conditions, the embryos were incubated for 2 hours at 70°C in 200 µl of Hybridization Buffer (HB) (50% deionized formamide, 5X SSC, 0.1% Tween-20, 50 µg/ml of heparin bile salts, 500 µg/ml of extracted RNase-free tRNA, pH 6.0). From 50 to 100 ng of DIG-labelled RNA probes targeting *dazl* transcripts (Forward: ACTAAAGTTGTAGCTGGGCCT, Reverse: CCTGAGTGGGCGTTAATGTT) were added to the HB and incubated overnight at 70°C. The next day, the probes were removed by four washes in increasing dilutions of 2 X SSC (NaCl 0.3M; Sodium citrate 0.03M) at 70°C (25, 50, 75, 100 %). The sample was then washed twice in 0.2% of SSC at 70°C 30 minutes each. The 0.2 X SSC was replaced by four serial dilutions in PBST 0.1% in order to remove any left-over probe. Washes were performed at room temperature with gentle agitation. The embryos were blocked in Blocking Buffer + Maleic Acid (Roche, 11585762001) for at least 3 hours before the anti-DIG antibody was added in a concentration of 1:5000 and incubated overnight at 4°C. The next day the antibody was washed five times in PBST 0.1% with gentle agitation at room temperature (30 minutes each) and fluorescently-tagged by horseradish peroxidase-catalysed signal amplification (Thermo Scientific, T20913)

Imaging

Embryos were placed in an agarose-coated petri dish and eventually embedded in agarose and imaged with a Zeiss 780 confocal or Z1 light sheet microscopes. Images were taken with the Zeiss ZEN pro 2.0 acquisition software with standard settings. Fixed samples were mounted in glycerol-based VectaShield (Vector laboratories, H-1000, UK) on a slide and covered with a glass slip. When preservation of the body shape was required, imaging dishes with glass bottom were used to avoid disintegration of the embryos.

Nuclear-cytoplasmic fractionation and qPCR

Embryos were set on ice and dissociated as described earlier. Nuclear and cytoplasmic fractions were separated by two washes in nuclei isolation buffer (Tris-HCl pH 7.4 10mM, NaCl 10mM, MgCl2 3mM and 0.1% IGEPAL CA-630) followed by 5 minutes centrifugation

at 500 x g. RNA was extracted from the two fractions with the RNAeasy Mini Extraction kit (Qiagen, 74044, UK), converted in cDNA via the SuperScript™ III Reverse Transcriptase (Thermo Fisher, 18080093, UK) and used for qPCR.

Reduced Representation Bisulfite Sequencing (RRBS-Seq)

Genomic DNA were extracted from 4500-7500 sorted somatic cells or PGCs at high, dome and prim-5 stage in duplicates and digested with MspI at 37°C for 3 hours. The fragment ends were repaired with Klenow exo at 37°C for 50 minutes. Then Methylated Illumina Pair-end Adaptors were ligated to gDNA fragments using T4 DNA ligase. The bisulfite conversion were performed using Zymo Research EZ DNA Methylation Gold kit. Libraries were PCR amplified for twenty cycles using Platinum Taq DNA polymerase and sequenced on Illumina HiSeq 2500 on 50bp single-end mode.

Patch Bisulfite PCR

Genomic DNA were extracted from 5000 sorted somatic cells or PGCs at prim-5, with three replicates for each cell types. Whole Genome Amplified genomic DNA (WGA gDNA) were generated from prim-5 WT Tu fish embryos using GE GenomiPhi V2 DNA Amplification Kits. 30 ng of DNA from sorted cells, 100 ng of genomic DNA from prim-5 embryos (non-bisulfite conversion control) and 100 ng of WGA gDNA (hypomethylation control) were digested with HpyCH4V and NlaIII at 37°C for 1 hour following by heat inactivation for 20 minutes at 65°C. Then custom universal primers were ligated to the targeted fragments using HiFi Taq DNA ligase with the help of gene specific designed oligo patches. The reaction was incubated at 95°C for 15 minutes followed by 30 seconds at 94°C and 4 minutes at 65°C for 25 cycles, and was held at 4°C. Unligated DNA fragments were removed by Exo I and Exo III treatment at 37°C for 1 hour followed by heat inactivation at 80°C for 20 minutes. Bisulfite Conversion were performed following Zymo EZ DNA Methylation Gold kit manufacturer's instruction. This step was skipped for non-bisulfite control sample. The eluted DNA were PCR amplified using EpiMark Hot Start Taq DNA polymerase. Libraries were sequenced on Miseq 250bp paired-end mode.

ATAC-seq library preparation and sequencing

Two biological repeats for PGCs and somatic cells at high (wild type), prim-5 (wild type) and morpholino-injected prim-5 (5mm and MO) were prepared. Cells were sorted into 500 µl of cold PBS Mg-, Ca- and immediately treated for ATAC. Nuclear isolation and transposition reaction occurred as described in Buenrostro et al., 2013. Tagmented DNA libraries were purified with 1.2X volume of AMPure XP beads. DNA bound to the beads was washed twice in 80% ethanol and eluted in 20 µl of water. Indexed fragments were checked in concentration by qPCR, profiled by Bioanalyzer, equimolarly pooled and sequenced on an Illumina Next-Seq 550.

RNA-seq library preparation and sequencing

cDNA was prepared from two biological repeats according to manufacturer's instructions as follows. Two hundred cells were sorted in 8.5µl of water (0.2U/µl RNase inhibitor). Immediately after collection, cells were added with 1µl of lysis buffer (0.2U/µl RNase inhibitor) and 1 ul of ERCC Mix1-2 (final dilution $1x10^{-6}$) and flash-frozen. Reverse transcription was performed following the SMART-Seq v4 protocol. Indexed fragments were checked in concentration by qPCR, profiled by Bioanalyzer, equimolarly pooled and sequenced on an Illumina Next-Seq 550.

ATAC-seq analysis

Paired-end ATAC reads were mapped to the genome using Bowtie2, not allowing discordant mapping of reads (--no-discordant) and insert sizes larger than 5000bp (--maxins 5000). Results were filtered for mapping quality (10) and for mapping to chromosomes 1 to 25 with the exclusion of the mitochondrial chromosome and of contigs present in danRer7/10 assemblies. There was no removal of subsequent read pairs mapping to the same locus (so called duplicate removal).

Mapped and filtered ATAC reads were corrected for Tn5 transposase overhang by adding 5bp to the position of the start of the first read and by subtracting 4bp from the end of the second read in the read pair as described in Buenrostro et al., 2013. Both thus obtained Tn5 cut sites were extended by a fixed amount. For genome browser visualisation 25bp was added to Tn5 cut sites yielding two 51bp-long regions for each read pair. For PCA and SOM analysis a shorter 5bp extension was used. No selection for particular insert size fraction was used.

Enhancer calling in somatic cells and PGCs

The set of putative enhancers was obtained from ATAC-seq in the following way. MACS2 was used with options -f BED -g 1.412e9 --keep-dup all --nolambda --nomodel to call peaks in each sample and replicate separately. Four replicate pairs in PGCs and somatic cells at high and at prim5 stages were used to identify peaks reproducible across replicates with Irreproducible Discovery Rate (IDR)2 approach. At 5% IDR the number of reproducible peaks were for soma/high: 12,862; PGC/high: 14,029; soma/prim5: 79,494; PGC/prim5: 61,641. After a further removal of peaks within 500bp from any known transcript start (ENSEMBL version 79/91) and of width greater than 1000bp remaining peak numbers dropped to soma/high: 4,935; PGC/high: 5,097; soma/prim5: 55,784; PGC/prim5: 36,071.

As the last step four peaks sets were merged in a union and thus unified peaks were once again filtered for length <= 1000bp. This yielded a total of 70,612 candidate enhancers.

Principal Component Analysis and Self Organising Map clustering

In order to assign open chromatin scores to enhancers, windows of size 601bp around enhancer centre were used. ATAC signal levels from genome browser track were extracted (sum of signal values in 601bp bins proportional to the number of 5' ends of reads falling into these bins and normalised to the total number of reads in each sample) with the help of genomation package3 and saved into a matrix. For PCA and SOM, levels were log-transformed and “centred” by subtracting the mean of each matrix column (sample). For SOM, an additional row (enhancer) centering was performed effectively making SOM operate on log-fold change values.

Bioinformatic analysis for RRBS

Fastq files were aligned to ZV10 genome and processed using bismark(--bowtie1). Methylation level data were collected using bismark_methylation_extractor with parameters of --bedGraph --cutoff 6 --merge_non_CpG –comprehensive. Following methylation data were analysed using methylKit package in R with three replicates for each cell types at different developmental stages.

842 **Bioinformatic analysis for Patch bisulfite PCR**

843 Adapters were removed using cutadapt with parameters of -a
844 AGTGTGGGAGGGTAGTTGGTGTT -A ACTCCCCACCTTCCTCATTCTCTAAGACGGTGT --
845 minimum-length 10 for Read 1 and Read 2. Adapter trimmed fastq files were then aligned
846 to ZV10 genome and processed using bismark (--bowtie2). Methylation level data were
847 collected using bismark_methylation_extractor with parameters of --bedGraph --cutoff 6
848 --merge_non_CpG –comprehensive. The output CpG coverage files were converted to
849 colorBED files using a custom script. The colorBED files were loaded and visualized on
850 UCSC genome browser.

851

852 **RNA-seq analysis**

853 Fastq files were checked for quality by fastqc and trimmed by trimmomatic. Sequencing
854 reads were aligned to the zebrafish genome (danRer9/10) by STAR (v.2.6.1) with the
855 following settings: *--seedSearchStartLmax 12 --outFilterScoreMinOverLread 0.3 --*
856 *alignSJoverhangMin 15 --outFilterMismatchNmax 33 --outFilterMatchNminOverLread 0 --*
857 *outFilterType BySJout --outSAMunmapped Within --outSAMattributes NH HI AS NM MD --*
858 *outSAMstrandField intronMotif --outWigType bedGraph --quantMode GeneCounts*.
859 Raw read counts were loaded into R and differential expression analysis over samples
860 and stages was performed using DESeq2 (v.1.6.3) and maSigPro packages.

861

862 **Statistics**

863 All experiments for which statistical analyses were performed were repeated three times.
864 All sequencing experiments were performed in biological duplicates with the exception
865 of ATAC-seq for dome and 10-somites stages. Data from independent biological repeats
866 were pooled together and the statistical distribution of the dataset was evaluated upon
867 Shapiro-Wilk test. For normally distributed datasets, the p-value was estimated upon t-
868 test, while for non-normally distributed datasets, Wilcoxon test was used.

869

REFERENCES

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Dev. Camb. Engl. *144*, 3646–3658.

Batchelder, C., Dunn, M.A., Choy, B., Suh, Y., Cassie, C., Shim, E.Y., Shin, T.H., Mello, C., Seydoux, G., and Blackwell, T.K. (1999). Transcriptional repression by the Caenorhabditis elegans germ-line protein PIE-1. Genes Dev. *13*, 202–212.

Beer, R.L., and Draper, B.W. (2013). nanos3 maintains germline stem cells and expression of the conserved germline stem cell gene nanos2 in the zebrafish ovary. Dev. Biol. *374*, 308–318.

Bender, L.B., Cao, R., Zhang, Y., and Strome, S. (2004). The MES-2/MES-3/MES-6 complex and regulation of histone H3 methylation in C. elegans. Curr. Biol. CB *14*, 1639–1643.

Blaser, H., Eisenbeiss, S., Neumann, M., Reichman-Fried, M., Thisse, B., Thisse, C., and Raz, E. (2005). Transition from non-motile behaviour to directed migration during early PGC development in zebrafish. J. Cell Sci. *118*, 4027–4038.

Bogdanovic, O., Fernandez-Miñán, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J.C., and Gómez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. *22*, 2043–2053.

Bogdanović, O., Smits, A.H., de la Calle Mustienes, E., Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat. Genet. *48*, 417–426.

Breiling, A., and Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin *8*, 24.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Butuči, M., Williams, A.B., Wong, M.M., Kramer, B., and Michael, W.M. (2015). Zygotic Genome Activation Triggers Chromosome Damage and Checkpoint Signaling in C. elegans Primordial Germ Cells. Dev. Cell *34*, 85–95.

Charlesworth, A., Wilczynska, A., Thampi, P., Cox, L.L., and MacNicol, A.M. (2006). Musashi regulates the temporal order of mRNA translation during Xenopus oocyte maturation. EMBO J. *25*, 2792–2801.

903 Chen, H.-H., Welling, M., Bloch, D.B., Muñoz, J., Mientjes, E., Chen, X., Tramp, C., Wu, J.,
904 Yabuuchi, A., Chou, Y.-F., et al. (2014). DAZL limits pluripotency, differentiation, and
905 apoptosis in developing primordial germ cells. Stem Cell Rep. *3*, 892–904.
906 Doitsidou, M., Reichman-Fried, M., Stebler, J., Köprunner, M., Dörries, J., Meyer, D.,
907 Esguerra, C.V., Leung, T., and Raz, E. (2002). Guidance of primordial germ cell migration
908 by the chemokine SDF-1. Cell*111*, 647–659.
909 Draper, B.W., McCallum, C.M., and Moens, C.B. (2007). nanos1 is required to maintain
910 oocyte production in adult zebrafish. Dev. Biol.*305*, 589–598.
911 Eddy, E.M. (1975). Germ plasm and the differentiation of the germ cell line. Int. Rev. Cytol.
912 *43*, 229–280.
913 Eno, C., and Pelegri, F. (2016). Germ Cell Determinant Transmission, Segregation, and
914 Function in the Zebrafish Embryo. In Insights from Animal Reproduction, R. Payan
915 Carreira, ed. (InTech), p.
916 Eno, C., Gomez, T., Slusarski, D.C., and Pelegri, F. (2018). Slow calcium waves mediate
917 furrow microtubule reorganization and germ plasm compaction in the early zebrafish
918 embryo. Dev. Camb. Engl. *145*.
919 Forbes, A., and Lehmann, R. (1998). Nanos and Pumilio have critical roles in the
920 development and function of Drosophila germline stem cells. Dev. Camb. Engl. *125*, 679–
921 690.
922 Gaydos, L.J., Rechtsteiner, A., Egelhofer, T.A., Carroll, C.R., and Strome, S. (2012).
923 Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate
924 gene expression in C. elegans germ cells. Cell Rep. *2*, 1169–1177.
925 Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-
926 Padilla, M.-E. (2009). TBP2 is essential for germ cell development by regulating
927 transcription and chromatin condensation in the oocyte. Genes Dev. *23*, 2210–2223.
928 Gerovska, D., and Araúzo-Bravo, M.J. (2016). Does mouse embryo primordial germ cell
929 activation start before implantation as suggested by single-cell transcriptomics
930 dynamics? Mol. Hum. Reprod. *22*, 208–225.
931 Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and
932 Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of
933 maternal mRNAs. Science *312*, 75–79.

Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., and Clark, A.T. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell *161*, 1425–1436.

Gorokhova, S., Bibert, S., Geering, K., and Heintz, N. (2007). A novel family of transmembrane proteins interacting with beta subunits of the Na,K-ATPase. Hum. Mol. Genet. *16*, 2394–2410.

Gross-Thebing, T., Yigit, S., Pfeiffer, J., Reichman-Fried, M., Bandemer, J., Ruckert, C., Rathmer, C., Goudarzi, M., Stehling, M., Tarbashevich, K., et al. (2017). The Vertebrate Protein Dead End Maintains Primordial Germ Cell Fate by Inhibiting Somatic Differentiation. Dev. Cell *43*, 704–715.e5.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. Cell *161*, 1437–1452.

Guo, H., Hu, B., Yan, L., Yong, J., Wu, Y., Gao, Y., Guo, F., Hou, Y., Fan, X., Dong, J., et al. (2017). DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. Cell Res. *27*, 165–183.

Guven-Ozkan, T., Nishi, Y., Robertson, S.M., and Lin, R. (2008). Global transcriptional repression in C. elegans germline precursors by regulated sequestration of TAF-4. Cell *135*, 149–160.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381–385.

Hadzhiev, Y., Qureshi, H.K., Wheatley, L., Cooper, L., Jasiulewicz, A., Van Nguyen, H., Wragg, J.W., Poovathumkadavil, D., Conic, S., Bajan, S., et al. (2019). A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation. Nat. Commun. *10*, 691.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature *451*, 730–733.

Harvey, S.A., Sealy, I., Kettleborough, R., Fenyes, F., White, R., Stemple, D., and Smith, J.C. (2013). Identification of the zebrafish maternal and paternal transcriptomes. Dev. Camb. Engl. *140*, 2703–2710.

966 Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M.,
967 Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic
968 reprogramming enables the transition from primordial germ cell to gonocyte. Nature *555*,
969 392–396.
970 Hosokawa, M., Shoji, M., Kitamura, K., Tanaka, T., Noce, T., Chuma, S., and Nakatsuji, N.
971 (2007). Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: domain
972 composition, intracellular localization, and function in male germ cells in mice. Dev. Biol.
973 *301*, 38–52.
974 Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H.,
975 Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., et al. (2007). A role for Piwi and piRNAs in
976 germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69–82.
977 Huang, H.-Y., Houwing, S., Kaaij, L.J.T., Meppelink, A., Redl, S., Gauci, S., Vos, H., Draper,
978 B.W., Moens, C.B., Burgering, B.M., et al. (2011). Tdrd1 acts as a molecular scaffold for Piwi
979 proteins and piRNA targets in zebrafish. EMBO J. *30*, 3298–3308.
980 Iguchi, N., Tobias, J.W., and Hecht, N.B. (2006). Expression profiling reveals meiotic male
981 germ cell mRNAs that are translationally up- and down-regulated. Proc. Natl. Acad. Sci. U.
982 S. A. *103*, 7712–7717.
983 Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S.,
984 Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial
985 germ cell fate. Cell *160*, 253–268.
986 Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B.
987 (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res. *21*, 1543–
988 1551.
989 Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic Loss and
990 Gain of Chromatin Architecture throughout Zebrafish Development. Cell Rep. *24*, 1-10.e4.
991 Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A.F., Slanchev, K., le Sage, C.,
992 Nagel, R., Voorhoeve, P.M., van Duijse, J., Ørom, U.A., et al. (2007). RNA-binding protein
993 Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273–1286.
994 Knaut, H., Pelegri, F., Bohmann, K., Schwarz, H., and Nüsslein-Volhard, C. (2000). Zebrafish
995 vasa RNA but not its protein is a component of the germ plasm and segregates
996 asymmetrically before germline specification. J. Cell Biol. *149*, 875–888.
997 Köprunner, M., Thisse, C., Thisse, B., and Raz, E. (2001). A zebrafish nanos-related gene is
998 essential for the development of primordial germ cells. Genes Dev. *15*, 2877–2885.

999 Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008).
1000 Complex genome-wide transcription dynamics orchestrated by Blimp1 for the
1001 specification of the germ cell lineage in mice. Genes Dev. *22*, 1617–1635.
1002 Ladstätter, S., and Tachibana, K. (2019). Genomic insights into chromatin reprogramming
1003 to totipotency in embryos. J. Cell Biol. *218*, 70–82.
1004 Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P.,
1005 and Hogan, B.L. (1999a). Bmp4 is required for the generation of primordial germ cells in
1006 the mouse embryo. Genes Dev. *13*, 424–436.
1007 Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P.,
1008 and Hogan, B.L. (1999b). Bmp4 is required for the generation of primordial germ cells in
1009 the mouse embryo. Genes Dev. *13*, 424–436.
1010 Lee, C.-Y.S., Lu, T., and Seydoux, G. (2017). Nanos promotes epigenetic reprograming of
1011 the germline by down-regulation of the THAP transcription factor LIN-15B. ELife *6*.
1012 Lesch, B.J., and Page, D.C. (2014). Poised chromatin in the mammalian germ line.
1013 Development *141*, 3619–3626.
1014 Levine, E., Cupp, A.S., Miyashiro, L., and Skinner, M.K. (2000). Role of transforming growth
1015 factor-alpha and the epidermal growth factor receptor in embryonic rat testis
1016 development. Biol. Reprod. *62*, 477–490.
1017 Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing Chromatin
1018 Regulatory Landscape during Mouse Preimplantation Development. Cell *165*, 1375–
1019 1388.
1020 Macleod, D., Clark, V.H., and Bird, A. (1999). Absence of genome-wide changes in DNA
1021 methylation during development of the zebrafish. Nat. Genet. *23*, 139–140.
1022 Mello, C.C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Priess, J.R. (1996). The PIE-1
1023 protein and germline specification in C. elegans embryos. Nature *382*, 710–712.
1024 Merkhofer, E.C., Hu, P., and Johnson, T.L. (2014). Introduction to cotranscriptional RNA
1025 splicing. Methods Mol. Biol. Clifton NJ *1126*, 83–96.
1026 Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J.-L., Bomane, A., Cosson, B.,
1027 Eyras, E., Rasko, J.E.J., et al. (2017). IRFinder: assessing the impact of intron retention on
1028 mammalian gene expression. Genome Biol. *18*, 51.
1029 Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K.
1030 (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish
1031 miR-430. Curr. Biol. CB *16*, 2135–2142.

1032 Murphy, P.J., Wu, S.F., James, C.R., Wike, C.L., and Cairns, B.R. (2018). Placeholder
1033 Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. Cell *172*,
1034 993-1006.e13.
1035 Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004a). Drosophila cup is an eIF4E
1036 binding protein that associates with Bruno and regulates oskar mRNA translation in
1037 oogenesis. Dev. Cell *6*, 69–78.
1038 Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004b). Drosophila cup is an eIF4E
1039 binding protein that associates with Bruno and regulates oskar mRNA translation in
1040 oogenesis. Dev. Cell *6*, 69–78.
1041 Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A
1042 Signaling Principle for the Specification of the Germ Cell Lineage in Mice. Cell *137*, 571–
1043 584.
1044 Ortega-Recalde, O., Day, R.C., Gemmell, N.J., and Hore, T.A. (2019). Zebrafish preserve
1045 global germline DNA methylation while sex-linked rDNA is amplified and demethylated
1046 during feminisation. Nat. Commun.*10*, 3053.
1047 Oulhen, N., Swartz, S.Z., Laird, J., Mascaro, A., and Wessel, G.M. (2017). Transient
1048 translational quiescence in primordial germ cells. Dev. Camb. Engl. *144*, 1201–1210.
1049 Ouyang, J.P.T., Folkmann, A., Bernard, L., Lee, C.-Y., Seroussi, U., Charlesworth, A.G.,
1050 Claycomb, J.M., and Seydoux, G. (2019). P Granules Protect RNA Interference Genes from
1051 Silencing by piRNAs. Dev. Cell *50*, 716-728.e6.
1052 Perino, M., and Veenstra, G.J.C. (2016). Chromatin Control of Developmental Dynamics
1053 and Plasticity. Dev. Cell *38*, 610–620.
1054 Raj, B., Gagnon, J.A., and Schier, A.F. (2018). Large-scale reconstruction of cell lineages
1055 using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT.
1056 Nat. Protoc. *13*, 2685–2713.
1057 Raz, E. (2003). Primordial germ-cell development: the zebrafish perspective. Nat. Rev.
1058 Genet. *4*, 690–700.

1059 Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T.M., Egelhofer, T.A., Wang, W., Kimura,
1060 H., Lieb, J.D., and Strome, S. (2010). The Histone H3K36 Methyltransferase MES-4 Acts
1061 Epigenetically to Transmit the Memory of Germline Gene Expression to Progeny. PLoS
1062 Genet. *6*, e1001091.
1063

1064 Riemer, S., Bontems, F., Krishnakumar, P., Gömann, J., and Dosch, R. (2015). A functional
1065 Bucky ball-GFP transgene visualizes germ plasm in living zebrafish. Gene Expr. Patterns
1066 GEP *18*, 44–52.
1067 Roovers, E.F., Kaaij, L.J.T., Redl, S., Bronkhorst, A.W., Wiebrands, K., de Jesus Domingues,
1068 A.M., Huang, H.-Y., Han, C.-T., Riemer, S., Dosch, R., et al. (2018). Tdrd6a Regulates the
1069 Aggregation of Buc into Functional Subcellular Compartments that Drive Germ Cell
1070 Specification. Dev. Cell *46*, 285-301.e9.
1071 Rothschild, S.C., Lahvic, J., Francescatto, L., McLeod, J.J.A., Burgess, S.M., and Tombes, R.M.
1072 (2013). CaMK-II activation is essential for zebrafish inner ear development and acts
1073 through Delta-Notch signaling. Dev. Biol. *381*, 179–188.
1074 Saitou, M., and Yamaji, M. (2012). Primordial germ cells in mice. Cold Spring Harb.
1075 Perspect. Biol. *4*.
1076 Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the
1077 specification of germ cell fate in mice. Nature *418*, 293–300.
1078 Seydoux, G., and Braun, R.E. (2006). Pathway to totipotency: lessons from germ cells. Cell
1079 *127*, 891–904.
1080 Siddall, N.A., McLaughlin, E.A., Marriner, N.L., and Hime, G.R. (2006). The RNA-binding
1081 protein Musashi is required intrinsically to maintain stem cell identity. Proc. Natl. Acad.
1082 Sci. U. S. A. *103*, 8402–8407.
1083 Skvortsova, K., Tarbashevich, K., Stehling, M., Lister, R., Irimia, M., Raz, E., and Bogdanovic,
1084 O. (2019). Retention of paternal DNA methylome in the developing zebrafish germline.
1085 Nat. Commun. *10*, 3054.
1086 Strasser, M.J., Mackenzie, N.C., Dumstrei, K., Nakkrasae, L.-I., Stebler, J., and Raz, E. (2008).
1087 Control over the morphology and segregation of Zebrafish germ cell granules during
1088 embryonic development. BMC Dev. Biol. *8*, 58.

1089 Strome, S., Kelly, W.G., Ercan, S., and Lieb, J.D. (2014). Regulation of the X chromosomes
1090 in Caenorhabditis elegans. Cold Spring Harb. Perspect. Biol. *6*.
1091
1092 Strome, S., and Updike, D. (2015). Specifying and protecting germ cell fate. Nat. Rev. Mol.
1093 Cell Biol. *16*, 406–416.

1094 Suzuki, A., Saba, R., Miyoshi, K., Morita, Y., and Saga, Y. (2012). Interaction between
1095 NANOS2 and the CCR4-NOT deadenylation complex is essential for male germ cell
1096 development in mouse. PloS One *7*, e33558.
1097 Suzuki, H., Saba, R., Sada, A., and Saga, Y. (2010). The Nanos3-3'UTR is required for germ
1098 cell specific NANOS3 expression in mouse embryos. PloS One *5*, e9300.
1099 Tada, H., Mochii, M., Orii, H., and Watanabe, K. (2012). Ectopic formation of primordial
1100 germ cells by transplantation of the germ plasm: direct evidence for germ cell
1101 determinant in Xenopus. Dev. Biol. *371*, 86–93.
1102 Tang, W.W.C., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A.,
1103 Chinnery, P.F., and Surani, M.A. (2015). A Unique Gene Regulatory Network Resets the
1104 Human Germline Epigenome for Development. Cell *161*, 1453–1467.
1105 Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear
1106 pore complex environment in the C. elegans germ line. J. Cell Biol. *192*, 939–948.
1107 Varley, K.E., and Mitra, R.D. (2010). Bisulfite Patch PCR enables multiplexed sequencing
1108 of promoter methylation across cancer samples. Genome Res. *20*, 1279–1287.
1109 Vasileva, A., Tiedau, D., Firooznia, A., Müller-Reichert, T., and Jessberger, R. (2009). Tdrd6
1110 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA
1111 expression. Curr. Biol. CB *19*, 630–639.
1112 Weidinger, G., Wolke, U., Köprunner, M., Klinger, M., and Raz, E. (1999). Identification of
1113 tissues and patterning events required for distinct steps in early migration of zebrafish
1114 primordial germ cells. Dev. Camb. Engl. *126*, 5295–5307.
1115 Weidinger, G., Stebler, J., Slanchev, K., Dumstrei, K., Wise, C., Lovell-Badge, R., Thisse, C.,
1116 Thisse, B., and Raz, E. (2003). dead end, a novel vertebrate germ plasm component, is
1117 required for zebrafish primordial germ cell migration and survival. Curr. Biol. CB *13*,
1118 1429–1434.
1119 White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy,
1120 D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course
1121 of embryonic development in zebrafish. ELife *6*.
1122 Wilhelm, J.E., Hilton, M., Amos, Q., and Henzel, W.J. (2003). Cup is an eIF4E binding protein
1123 required for both the translational repression of oskar and the recruitment of Barentsz.
1124 J. Cell Biol. *163*, 1197–1204.
1125 Williamson, A., and Lehmann, R. (1996). Germ cell development in Drosophila. Annu. Rev.
1126 Cell Dev. Biol. *12*, 365–391.

1127 Wu, S.-F., Zhang, H., and Cairns, B.R. (2011). Genes for embryo development are packaged
1128 in blocks of multivalent chromatin in zebrafish sperm. Genome Res. *21*, 578–589.
1129 Ying, Y., Qi, X., and Zhao, G.Q. (2001). Induction of primordial germ cells from murine
1130 epiblasts by synergistic action of BMP4 and BMP8B signaling pathways. Proc. Natl. Acad.
1131 Sci. U. S. A. *98*, 7858–7862.
1132 Yoon, C., Kawakami, K., and Hopkins, N. (1997). Zebrafish vasa homologue RNA is
1133 localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the
1134 primordial germ cells. Dev. Camb. Engl. *124*, 3157–3165.
1135 Zhang, B., Wu, X., Zhang, W., Shen, W., Sun, Q., Liu, K., Zhang, Y., Wang, Q., Li, Y., Meng, A.,
1136 et al. (2018). Widespread Enhancer Dememorization and Promoter Priming during
1137 Parental-to-Zygotic Transition. Mol. Cell *72*, 673-686.e6.
1138 Zhang, W., Liu, Y., Zhang, M., Zhu, C., and Lu, Y. (2017). Quantitative reproducibility
1139 analysis for identifying reproducible targets from high-throughput experiments. BMC
1140 Syst. Biol. *11*, 73.

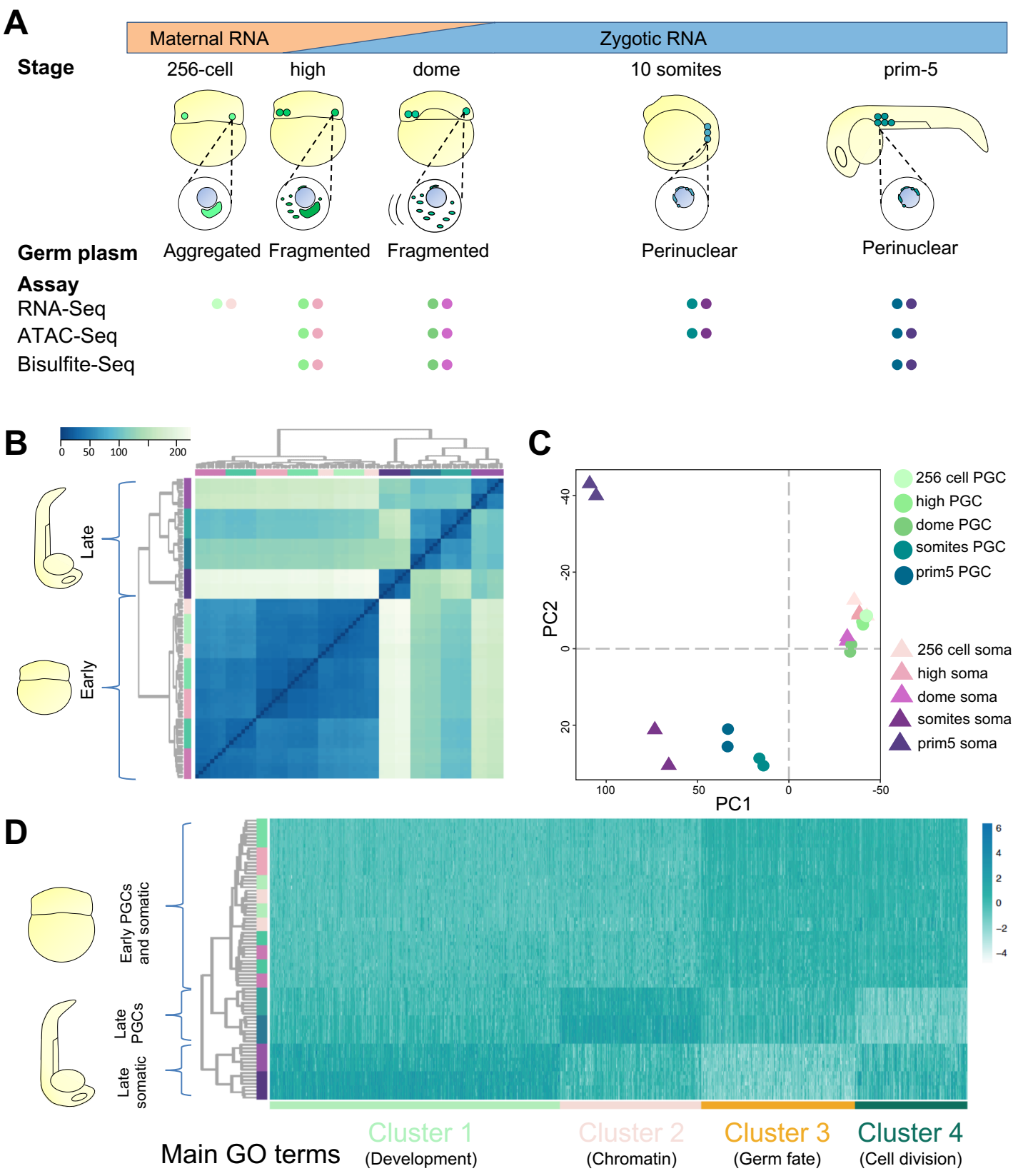


Figure 1. **Characterisation of PGC transcriptome highlights early developmental similarities and late divergence between PGCs and somatic cells**. (A) Key developmental stages used in the study are shown. Time points were selected according to various phases of germ plasm distribution/PGCs localisation. The early stages span zygotic genome activation including the first wave at 256-cell stage. NGS assays performed for each time point are shown as coloured dots; PGCs and somatic cells are in shades of green and purple respectively. (B-C) Unsupervised hierarchical clustering heatmap and two-dimensional PCA plot show developmental trends of PGC and somatic cell transcriptomes during development. (D) Four clusters of differential gene expression reported as heatmap upon k-mean-based clustering over development and cell type.

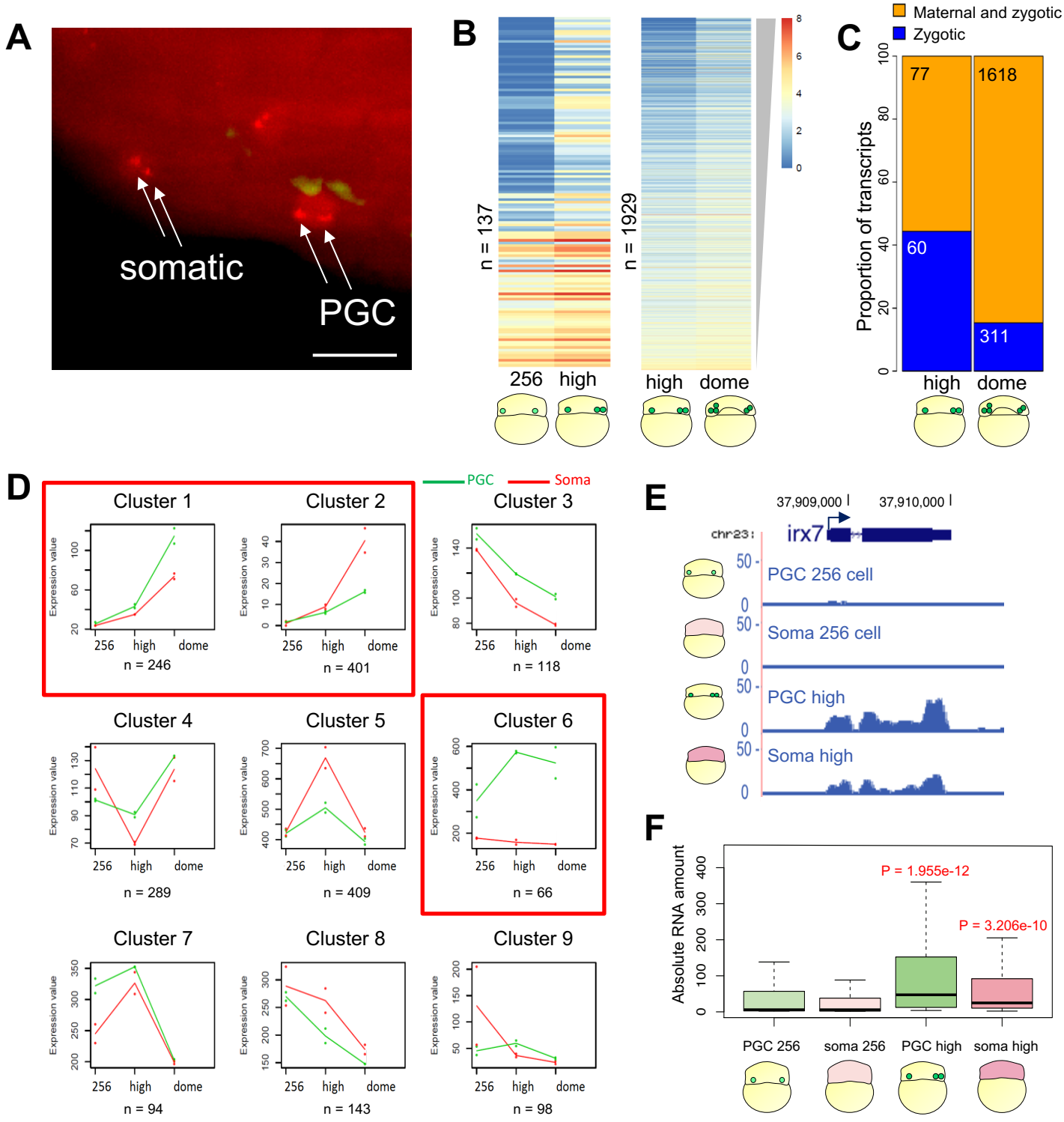


Figure 2. PGCs do not delay the major wave of transcriptional activation. (A) Maximum intensity projection of a multi-stack image showing *miR-430* transcription foci (arrows) detected by fluorescently-tagged Morpholinos (red) in germ plasm-carrying cells and somatic cells at 512-cell stage. Germ plasm-localised GFP-Buc (green) marks PGCs. Scale bar is 30 µm. (B) Heatmap of upregulated genes sorted by $\log_2$FC in the germ plasm-carrying cells between 256-cell and high stages and high and dome stages. Colours represent the $\log_2$(tpm+1). P adjusted < 0.1. (C) Proportion of zygotic and maternal/zygotic genes upregulated in the germ plasm-carrying cells at high and dome stages. (D) Clusters of gene expression trends among three developmental stages in PGCs and somatic cells. Median profiles of gene expression values are plotted for each trend in green (PGCs) and red (somatic cells) and represent the gene read counts normalised by the DESeq2-calculated size factor. Red squares highlight clusters supporting transcriptional activation in the PGCs. (E) Genome browser view of normalised RNA-seq reads for *irx7* gene. (F) Absolute RNA concentration normalised to internal control (ERCC spike-in) for genes upregulated in germ plasm-carrying cells between 256-cell and high stage. P-values are calculated by Wilcoxon test.

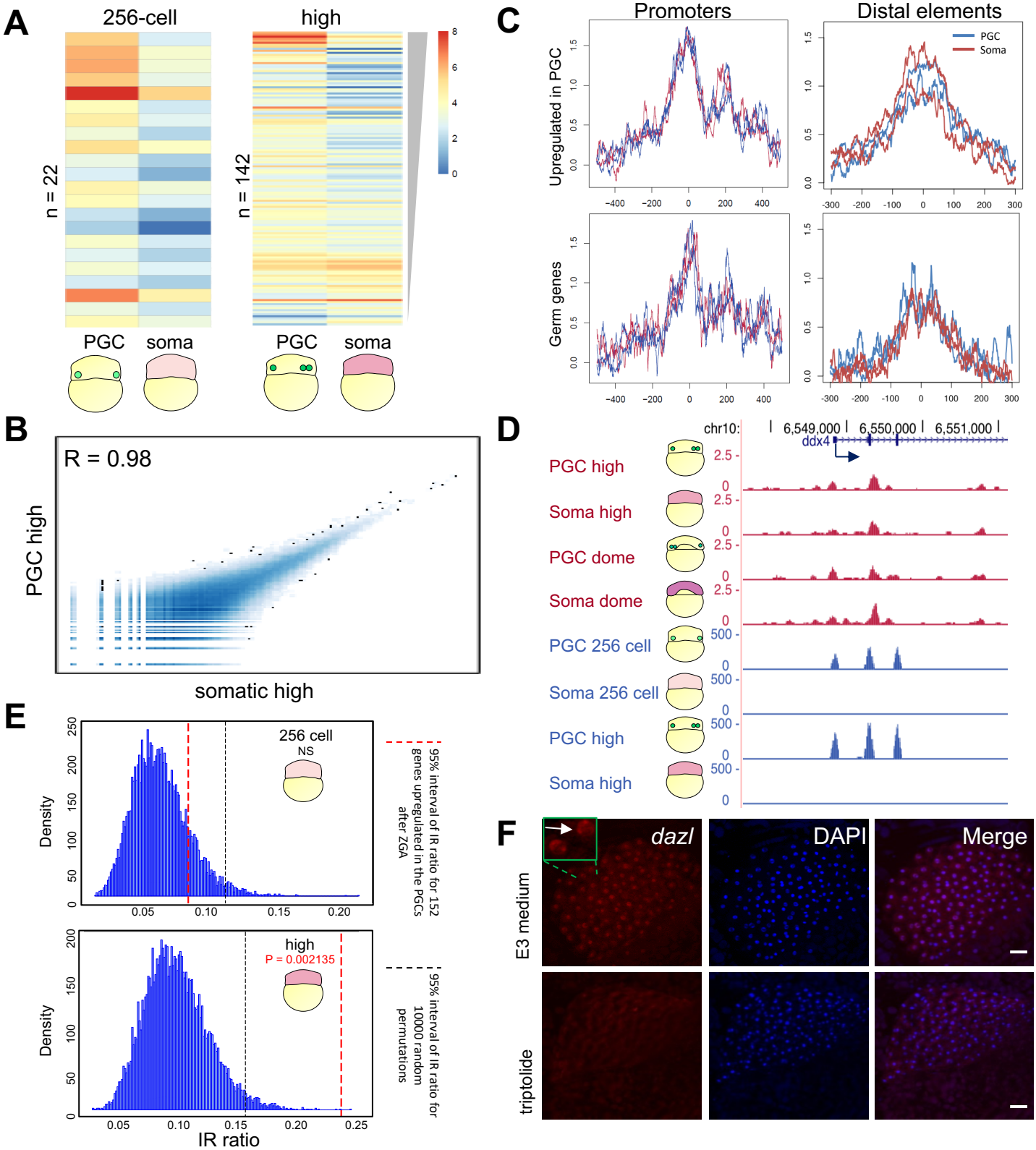


Figure 3. Differential transcriptome between PGCs and somatic cells at early stages is not caused by differential transcription. (A) Heatmap of differentially expressed genes sorted by log_2FC between PGCs and somatic cells at 256-cell and high stage. Colours represent log_2(tpm+1) values. P adjusted < 0.05. (B) Correlation plot for open chromatin regions detected in PGCs and somatic cells at high stage. (C) Average accessibility signal at promoters and distal elements of PGC and somatic cells for subgroups of genes at high stage. Promoters are aligned to transcription start site, while distal elements are aligned to peak centre. (D) Genome browser view of normalised ATAC-seq (magenta) and RNA-seq (blue) reads for PGC and somatic cells at stages spanning ZGA. (E) Intron retention ratio before and after ZGA in the somatic cells of genes upregulated in the PGCs after ZGA. P-value is calculated by t-test. Statistical significance was calculated upon 10000 random permutation whose density is shown before and after ZGA. Black and red dashed lines show 95% significance interval for the 10000 permutations and the gene subgroup respectively. NS = not significant. (F) In-situ hybridization for *dazl* pre-mRNA at high stage. Scale bar is 50 µm.

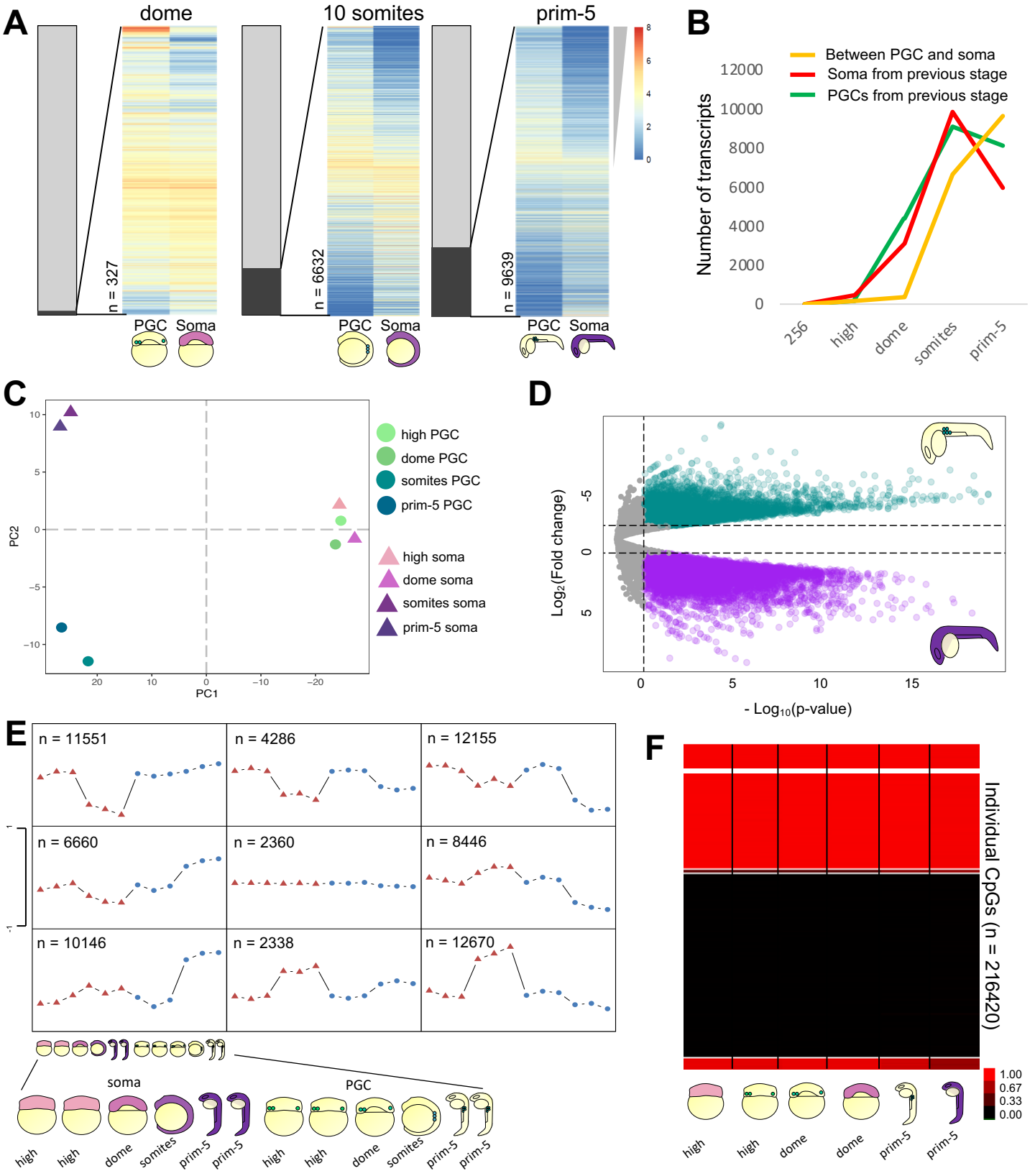


Figure 4. Gradual acquisition of germ identity is accompanied by epigenetic changes. (A) Heatmap of differentially expressed genes sorted by log_2FC between PGCs and somatic cells. Grey bars represent the proportion of differentially expressed genes over the total transcriptome. Colours represent $log_2(tpm+1)$ values. (B) Line-chart for gene counts. Upregulated genes from previous stage are in red (somatic) or green (PGCs). The orange line shows number of genes differentially expressed between PGCs and somatic cells at each stage. (C) Two-dimensional PCA plot of ATAC-seq profiles. (D) Volcano plot for regions of open chromatin between PGCs and somatic cells at prim-5 stage (log_2FC threshold = +/-1, p adjusted < 0.05). (E) Self-organising map of open chromatin regions. PGCs and somatic cells are shown as blue circles and red triangles respectively. Schematic of embryos as in Figure 1A and 1B (F) Methylation status in identified CpGs in PGCs and somatic cells at dome and prim-5 stages.

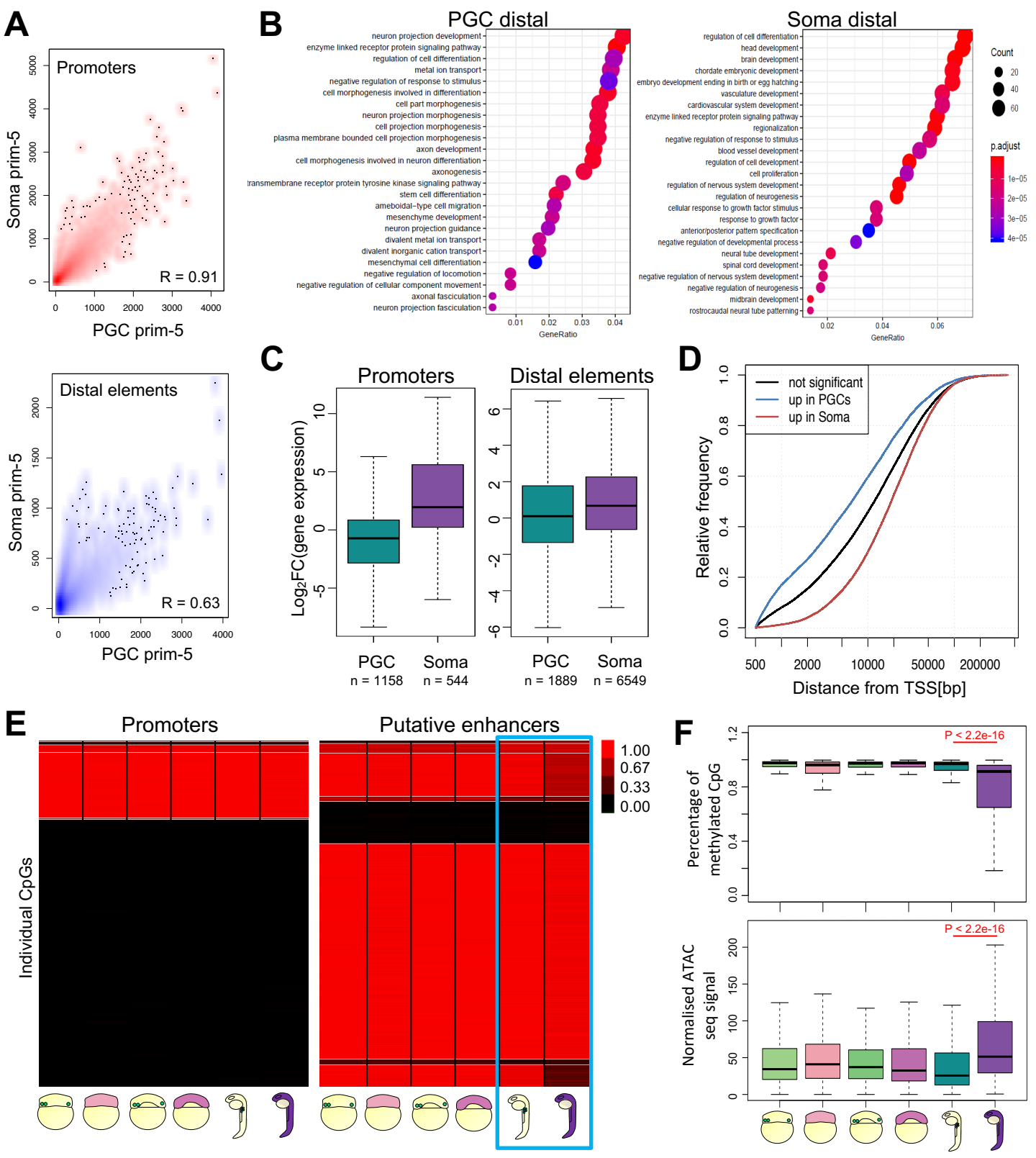


Figure 5. PGCs do not open chromatin at regions identified as putative enhancers. (A) Correlation scatter plots of ATAC peaks between PGCs and somatic cells for promoters and distal elements at prim-5 stage. (B) Developmental processes GO analysis for genes associated with differential chromatin accessible putative enhancers in PGCs and somatic cells. (C) Fold change of gene expression for genes associated with ATAC peaks upregulated in PGCs or somatic cells. Negative and positive fold changes represent gene upregulation in PGCs and somatic cells respectively. (D) Cumulative frequency of open chromatin elements in relation to distance from the closest TSS. P adjusted < 0.05. Colours indicate elements near differentially expressed genes as indicated. (E) Heatmap of DNA methylation for CpGs overlapping promoters (left) and H3K4me1/H3K27ac-rich genomic sites (putative enhancers) in PGCs and somatic cells at high, dome and prim-5 stages. Light blue box highlights putative enhancers at prim-5 stages. (F) Quantification of methylated CpGs and chromatin accessibility at putative enhancer regions. P-value was calculated by Wilcoxon test. Outlier are omitted.

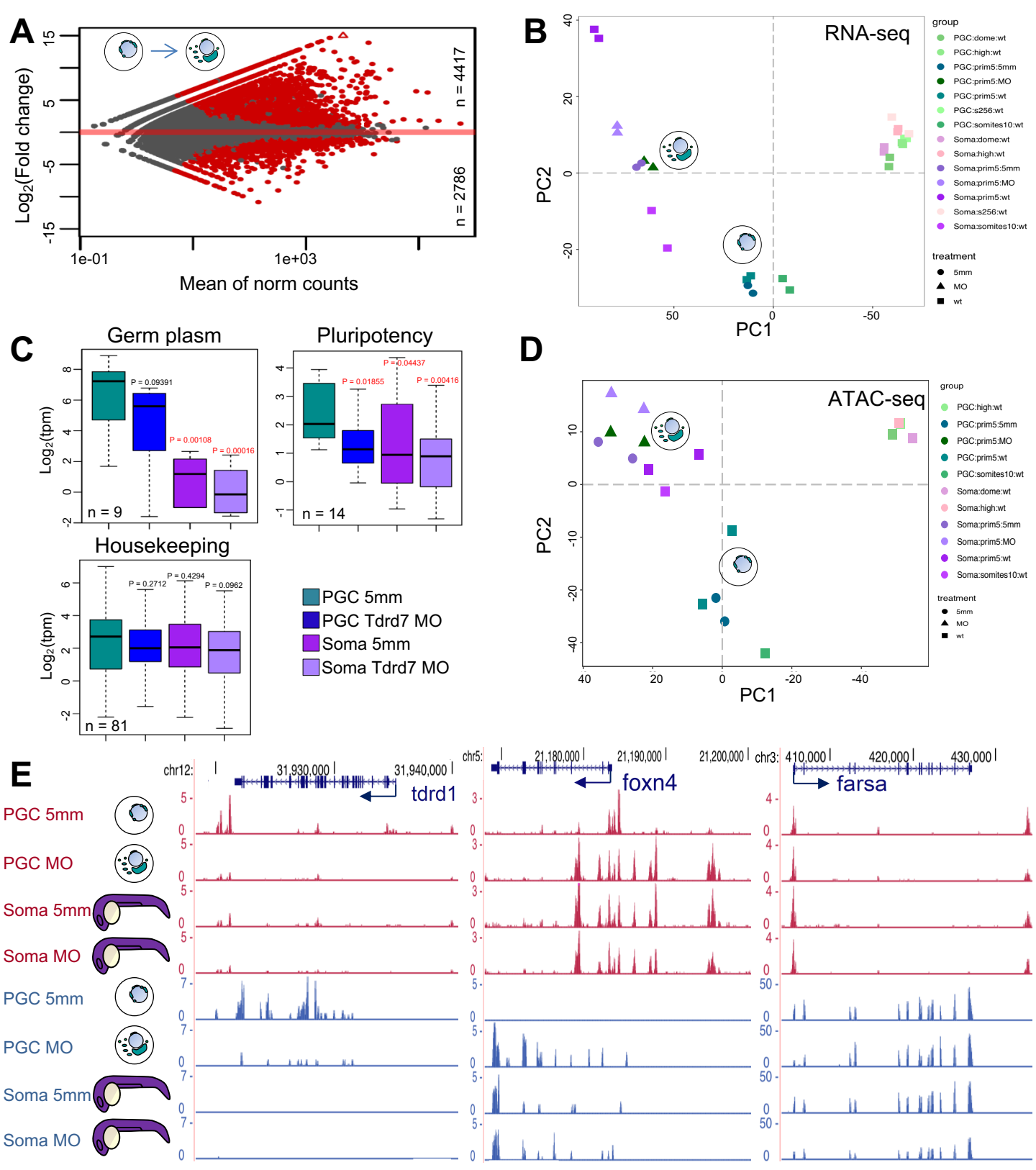


Figure 6. Tdrd7 is required for maintaining PGC fate. (A) Differential expression analysis for genes upregulated and downregulated by Tdrd7 KD in PGCs at prim-5 stage. Significantly differentially expressed genes are in red. P adjusted < 0.1, log_2FC < -1/> 1. (B) Global transcriptional variance shown as PCA plot for wild type and MO-injected PGCs and somatic cells. (C) Boxplots reporting normalised transcript levels (tpm values) for various gene subsets in MO-injected PGCs and somatic cells. P-values against control is shown. Red colour indicates significance ($p < 0.05$) based on Wilcoxon test. (D) PCA plots for only promoters and putative enhancers show PGCs diverging towards somatic-like chromatin state. (E) Genome browser view of ATAC profiles after morpholino injections. Open chromatin (ATAC-seq) is shown in magenta and transcript levels (RNA-seq) are shown in blue. Arrows show transcriptional directionality.

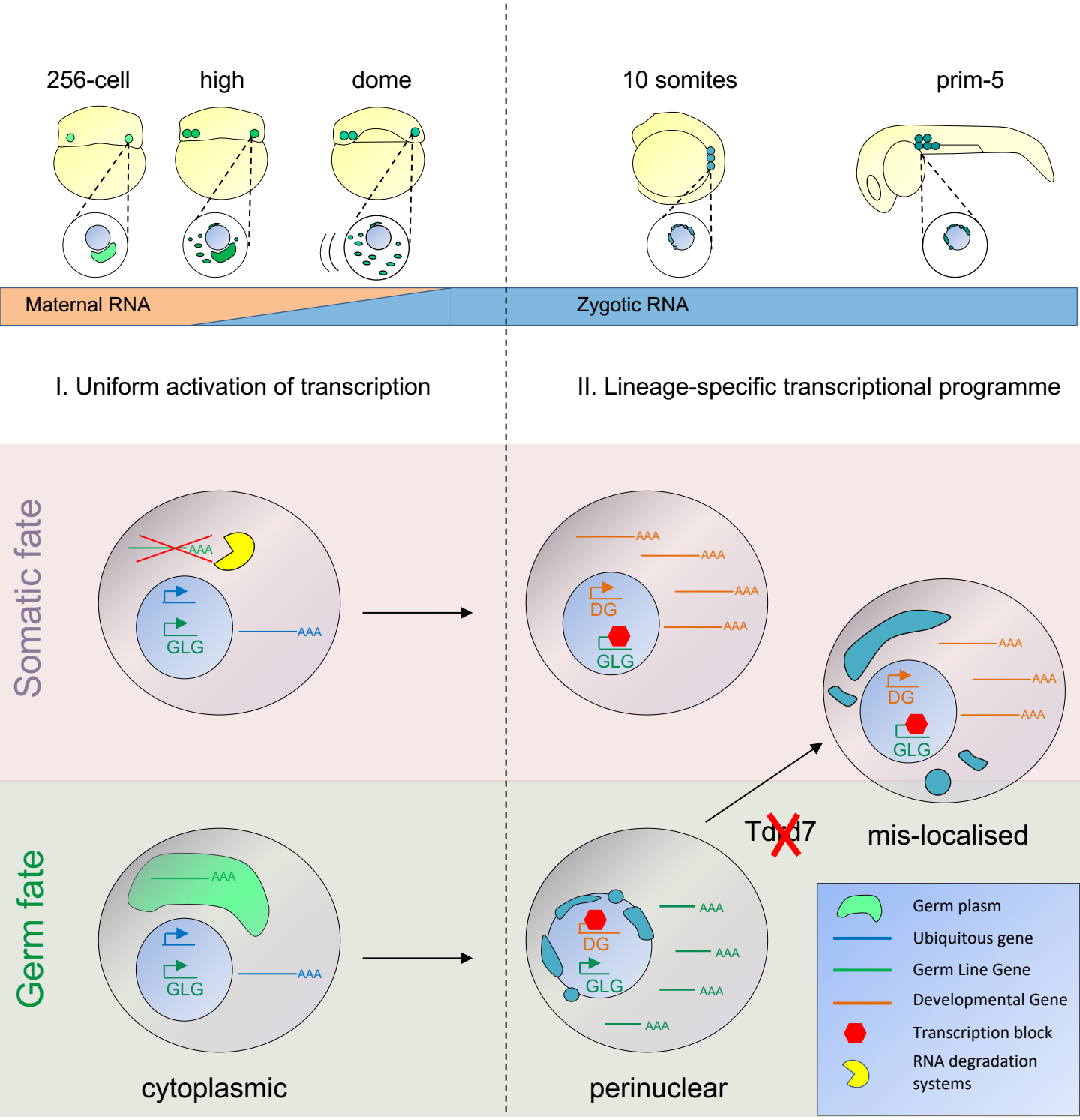
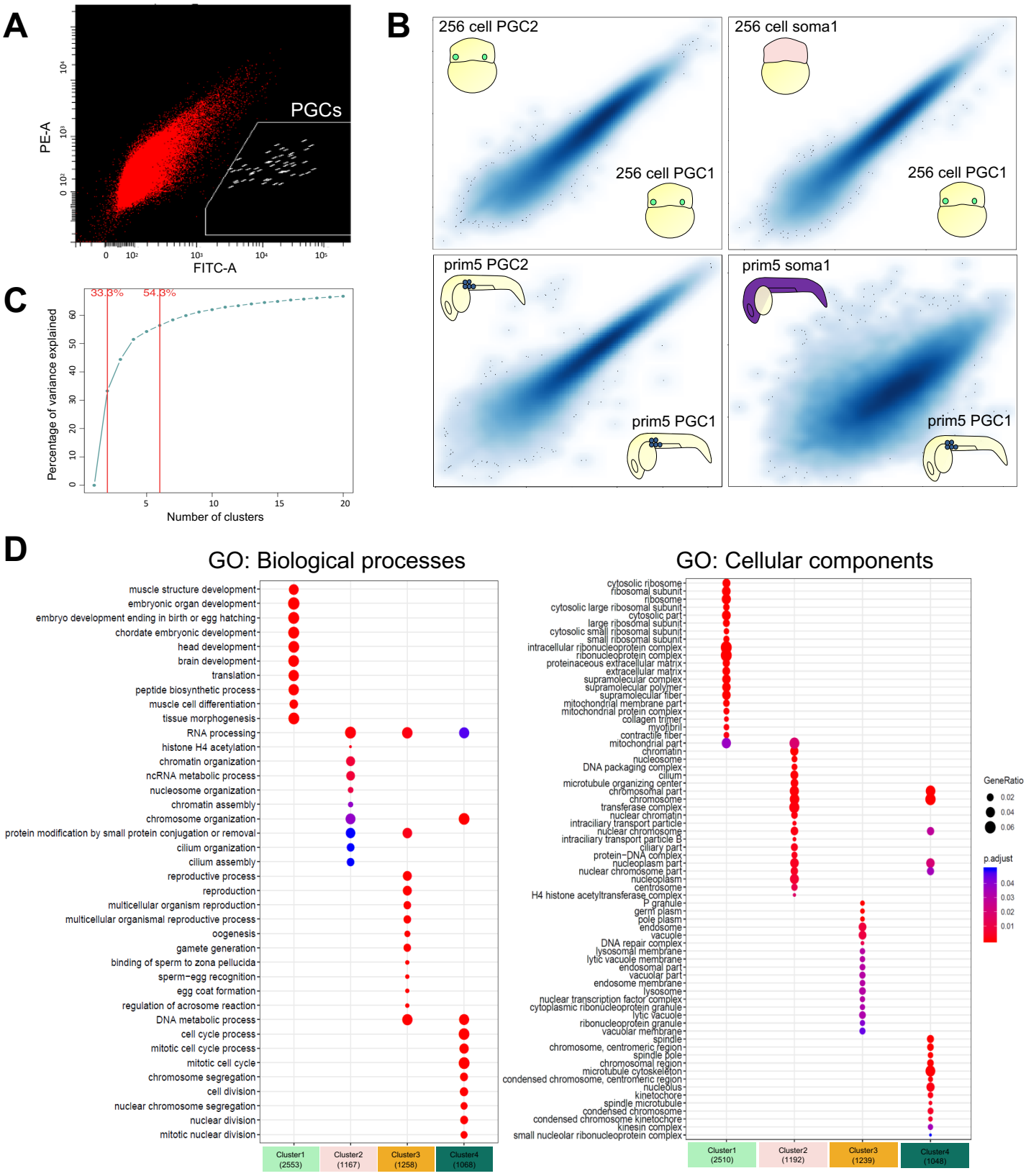
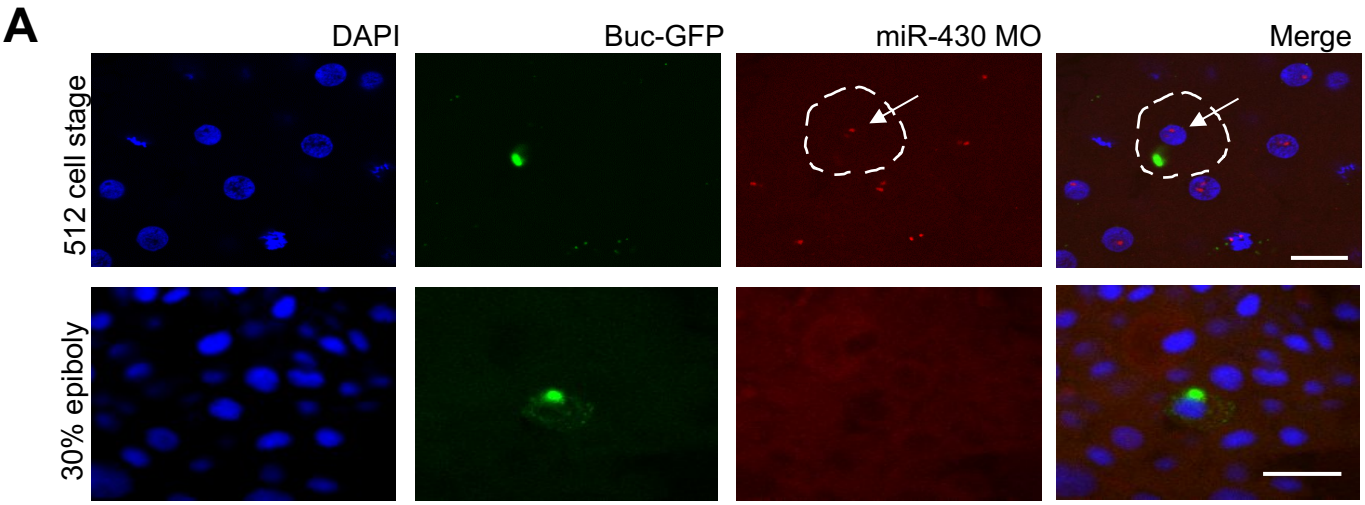
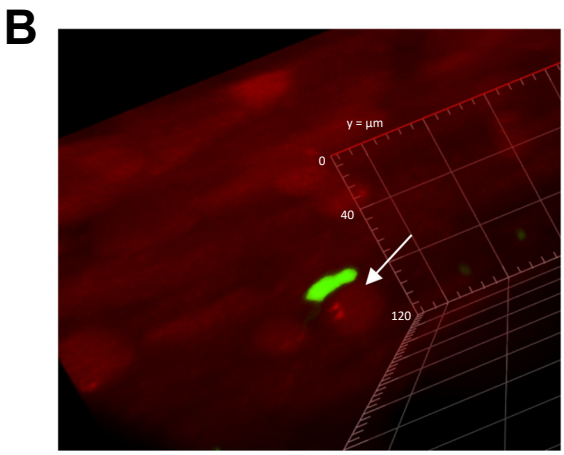


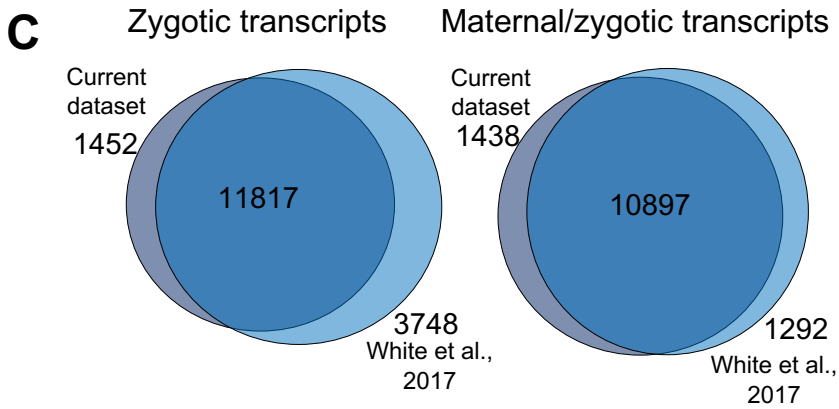
Figure 7. Two functions of the germ plasm during germ fate acquisition. During zygotic genome activation, both somatic cells and PGCs activate the genome similarly, including the activation of PGC specific genes. However, PGC specific genes are only stabilised in germ plasm-containing cells upon protection from the zygotic degradation machinery (yellow). During PGC migration, germ plasm components re-localise to the nuclear membrane. This coincides with acquisition of a PGC-specific transcriptional programme, involving transcriptional repression of developmental pathways blocking somatic differentiation. This process is Tdrd7 dependent as indicated by loss of germ plasm re-localisation and reactivation of somatic transcriptional programmes.

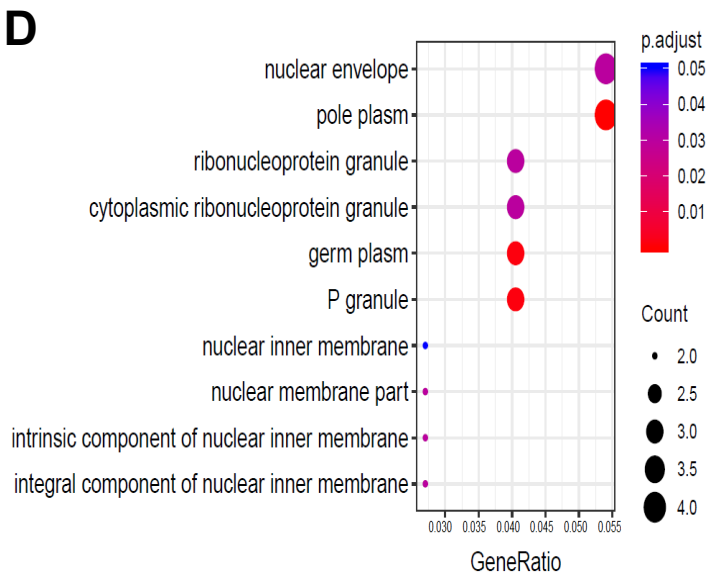


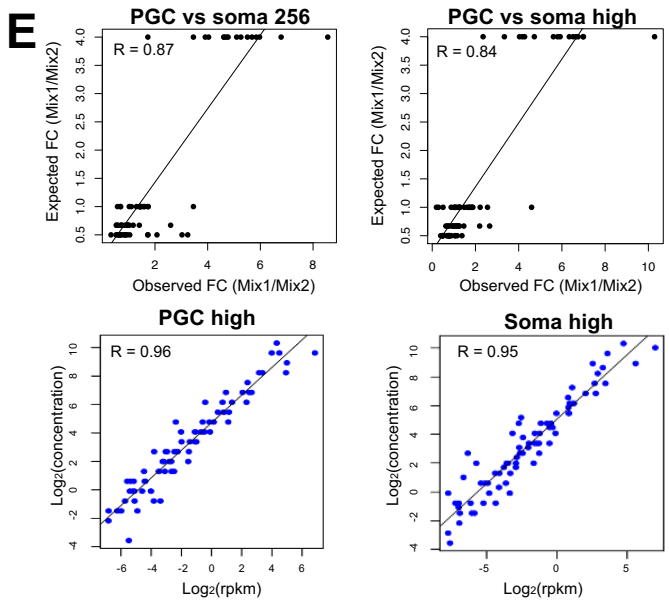
Supplementary Figure 1. **Characterisation of PGC transcriptome highlights early developmental similarities and late divergence between PGCs and somatic cells**. (A) FACS profile of GFP-expressing PGCs at prim-5 stage. GFP-positive and –negative cells are in white and red respectively. (B) Transcriptome correlation profiles after RNA-seq reads normalisation. Read counts for each gene were normalised for library size and gene length (tpm) and correlation among stages and cell types were compared. Reads are shown as Log_2(tpm+1). (C) Justification of number of k-means selected based on inter-sample variance. (D) Biological processes and cellular components GO analysis for the four identified expression groups by k-mean clustering (p adjusted < 0.05).



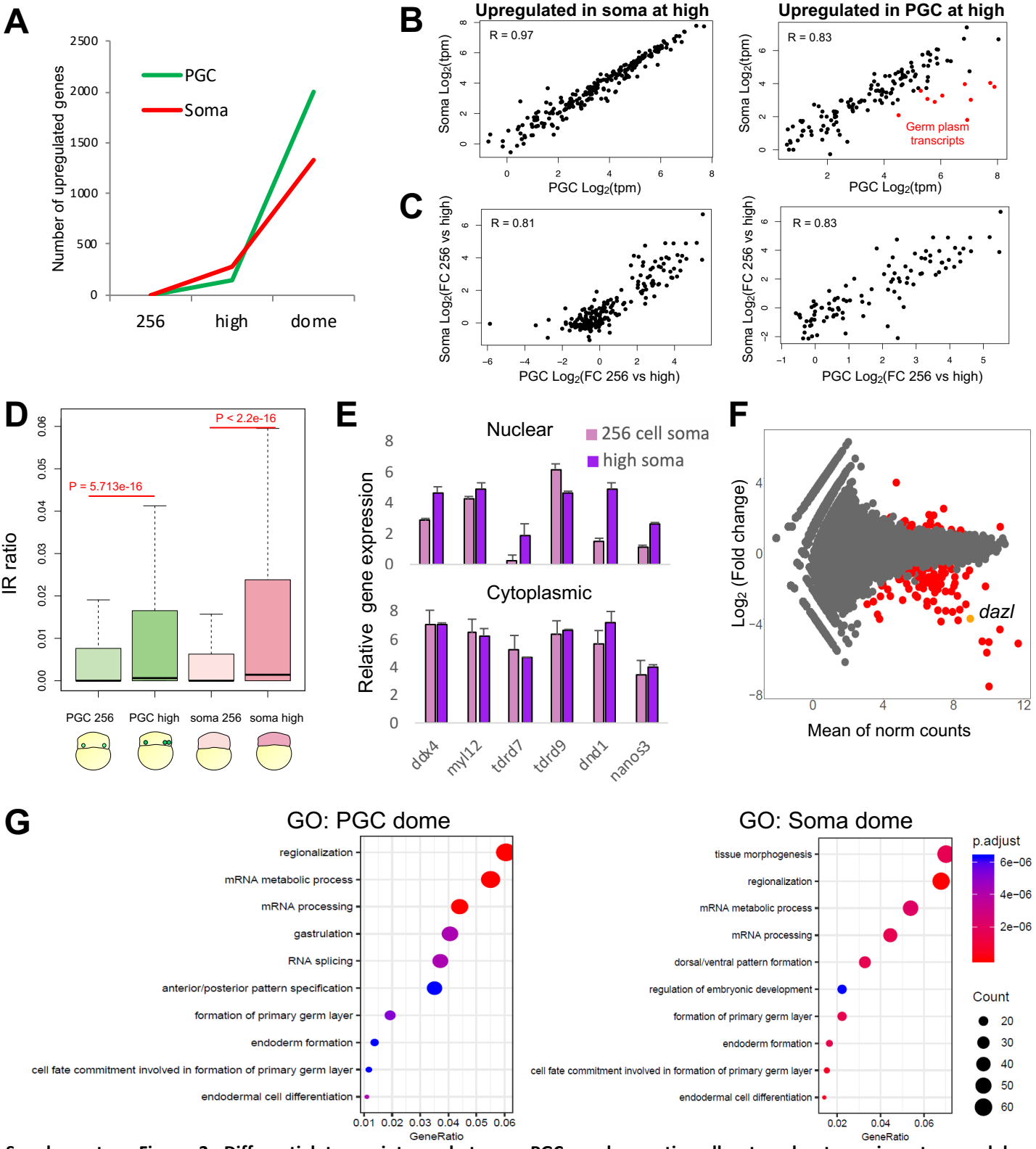




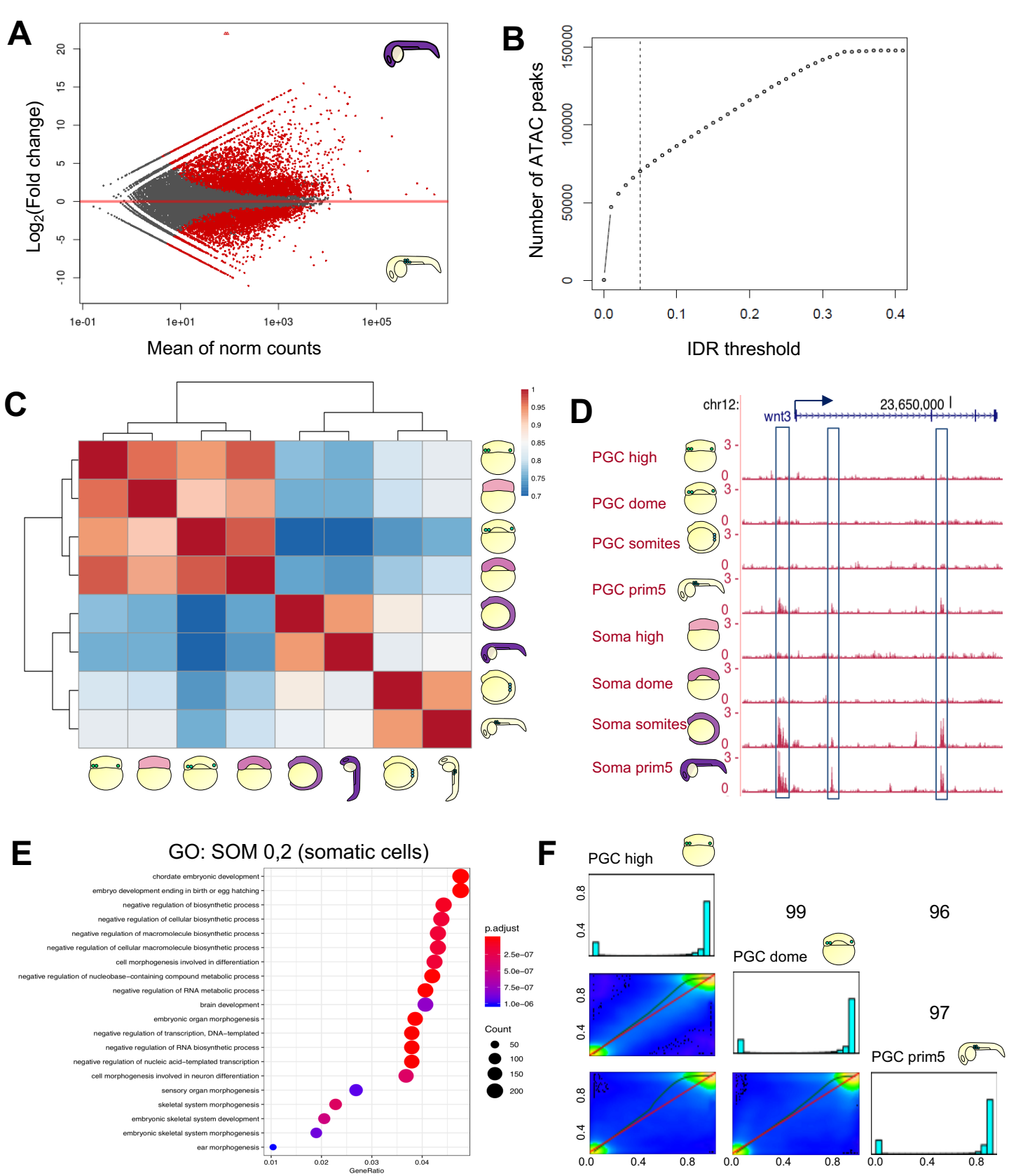




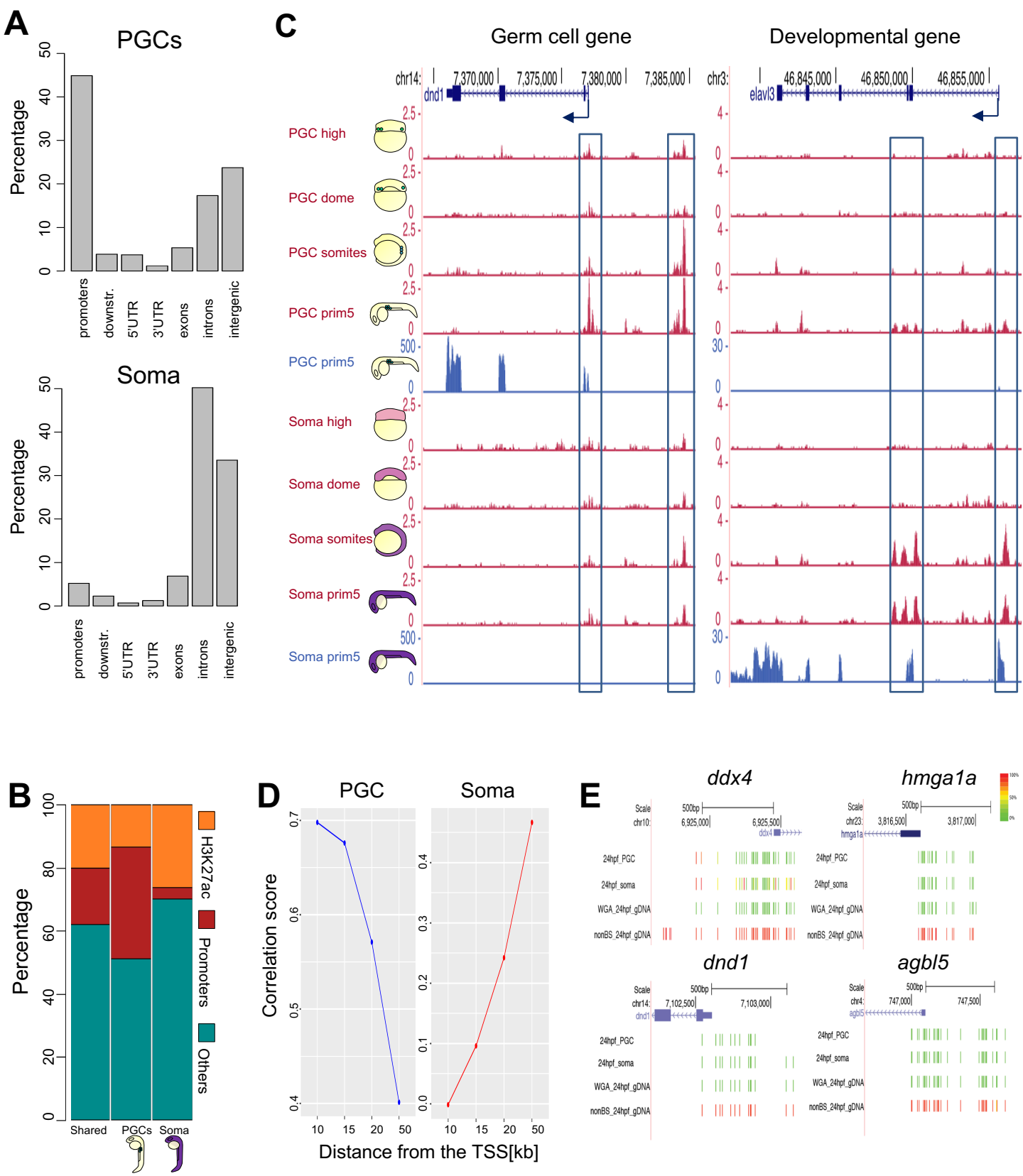
Supplementary Figure 2. PGCs do not delay transcriptional activation. (A) Maximum intensity projection of a multi-stack image of *miR-430* transcriptional activity at 512-cell stage and 30% epiboly. PGC boundaries are highlighted as white, dashed lines. Arrow indicates the nucleus of a germ plasm-carrying cell where transcriptional foci are detected. Scale bar is 30 µm. (B) 3D rendering of a *miR-430*-expressing nucleus (arrow) and a GFP-tagged germ granule. White arrow indicates the germ cell nucleus. (C) Venn diagram for predicted zygotic/maternal and zygotic genes from two independent RNA-seq datasets. (D) Cellular components GO analysis for genes upregulated in PGCs after the first wave of ZGA (from 256-cell to high stage). P adjusted < 0.05. (E) Representative regression scatter plots between expected vs observed ERCC mix ratios (top) and ERCC concentrations vs ERCC reads (bottom). The tpm threshold for normalised ERCC reads was 1.



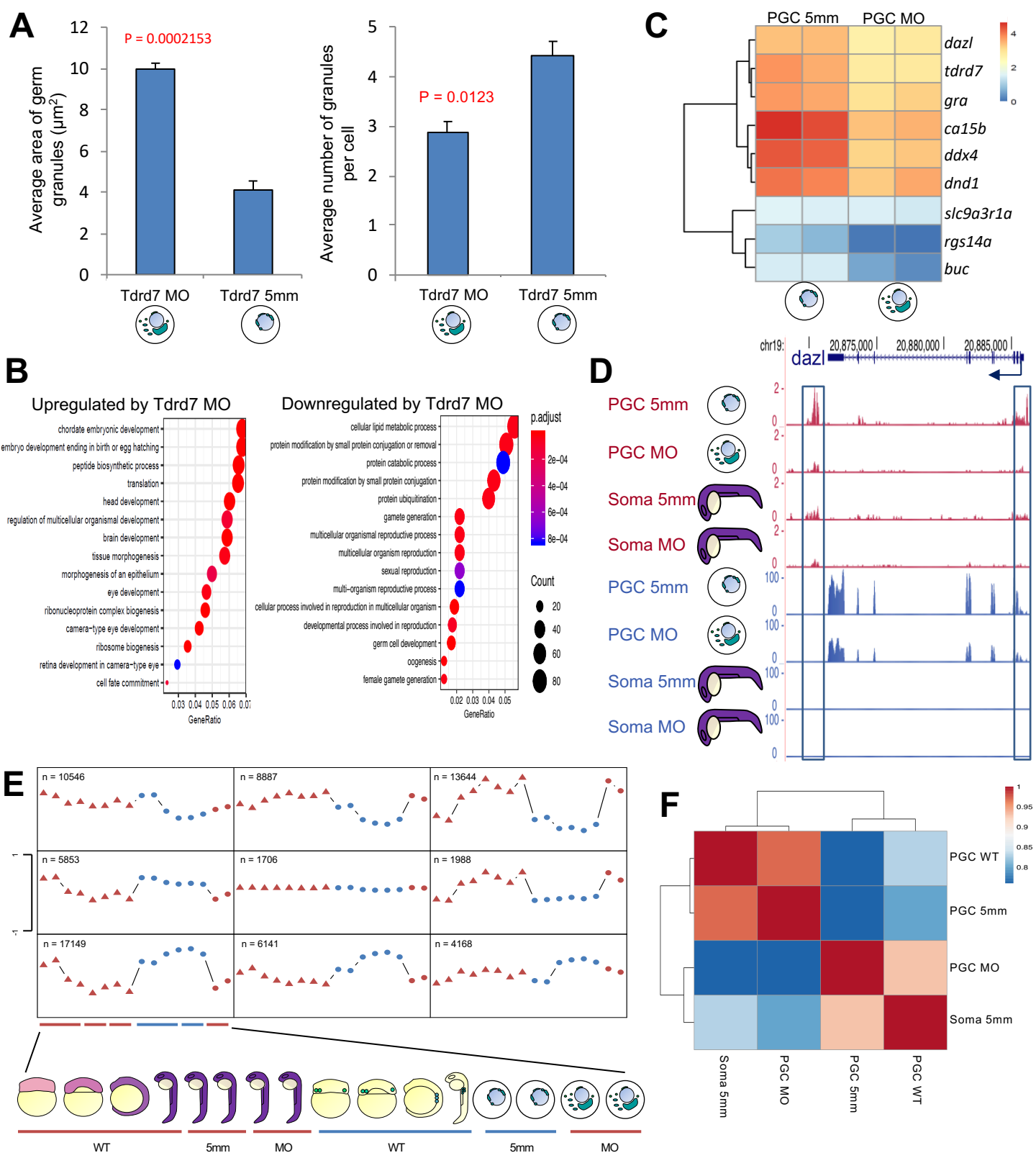
Supplementary Figure 3. Differential transcriptome between PGCs and somatic cells at early stages is not caused by differential transcription. (A) Count of upregulated genes from previous stage in PGCs and somatic cells at high and dome stages. (B) Correlation of normalised RNA-seq reads (tpm) for genes upregulated from the previous stage in the PGCs or somatic cells only at high stage. Germ plasm genes are in red. (C) Correlation between fold change increases from 256-cell to high stages in PGCs and somatic cells for gene subsets. (D) Intron retention ratio for all transcripts in PGCs (green) and somatic cells (purple) before and after ZGA. T-test was used to calculate p-values. (E) Relative gene expression fold change normalised to a reference gene in somatic cells for nuclear and cytoplasmic fractions. Error bars show standard errors. (F) Differential expressed genes between 256-cell and high stages in PGCs. Significantly differentially expressed genes are in red. P adjusted < 0.1, $\log_2$FC < -1/> 1. (G) Biological processes GO analysis for genes upregulated from high to dome in PGCs (left) and somatic cells (right).



Supplementary Figure 4. Gradual acquisition of germ identity is accompanied by epigenetic changes. (A) Differentially expressed genes in PGCs and somatic cells at prim-5 stage. Significantly differentially expressed genes are in red. P adjusted < 0.1. (B) Reproducible ATAC peaks identified in function of the IDR threshold. (C) Unsupervised correlation heatmap for developmental stages and cell types after selection of reproducible ATAC peaks. (D) Genome browser view showing ATAC-seq tracks (magenta) of *wnt3* gene in PGCs and somatic cells across early development. Blue boxes highlight putative regulatory elements. Arrows show transcriptional directionality. (E) GO terms for genes associated with open chromatin regions upregulated in late somatic cells. (F) CpG base Pearson correlation shown as heatmap and histogram of CpG frequency.



Supplementary Figure 5. PGCs do not open chromatin at regions identified as putative enhancers. (A) Percentage of differentially accessible ATAC peaks in PGCs and somatic cells at prim-5 stage overlapping gene features. Promoters include 1kb up- and downstream from the TSS. (B) Genome browser view of open chromatin regions (ATAC-seq in magenta) and transcript levels (RNA-seq in blue) in PGCs and somatic cells. Blue boxes highlight putative regulatory elements or promoters. Arrows show transcriptional directionality. (C) Percentage of ATAC peaks associated with H3K27ac and promoters in PGCs and somatic cells at prim-5 stage. (D) Correlation score between chromatin accessibility and gene transcription in function of distance from the TSS. Correlation score is represented as the absolute ATAC log_2FC value associated to genes upregulated in PGCs (blue) or somatic cells (red) (E) Genome browser view of four representative promoters upregulated in late PGCs versus somatic cells. CpG methylation levels are reported as coloured bars.



Supplementary Figure 6. **Tdrd7 translation is required for maintaining PGC specification.** (A) Phenotypic effect of Tdrd7 KD PGCs. Number and size of germ granules were measured. P-value against control is shown. Red colour indicates significant difference ($p < 0.005$) based on Wilcoxon test. (B) GO terms for genes upregulated and downregulated by Tdrd7 translational inhibition in PGCs. (C) Gene expression heatmap for germ plasm-localised transcripts. Colour intensities indicate $\log_2$(tpm+1). (D) Genome browser view of ATAC profiles after morpholino injections. Open chromatin (ATAC-seq) is shown in magenta and normalised transcript levels (RNA-seq) are shown in blue. Arrows show transcriptional directionality. (E) Self-organising map of reproducible open chromatin regions. Wild type (WT) PGCs and somatic cells are represented as circles and triangles respectively. Tdrd7 KD PGCs are shown as red circles. (F) Unsupervised correlation heatmap between wild type (WT) mismatch- and *tdrd7*-targeting morpholinos.