Dual role of Ovol2 on the germ cell lineage segregation during gastrulation in mouse embryogenesis

内藤, 優希

© 2022. Published by The Company of Biologists Ltd | Development (2022) 149, dev200319. doi:10.1242/dev.200319

RESEARCH ARTICLE

Dual role of *Ovol2* on the germ cell lineage segregation during gastrulation in mouse embryogenesis

Yuki Naitou[1], Go Nagamatsu[1], Nobuhiko Hamazaki[1], Kenjiro Shirane[1], Masafumi Hayashi[1], Makoto Hayashi[2,3], Satoru Kobayashi[2,3] and Katsuhiko Hayashi[1,4,*]

ABSTRACT

In mammals, primordial germ cells (PGCs), the origin of the germ line, are specified from the epiblast at the posterior region where gastrulation simultaneously occurs, yet the functional relationship between PGC specification and gastrulation remains unclear. Here, we show that OVOL2, a transcription factor conserved across the animal kingdom, balances these major developmental processes by repressing the epithelial-to-mesenchymal transition (EMT) that drives gastrulation and the upregulation of genes associated with PGC specification. *Ovol2a*, a splice variant encoding a repressor domain, directly regulates EMT-related genes and, consequently, induces re-acquisition of potential pluripotency during PGC specification, whereas *Ovol2b*, another splice variant missing the repressor domain, directly upregulates genes associated with PGC specification. Taken together, these results elucidate the molecular mechanism underlying allocation of the germ line among epiblast cells differentiating into somatic cells through gastrulation.

This article has an associated 'The people behind the papers' interview.

KEY WORDS: Gastrulation, *Ovol2*, Primordial germ cells

INTRODUCTION

In mice, primordial germ cells (PGCs) are specified in the posterior epiblasts at around embryonic day (E) 6.0 in response to BMP4 and WNT3 (Lawson et al., 1999; Liu et al., 1999; Ohinata et al., 2009). During this period, PGCs start to express a specific set of transcription factors that repress the somatic cell program, re-acquire latent pluripotency, trigger epigenetic reprogramming and thereby determine the PGC fate. The mechanisms underlying the PGC-specification process, including key signaling pathways, such as BMP and WNT, and transcription factors, such as BLIMP1 (PRDM1) and TFAP2C, are largely conserved among mammalian species, including, mice, rabbits, pigs and primates (Ohinata et al., 2005; Kobayashi et al., 2017, 2021; Kojima et al., 2017). Notably and without exception, PGCs form at the posterior part of the epiblast in these animals, as a set of cytokines initiates another key event in embryogenesis: gastrulation. Therefore, PGC specification essentially entails gastrulation in mammalian embryogenesis.

During gastrulation, epiblast cells undergo epithelial-to-mesenchymal transition (EMT) along the primitive streak followed by bilateral ingression underneath the epiblast layer. During this process, epiblast cells swiftly lose expression of pluripotency-associated genes and then differentiate into the primary germ layers (Nakamura et al., 2016; Hamidi et al., 2020). Considering re-acquisition of latent pluripotency with constant expression of pluripotency-associated genes in PGCs, mutually exclusive programs are executed in a group of cells in the posterior epiblast under the influence of BMP4 and WNT3. However, how these cells are properly sorted into PGCs or somatic cells remains an enigma. Recent findings have shown that both PGC and somatic cell precursors express genes involved in nascent mesoderm differentiation. For example, *T* (brachyury) is expressed in the posterior epiblast prior to gastrulation and is required for the formation of both PGCs and mesoderm (Aramaki et al., 2013, 2021). These findings indicate that PGC specification and gastrulation are initiated by an identical transcription network, and then regulated by the balance of transcription factors.

A plausible approach to the mechanism of the fate determination would be to focus on whether or not epiblast cells undergo EMT, as EMT-related genes such as *Snai1* and *Vim* are repressed in PGCs (Kurimoto et al., 2008); vice versa, *Cdh1* (E-cadherin), a key molecule maintaining the epithelium, is required for PGC differentiation (Okamura et al., 2003). Based on previous studies, genes such as *Snai1*, *Snai2*, *Zeb1*, *Zeb2*, *Twist1* and *Twist2*, have been well characterized as key transcription factors promoting EMT in various cellular contexts (Peinado et al., 2007; De Craene and Berx, 2013; Dongre and Weinberg, 2019). However, genes such as *Elf5*, *Gata3*, *Grhl2*, *Pou5f1*, *Klf4* and *Ovol2* have been identified as transcription factors that counteract EMT (Li et al., 2010; Chakrabarti et al., 2012; Cieply et al., 2012; Watanabe et al., 2014; Takaku et al., 2016; Jägle et al., 2017). Among them, *Ovol2* might have an important role in PGC specification, as evidenced by the fact that *Ovol2*-deficient embryos exhibit a reduced number of PGCs (Hayashi et al., 2017).

Ovo genes encode C2H2 zinc-finger transcription factors and are widely conserved among flies, nematodes, mice and humans (Mevel-Ninio et al., 1991; Chidambaram et al., 1997; Dai et al., 1998; Lü et al., 1998). The *Drosophila ovo* gene plays pivotal roles in the development and sex determination of germ cells (Oliver et al., 1987, 1990, 1994; Mevel-Ninio et al., 1991; Lü et al., 1998; Andrews et al., 2000). In mice, the Ovol gene family comprises *Ovol1*, *Ovol2* and *Ovol3*, and, furthermore, the *Ovol2* gene locus generates at least three splicing variants: *Ovol2a*, *Ovol2b* and *Ovol2c*. Based on the findings of the *Drosophila ovo* gene, the splicing variants have a unique role in germ cell and epidermal cell development in a highly cell context-

[1]Department of Stem Cell Biology and Medicine, Graduate School of Medical Sciences, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan. [2]Life Science Center for Survival Dynamics, Tsukuba Advanced Research Alliance (TARA), University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. [3]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. [4]Department of Germline Genetics, Graduate School of Medicine, Osaka University, Suita 565-0871, Osaka, Japan.

*Author for correspondence (hayashi.katsuhiko.104@m.kyushu-u.ac.jp)

K.H., 0000-0002-2479-2941

Handling Editor: Haruhiko Koseki
Received 3 November 2021; Accepted 27 December 2021

Development (2022) 149, dev200319. doi:10.1242/dev.200319

dependent manner (Payre et al., 1999; Andrews et al., 2000). These findings imply that mouse Ovol genes would have a unique role in PGC specification. Thus, in this study, we investigate the functional involvement of mouse Ovol genes in the lineage segregation between PGCs and somatic cells during gastrulation.

RESULTS

Ovol2 is involved in an initial step of PGC specification

To clarify mechanistic insights into *Ovol2*, we first investigated the number of cells expressing nascent PGC markers, such as *Blimp1* (also known as *Prdm1*) and *Tfap2c*, in *Ovol2* mutant embryos (*Ovol2*$^{-/-}$) during gastrulation. We crossed *Ovol2*$^{+/-}$ males harboring the *Blimp1-mVenus* (BV) reporter gene (Ohinata et al., 2008) with *Ovol2*$^{+/-}$ females. Immunofluorescence analysis showed that the differentiation of BV-positive PGCs was severely disturbed in *Ovol2*$^{-/-}$ embryos at E7.5 (Fig. 1A). Compared with wild-type embryos holding a cluster of PGCs expressing both BV and TFAP2C, *Ovol2*$^{-/-}$ embryos showed sparse formation of PGCs with weak levels of BV and TFAP2C expression (Fig. 1B). The number of BV-positive cells was reduced in *Ovol2*$^{-/-}$ embryos around gastrulation (Fig. 1C), a far earlier stage than the somite stage reported previously (Hayashi et al., 2017). These results demonstrated that PGC specification is hampered in the *Ovol2*$^{-/-}$ embryos at the gastrulation stage.

As the number of nascent PGCs is reduced in *Ovol2*$^{-/-}$ gastrulating embryos, we employed an *in vitro* differentiation system, in which PGC specification is faithfully reproduced in a series of differentiation stages from embryonic stem cells (ES cells) to PGC-like cells (PGCLCs) via epiblast-like cells (EpiLCs) (Hayashi et al., 2011). Taking advantage of this culture system, we assessed the individual involvement of splicing variants of *Ovol2* (*Ovol2a*, *Ovol2b* and *Ovol2c*) and other Ovol family genes (*Ovol1* and *Ovol3*) in PGCLC differentiation. Quantitative PCR (Q-PCR) analysis revealed that *Ovol1* expression was continuously increased from ES cells to PGCLCs, and *Ovol2a* and *Ovol2b* expression reached a peak in EpiLCs and PGCLCs at 3 days of culture, respectively; *Ovol2c* expression was barely detectable and *Ovol3* expression was constant throughout the differentiation process (Fig. 1D). We interrogated expression patterns of Ovol family genes in epiblast cells from E4.5 to E6.5 in the single-cell RNA sequence (RNA-seq) dataset provided by SC3-seq (Fig. S1A) (Nakamura et al., 2015, 2016). Considering that ES cells and EpiLCs correspond to E4.5 and E5.5 epiblast cells, respectively, and E6.5 epiblast contains early PGCs, *Ovol1* and *Ovol2* expression was largely consistent between *in vitro* and their *in vivo* counterparts. However, *Ovol3* expression was not detectable in the dataset, possibly owing to the 3′UTR of *Ovol3* being indistinguishable from that of *Polr2i* (91 bp are completely matched).

Next, we induced PGCLCs from individual knockout (KO) (*Ovol1*$^{-/-}$, *Ovol2*$^{-/-}$ or *Ovol3*$^{-/-}$) and triple KO (TKO) ES cells harboring the BV reporter (Fig. S1B). At day 2 of PGCLC induction, BV expression in *Ovol2*$^{-/-}$ and TKO aggregates was clearly weaker than that in wild type, whereas BV expression in *Ovol1*$^{-/-}$ and *Ovol3*$^{-/-}$ appeared to be comparable with that in wild type (Fig. 1E). Quantification of the fluorescence intensity showed that the levels of the BV intensities in *Ovol2*$^{-/-}$ and TKO aggregates were significantly reduced (Fig. 1F). Q-PCR analysis of *Ovol1*$^{-/-}$, *Ovol2*$^{-/-}$, *Ovol3*$^{-/-}$ and TKO showed that endogenous *Blimp1* expression was downregulated not only in *Ovol2*$^{-/-}$ and TKO but also in *Ovol3*$^{-/-}$ (Fig. 1G). Of note, we observed a distinctive expression pattern for *Tfap2c*, a functional marker of PGC specification (Weber et al., 2010), between *Ovol2*$^{-/-}$ or TKO and *Ovol3*$^{-/-}$. To further characterize BV-positive cells in the KO aggregates, we examined expression of BV and TFAP2C by immunofluorescence analysis. The number of cells expressing both BV and TFAP2C at high levels was significantly decreased in *Ovol2*$^{-/-}$, *Ovol3*$^{-/-}$ and TKO (Fig. 1H,I). In contrast, *Ovol3*$^{-/-}$ aggregates contained larger percentages of TFAP2C-high- and BV-low-expressing cells (Fig. S1C), which might be responsible for the abundance of transcripts of *Tfap2c* in *Ovol3*$^{-/-}$ detected by Q-PCR (Fig. 1G). This unique phenotype is not consistent with the PGC differentiation process *in vivo*, in which the expression of BV and TFAP2C are correlated (Fig. 1B), indicating a distinctive role for *Ovol3* in PGC specification. At day 4 of PGCLC induction, the number of BV-positive cells in *Ovol2*$^{-/-}$ and TKO aggregates decreased and the number in *Ovol3*$^{-/-}$ was slightly restored (Fig. S1D,E). The number of BV and TFAP2C double-positive cells decreased in *Ovol2*$^{-/-}$ and TKO aggregates (Fig. S1F,G). Importantly, the extent of the decrease in the number of BV and TFAP2C double-positive cells in *Ovol2*$^{-/-}$ aggregates was similar to that in *Ovol2*$^{-/-}$ embryos (Fig. 1B, Fig. S1G) (Hayashi et al., 2017). These results demonstrate that *Ovol2* has a dominant role in PGC specification among the Ovol family genes.

Identification of molecular pathway downstream of *Ovol2*

To identify the molecular pathway downstream of each Ovol gene, we further investigated the transcriptomes of *Ovol1*$^{-/-}$, *Ovol2*$^{-/-}$, *Ovol3*$^{-/-}$ and TKO. Principal component analysis (PCA) showed that the effect of the gene disruption was subtle in ES cells and EpiLCs, but it appeared clearly after PGCLC induction (Fig. 2A). Furthermore, PCA using only PGCLCs showed that these transcriptomes could be divided into three groups: wild type and *Ovol1*$^{-/-}$, *Ovol2*$^{-/-}$ and TKO, and *Ovol3*$^{-/-}$ (Fig. 2B). Analysis of genes reflecting on each group showed that genes expressed in nascent mesoderm but downregulated in PGCs, such as *Hand1*, *Vim* and *Bmp4* (Kurimoto et al., 2008), were enriched in the direction to the group of *Ovol2*$^{-/-}$ and TKO (Fig. 2C). In contrast, genes involved in both PGC and nascent mesoderm differentiation, such as *T*, *Eomes* and *Wnt3* (Aramaki et al., 2013; Senft et al., 2019), were enriched in the direction of wild type and *Ovol1*$^{-/-}$. In addition, unexpectedly, pluripotency-associated genes, such as *Nanog*, *Dppa3* and *Zfp42*, were enriched in the direction of *Ovol3*$^{-/-}$. It is noteworthy that EMT-related genes, such as *Zeb1*, *Snai1* and *Snai2*, were enriched in the group of *Ovol2*$^{-/-}$ and TKO, whereas genes representing the maintenance of cellular polarity, such as *Cdh1*, *Cldn6* and *Cldn7*, were clearly enriched in the opposite direction. These expression profiles suggested that *Ovol2*$^{-/-}$ and TKO accelerate mesoderm differentiation through the progression of EMT.

Consistent with the PCA results, differentially expressed genes (DEGs) compared with wild type were largely overlapped between *Ovol2*$^{-/-}$ and TKO (Fig. 2D), whereas they were scarcely overlapped between TKO and either *Ovol1*$^{-/-}$ or *Ovol3*$^{-/-}$ (Fig. S2A). Gene ontology (GO) analysis of 198 commonly upregulated DEGs in both *Ovol2*$^{-/-}$ and TKO exhibited terms such as ‘cell differentiation’ and ‘positive regulation of EMT’. On the other hand, 173 commonly downregulated DEGs belonged to ‘stem cell population maintenance’ and ‘cell adhesion’. Considering the involvement of OVOL2 in the inhibition of EMT (Lee et al., 2014; Watanabe et al., 2014; Wu et al., 2017), these results indicate that the defective PGCLC differentiation in *Ovol2*$^{-/-}$ and TKO was attributable to aberrant regulation of EMT and/or maintenance of pluripotency. This idea was supported by RNA-seq analyses showing downregulation of pluripotency-associated genes and epithelial adhesion genes, and upregulation of

Development (2022) 149, dev200319. doi:10.1242/dev.200319

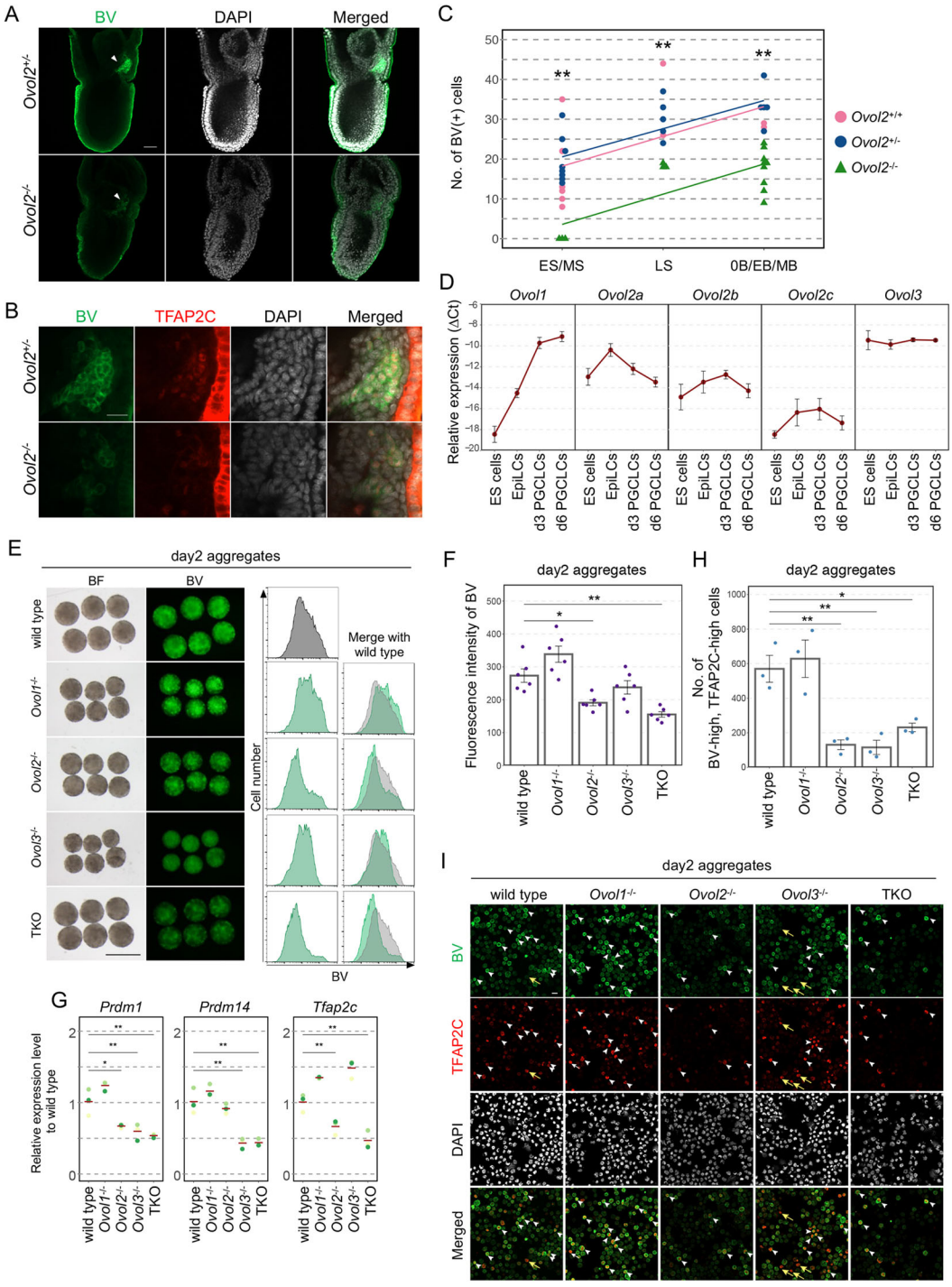

Fig. 1. Mouse Ovol genes contribute to PGC specification. (A) PGC specification in *Ovol2*$^{+/-}$ and *Ovol2*$^{-/-}$ embryos at the EB/MB stage. BV (green), DAPI (white) and merged images are shown. Arrowheads indicate BV-positive PGCs. Scale bar: 50 µm. (B) A cluster of PGCs in the *Ovol2*$^{+/-}$ and *Ovol2*$^{-/-}$ embryos. BV (green), TFAP2C (red), DAPI (white) and merged images are shown. Scale bar: 20 µm. (C) Number of BV-positive cells in embryos. The lines of best fit are shown. There were fewer BV-positive cells in *Ovol2*$^{-/-}$ than in wild-type and *Ovol2*$^{+/-}$ embryos. ***P*<0.01 (unpaired, two-tailed Student's *t*-test). ES, early streak stage; MS, midstreak stage; LS, late streak stage; 0B, no bud stage; EB, early bud stage; MB, midbud stage. (D) Expression of Ovol genes during PGCLC induction. Mean ΔCt values±s.d. determined from three independent Q-PCR analyses are shown. (E) PGCLC induction from Ovol gene KO cells. Bright field (BF) and BV images (left), and FACS analysis of BV expression (right) in aggregates at day 2 of PGCLC induction. Scale bar: 500 µm. (F) Quantification of fluorescent intensities of BV in Ovol gene-KO and wild-type aggregates at day 2. Data are mean±s.d. **P*<0.05; ***P*<0.01 (Tukey's test). (G) Q-PCR analysis of PGC-related genes in day 2 aggregates. For each gene, values relative to wild type are shown. The expression levels were determined by experiments using triple biological samples. The lines indicate the mean values. **P*<0.05; ***P*<0.01 (Tukey's test). (H) Quantification of PGCLC induction in *Ovol1*$^{-/-}$, *Ovol2*$^{-/-}$, *Ovol3*$^{-/-}$ and TKO. The numbers of cells exhibiting high expression of both BV and TFAP2C at day 2 of PGCLC induction are shown. Data are mean±s.d. from three independent experiments. **P*<0.05; ***P*<0.01 (Tukey's test). (I) The immunofluorescence analysis of BV and TFAP2C. BV (green), TFAP2C (red) and DAPI (white) in cells at day 2 of PGCLC induction are shown. Arrowheads and yellow allows indicate representative cells highly expressing both BV and TFAP2C, and cells highly expressing TFAP2C but not BV, respectively. BV- and TFAP2C-positive cells are sparce in *Ovol2*$^{-/-}$, *Ovol3*$^{-/-}$ and TKO. Scale bar: 20 µm.

Development (2022) 149, dev200319. doi:10.1242/dev.200319

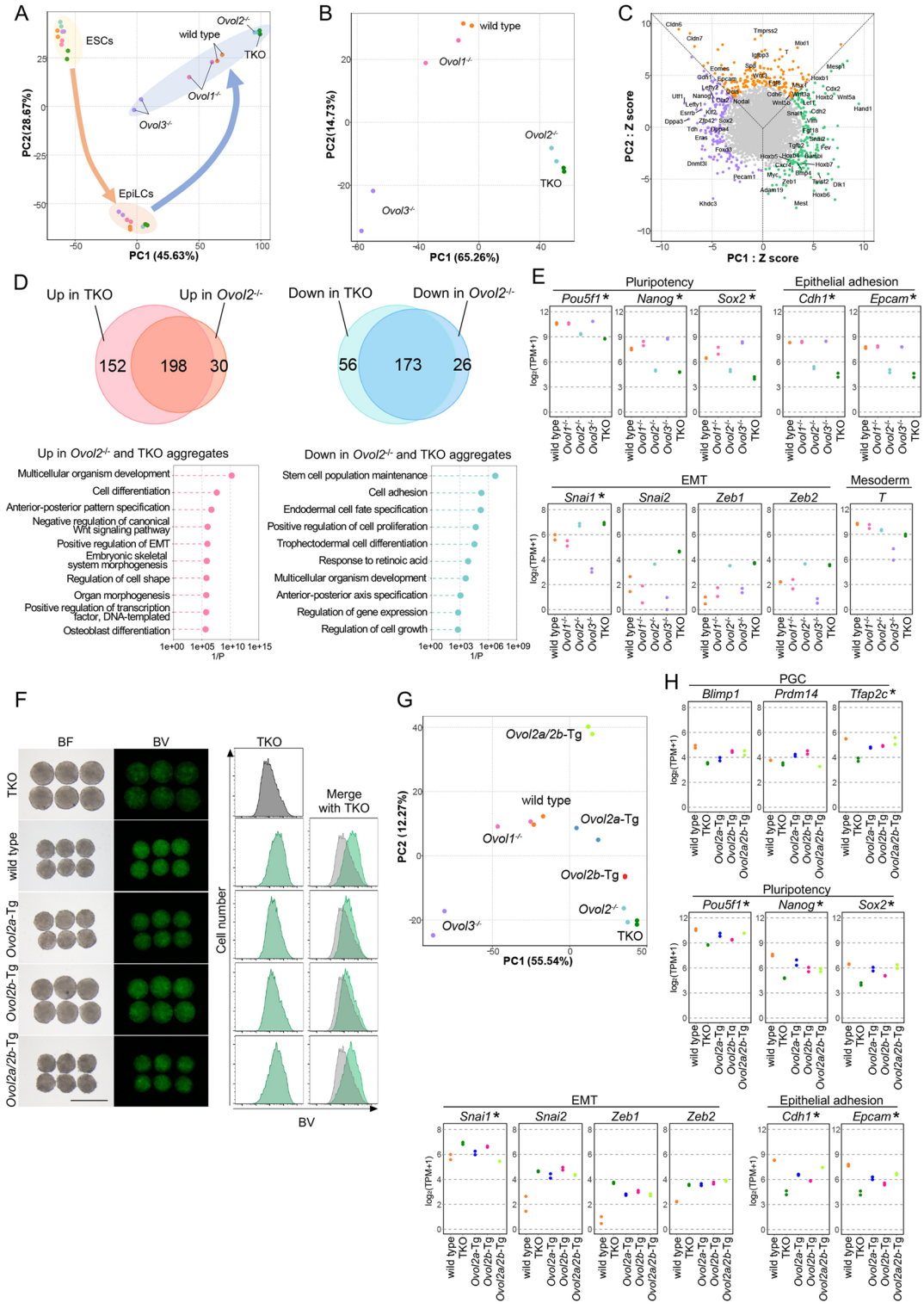

Fig. 2. *Ovol2* plays a dominant role in PGC specification through repression of EMT-related genes. (A) Principal component analysis (PCA) of the transcriptomes of ES cells, EpiLCs and day 2 aggregates of Ovol gene-KO and wild type. (B) PCA of transcriptomes of Ovol gene-KO and wild-type day 2 aggregates. (C) A scatterplot of the Z-normalized loading scores of the genes for PC1 and PC2. Genes with a radius greater than four s.d. values (476 genes) are shown. Key genes are annotated. (D) Venn diagrams representing the overlaps of DEGs. Comparisons of DEGs upregulated (left) or downregulated (right) in *Ovol2*$^{-/-}$ and TKO day 2 aggregates, compared with wild type, are shown. GO analyses of genes commonly upregulated (left) or downregulated (right) are shown below. (E) Expression levels of genes involved in pluripotency, epithelial adhesion, EMT and mesoderm in day 2 aggregates. The $\log_2$(TPM+1) values determined by two independent RNA-seq analyses are shown. Asterisks indicate DEGs in TKO compared with wild type (>2-fold difference). (F) PGCLC induction from *Ovol2*-Tg ES cell lines. Images (left) and FACS analysis (right) in wild-type, parental TKO, *Ovol2a*-, *Ovol2b*- and *Ovol2a*/*2b*-Tg aggregates at day 2 are shown. Scale bar: 500 µm. (G) PCA of transcriptomes of the *Ovol2*-Tg aggregates at day 2. (H) Expression levels of genes involved in PGC specification, pluripotency, EMT and epithelial adhesion in the *Ovol2*-Tg aggregates at day 2. The $\log_2$(TPM+1) values determined by two independent RNA-seq analyses are shown. Asterisks indicate DEGs in *Ovol2a*/*2b*-Tg compared with TKO (>2-fold difference).

Development (2022) 149, dev200319. doi:10.1242/dev.200319

EMT-related genes (Fig. 2E, Fig. S2B). The downregulation of *Cdh1* in *Ovol2*$^{-/-}$ and TKO was confirmed by immunofluorescence analysis (Fig. S2C). These results indicated that a key molecular pathway involving *Ovol2* during PGC specification is maintenance of cell adhesion, which is lost during EMT entailing mesoderm induction.

Next, we further characterized the function of *Ovol2* variants in PGC specification by enforced expression of *Ovol2a* and/or *Ovol2b* in TKO (Fig. S2D). With enforced expression of the variants, BV signals were restored in all *Ovol2a*-transgenic (Tg), *Ovol2b*-Tg and *Ovol2a/2b*-Tg aggregates, compared with the parental TKO aggregates (Fig. 2F). FACS analyses showed that the level of BV expression in *Ovol2a/2b*-Tg aggregates was higher than that in *Ovol2a*- or *Ovol2b*-Tg aggregates, suggesting a synergistic effect of these variants. RNA-seq analysis followed by PCA revealed that aggregates of *Ovol2a*-, *Ovol2b*- and *Ovol2a/2b*-Tg at day 2 of culture became closer, albeit not identical, to wild type (Fig. 2G). Of note, the similarity of gene expression profiles of *Ovol2a*-Tg aggregates and wild type was greater than that between the profiles of *Ovol2b*-Tg aggregates and wild type. In *Ovol2a*-, *Ovol2b*- and *Ovol2a/2b*-Tg aggregates at day 2 of culture, the expression of *Tfap2c* was significantly restored and so were pluripotency-associated genes, such as *Pou5f1*, *Nanog* and *Sox2* (Fig. 2H). Expression of some EMT-related genes, such as *Snai1* and *Zeb1*, was partially repressed in these transgenic aggregates, and the extent of this repression was slightly larger in *Ovol2a*-Tg than in *Ovol2b*-Tg. Consistent with the repression of the EMT-related gene expression, expression of the epithelial adhesion genes *Cdh1* and *Epcam* was partially but significantly restored in these transgenic aggregates (Fig. 2H). These results indicate that *Ovol2a* and *Ovol2b* promote PGC specification by preventing EMT and promoting pluripotency.

Incomplete compensation for *Ovol2* disruption by *Cdh1*

CDH1-mediated cell-cell interaction is essential for PGC formation at around E7 (Okamura et al., 2003). Consistently, the level of CDH1 was correlated with BV protein expression in aggregates at day 2 of PGCLC induction (Fig. S2C). Thus, we examined whether a scarcity of CDH1 in *Ovol2*$^{-/-}$ and TKO is the main cause of the defective PGCLC differentiation by enforced expression of *Cdh1* in TKO ES cells (*Cdh1*-Tg) followed by PGCLC induction (Fig. S3A). In *Cdh1*-Tg aggregates at day 2 of PGCLC induction, BV expression was restored to a level comparable with that in the wild-type aggregates (Fig. 3A). However, despite BV expression, *Cdh1*-Tg aggregates had a closely similar transcriptome to TKO aggregates: the difference in number of DEGs between *Cdh1*-Tg and TKO (134 genes) was much smaller than that between *Cdh1*-Tg and wild type (587 genes) (Fig. 3B). PGC-related genes, such as *Blimp1* and *Tfap2c*, were restored in *Cdh1*-Tg, whereas pluripotency-associated genes and EMT-related genes were not nominated as DEGs ($\log_2$ fold change >1, FDR <0.001) (Fig. 3C). Indeed, compared with the PGC-related genes, the changes in expression of pluripotency-associated genes (*Pou5f1*, *Nanog* and *Sox2*) and EMT-related genes (*Snai1*, *Snai2* and *Zeb1*) were subtle (Fig. 3D). PCA confirmed that genes whose expression was altered in *Cdh1*-Tg were *Cdh1* itself and genes that are irrelevant to PGC specification (Fig. 3E, Fig. S3B). These results suggest that enforced expression of *Cdh1* partially restores PGCLC induction in TKO through upregulation of the PGC-related genes. On the other hand, transcriptional regulation of pluripotency-associated genes and EMT-related genes was independent of *Cdh1* expression, therefore indicating that the *Ovol2*-mediated gene regulatory network is required for PGC(LC) induction in a *Cdh1*-independent manner.

Identification of direct targets of OVOL2A and OVOL2B

To explore the genome-wide targets for OVOL2A and OVOL2B, we performed chromatin immunoprecipitation sequencing (ChIP-seq). For this purpose, we used *Ovol2a*-Tg and *Ovol2b*-Tg aggregates at day 2 of PGCLC induction and antibodies against FLAG-tag fused to exogenous OVOL2A and OVOL2B. ChIP-seq analyses using biologically duplicated samples detected 1215 and 5157 peaks as candidates for OVOL2A- and OVOL2B-binding sites, respectively (Fig. 4A). Nearly all of the peaks of OVOL2A-binding sites were overlapped with those for OVOL2B, except in the case of five genes specific to OVOL2A (Fig. 4A,B). *De novo* motif-finding analysis identified a binding consensus sequence (CCGYTA) of both OVOL2A and OVOL2B (Fig. 4C), which is consistent with the fact that these variants harbor the identical zinc-finger domain. These consensus sequences were almost identical to a known OVOL1/2 binding sequence (CCGTTA) (Nair et al., 2007; Watanabe et al., 2014).

We tested whether these binding peaks are associated with gene expression dynamics. By referring to the published datasets (Kurimoto et al., 2015), we analyzed histone marks, including trimethylated histone H3 lysine 4 (H3K4me3), trimethylated histone H3 lysine 27 (H3K27me3) and acetylated histone H3 lysine 27 (H3K27ac) around these peaks for OVOL2A and OVOL2B in day 2 aggregates. Genomic regions highly mapped as OVOL2A- and OVOL2B-binding sites were correlated with enrichment of H3K4me3, and H3K27ac was slightly enriched in these regions (Fig. 4D). However, H3K27me3 was enriched in genomic regions flanking OVOL2A- and OVOL2B-binding sites that were moderately mapped. Next, we assigned genes harboring a OVOL2A- or OVOL2B-binding peak within a 50 kb region flanking the longest transcripts detected, and then interrogated their expression dynamics upon enforced expression of *Ovol2a* or *Ovol2b*. Unsupervised hierarchical clustering (UHC) of transcriptionally altered OVOL2A- or OVOL2B-binding genes revealed that these genes were classified into four (cluster 1A-4A) or five (cluster 1B-5B) clusters in the comparison between wild type and either *Ovol2a*-Tg or *Ovol2b*-Tg, respectively (Fig. 4E). Genes consistently up- or downregulated in both the wild type and *Ovol2a*-Tg were enriched in cluster 2A or cluster 4A, respectively; cluster 2A included WNT-related genes, such as *Wnt3* and *Sp5*, and cluster 4A included EMT-related genes, such as *Zeb1* and *Zeb2* (Fig. 4E). Genes consistently up- or downregulated in both wild type and *Ovol2b*-Tg were enriched in cluster 1B or cluster 4B, respectively. Of note, cluster 1B included exclusively PGC-related and pluripotency-associated genes, such as *Blimp1*, *T* and *Nanog*. The *Zeb1* locus showed a peak of OVOL2A and OVOL2B with bivalent histone modification at the transcription start site (TSS), and *Blimp1*, *Nanog* and *T* loci showed peaks of OVOL2B with H3K27ac at the region upstream of each TSS (Fig. 4F, Fig. S4A). Among genes harboring a OVOL2A- or OVOL2B-binding peak within the 0.5 kb region flanking the TSS, 236 OVOL2A-binding genes and 656 OVOL2B-binding genes were associated with both H3K4me3 and H3K27me3, so-called bivalent histone modification (Fig. S4B). Of the 236 and 656 genes, 119 and 303 genes, respectively, overlapped with the genes with the bivalent histone modification in PGCLCs at day 2 of induction (Fig. S4B) (Kurimoto et al., 2015), indicating that OVOL2A and/or OVOL2B play a role in the establishment of the PGC-specific epigenetic landscape.

Development (2022) 149, dev200319. doi:10.1242/dev.200319

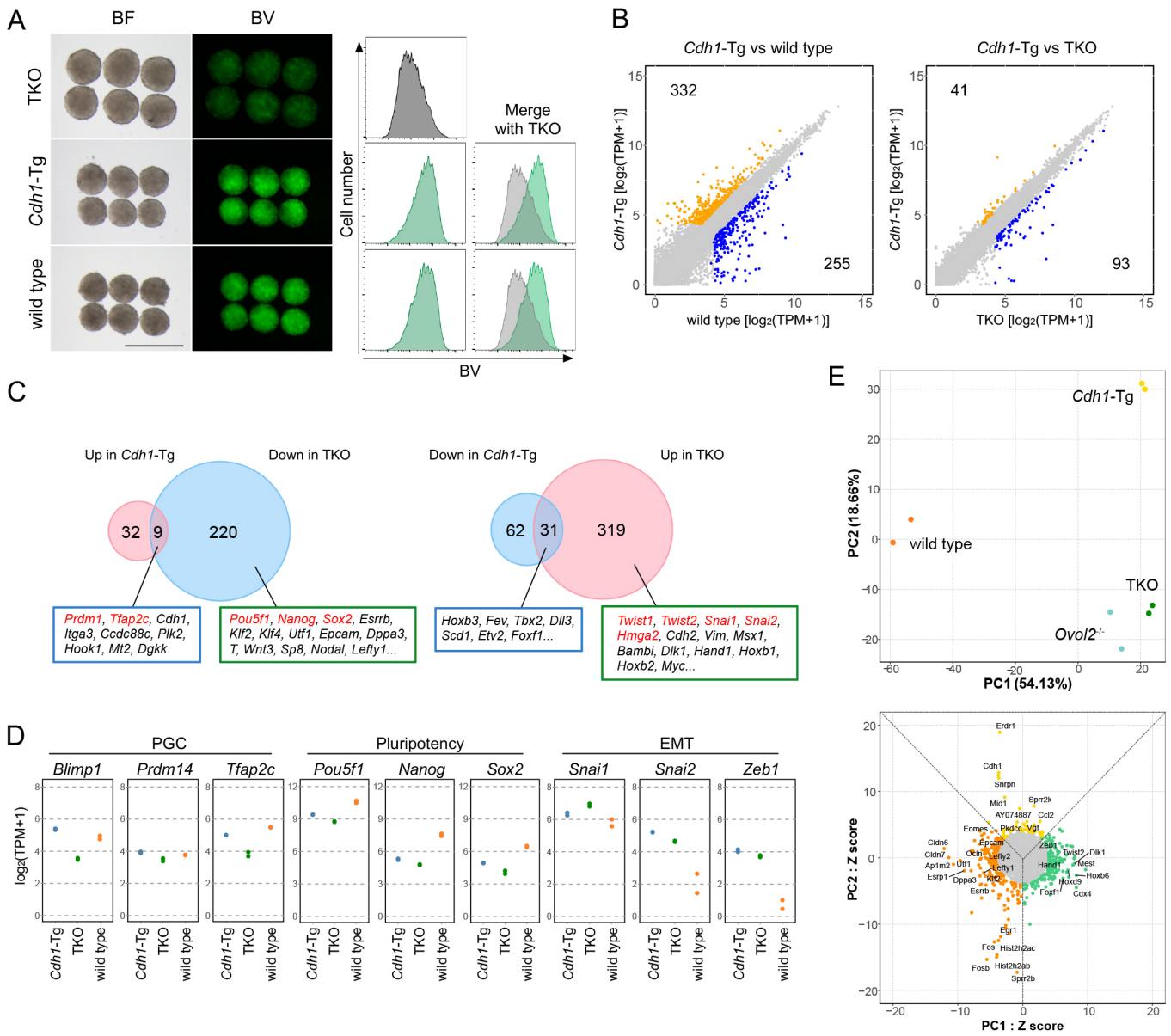

Fig. 3. *Cdh1* expression can restore PGC-related genes but not pluripotency-associated genes. (A) PGCLC induction from TKO with enforced expression of *Cdh1*. Bright-field (BF) and BV images (left), and FACS analysis of BV expression (right) in parental TKO, *Cdh1*-Tg and wild-type day 2 aggregates are shown. Scale bar: 500 µm. (B) Scatterplot comparison of gene expression profiles of TKO, *Cdh1*-Tg and wild-type day 2 aggregates. DEGs (>2-fold difference) are shown with orange or blue dots. The numbers of DEGs are indicated. (C) Venn diagrams representing the overlap of DEGs. The overlap between genes upregulated in *Cdh1*-Tg and downregulated in TKO (left), and downregulated in *Cdh1*-Tg and upregulated in TKO (right) day 2 aggregates are shown. Key genes are annotated; genes particularly relevant to PGC specification or EMT are shown in red. (D) Expression levels of genes involved in PGC specification, pluripotency, EMT and epithelial adhesion in the *Chd1*-Tg aggregates at day 2. The $\log_2$(TPM+1) values determined by two independent RNA-seq analyses are shown.
(E) Transcriptomic analysis of TKO, *Cdh1*-Tg and wild-type aggregates at day 2. PCA (top) and a scatterplot of the Z-normalized loading scores of the genes for the PC1 and PC2 (bottom) are shown. In the Z-score scatter plot, genes with a radius greater than four s.d. values (375 genes) are shown. Key genes are annotated.

To confirm the functional involvement of OVOL2A and OVOL2B in transactivation of these genes, we performed luciferase reporter analyses using genomic fragments upstream of each gene. Transcriptional activity of the luciferase gene with the ∼1000 bp upstream of *Zeb1* was clearly repressed by expression of OVOL2A in HEK293T cells, whereas it was not repressed by expression of OVOL2B (Fig. 4G, Fig. S4C). These results demonstrate that OVOL2A but not OVOL2B directly represses *Zeb1*. Next, we examined the transcriptional effects of genomic fragments upstream of *Blimp1*, *T* and *Nanog*. To stringently assess the transcriptional activity in an appropriate cell context, each luciferase reporter construct was integrated into the genome of *Ovol2b*-Tg (Fig. S4C). Using these *Ovol2b*-Tg lines that harbor a similar copy number of the reporter construct (Fig. S4D), luciferase activities were determined at day 1 of PGCLC induction with or without doxycycline (Dox), which transactivates the exogenous *Ovol2b* gene. In *Ovol2b*-Tg with *Blimp1-luc*, luciferase activity was elevated in a presence of Dox, whereas other genomic constructs did not enhance transactivation (Fig. 4H). The upregulation of *Blimp1-luc* was OVOL2 dependent, as deletion of the OVOL2B-binding

Development (2022) 149, dev200319. doi:10.1242/dev.200319

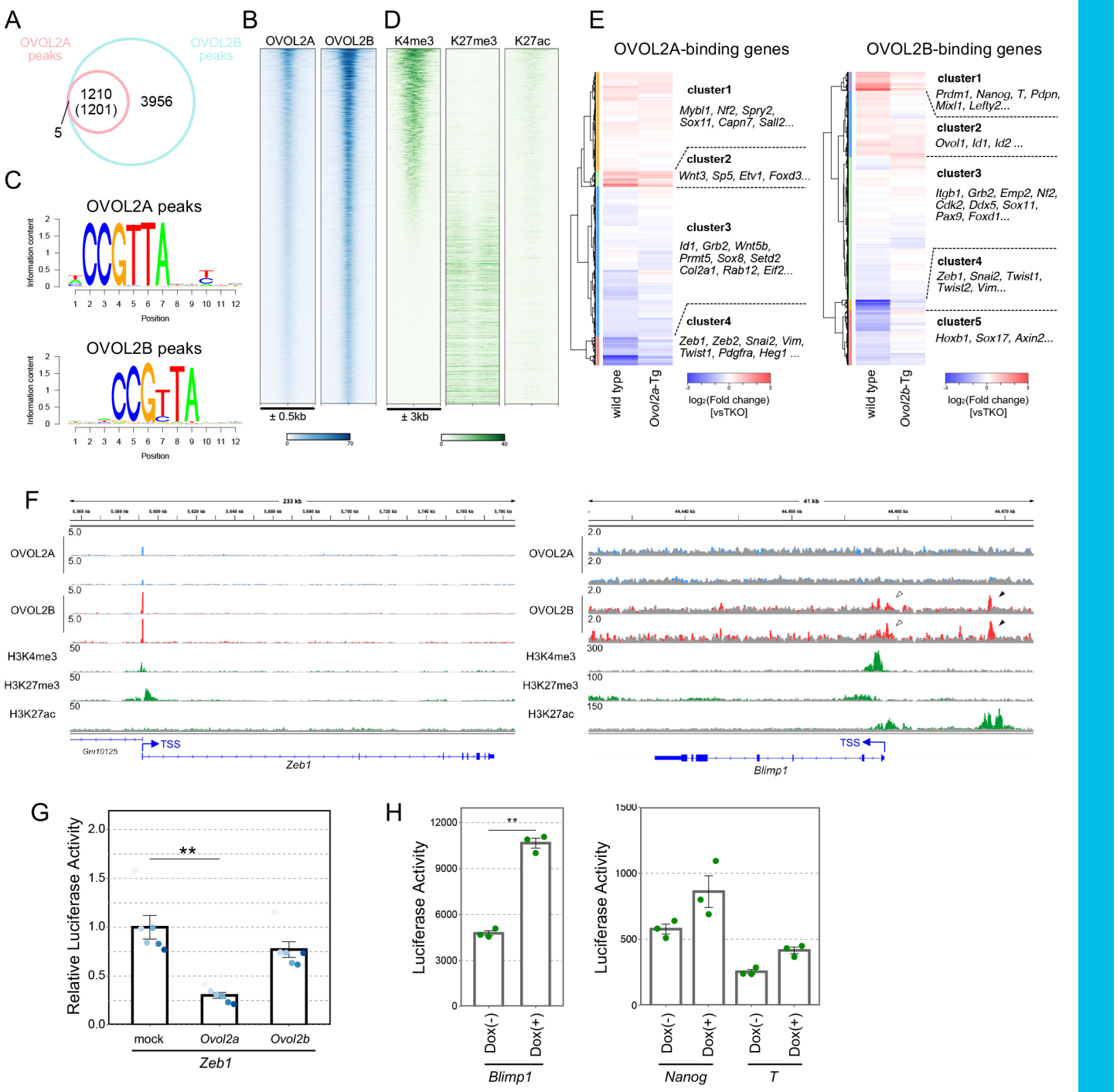

Fig. 4. *Ovol2a* and *Ovol2b* directly regulate *Zeb1* and *Blimp1*. (A) Venn diagram of binding peaks of OVOL2A and OVOL2B. The result shown is based on two independent experiments. (B) Heat map representation of OVOL2A and OVOL2B peaks around the 0.5 kb genomic region flanking the binding peaks. (C) Motifs enriched in the OVOL2A- or OVOL2B-binding regions. (D) Heat map representation of the histone modifications around the OVOL2A- or OVOL2B-binding regions. Each histone modification level around the 3.0 kb genomic region flanking the OVOL2A- or OVOL2B-binding peak is shown. (E) Identification of OVOL2A or OVOL2B target genes associated with gene expression. OVOL2A or OVOL2B target genes whose expression was significantly changed in *Ovol2a*- (left) or *Ovol2b*- (right) Tg, respectively, and wild-type day 2 aggregates, compared with TKO, are sorted by unsupervised hierarchical clustering (UHC). Each cluster is defined by the UHC dendrogram. Representative genes in the clusters are shown. (F) OVOL2A or OVOL2B binding in *Blimp1* and *Zeb1* loci. ChIP-seq tracks of the OVOL2A- (blue) or OVOL2B- (red) binding region; H3K4me3, H3K27me3 and H3K27ac around *Zeb1* and *Blimp1* are shown as counts per million reads (CPM). Gray in the tracks shows the CPM from input. Black and white arrowheads in the OVOL2B tracks indicate the peaks around enhancer and TSS of *Blimp1* locus, respectively. (G) Luciferase analysis of OVOL2A- or OVOL2B-mediated repression of *Zeb1* expression. Transcriptional activity of the luciferase reporter bearing the OVOL2-binding element (OBE) of *Zeb1*, shown in F, upon *Ovol2a* or *Ovol2b* expression. ***P*<0.01 (Mann–Whitney *U*-test). Data are mean±s.d. from six independent experiments. (H) Luciferase analysis of OVOL2A- or OVOL2B-mediated activation of PGC-related genes expression. Transcriptional activities of the luciferase reporter bearing OBEs around *Blimp1*, *Nanog* or *T* in day 2 aggregates, upon *Ovol2b* expression, are shown. The genomic fragment of *Blimp1* used for this analysis is shown in F, and those of *Nanog* and *T* for the reporter construct are shown in Fig. S4A. Data are mean±s.d. from three independent experiments. ***P*<0.01 (paired, two-tailed Student's *t*-test).

Development (2022) 149, dev200319. doi:10.1242/dev.200319

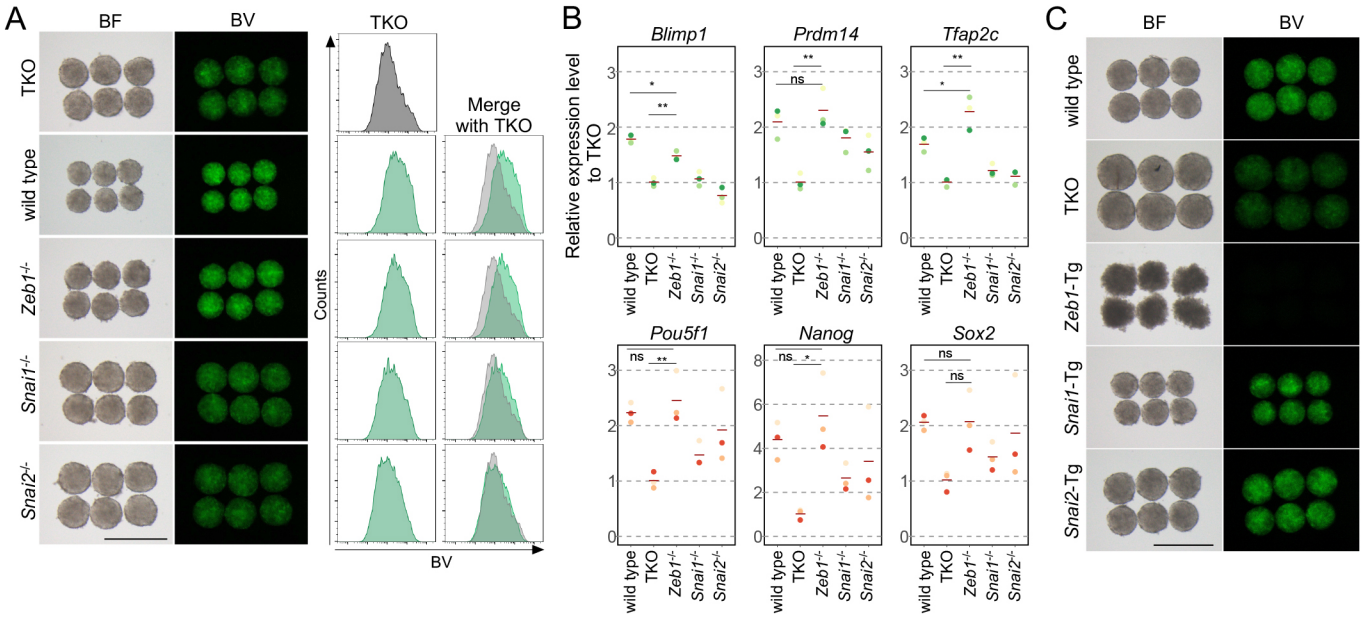

Fig. 5. The repression of EMT promotes PGC specification. (A) PGCLC induction from EMT-related gene-disrupted TKO ES cells. Bright-field (BF) and BV images (left), and FACS analysis of BV expression (right) in day 2 aggregates of the genotype indicated are shown. Scale bar: 500 µm. (B) Q-PCR analysis of PGC-related genes and pluripotency-associated genes in day 2 aggregates. For each gene, relative values in the genotype indicated, compared with TKO, are shown. The expression levels were determined by experiments using triple biological samples. The horizontal lines indicate the mean values. **P*<0.05; ***P*<0.01 (Tukey's test). (C) PGCLC induction from ES cells with enforced expression of EMT-related genes. BF and BV images in day 2 aggregates from ES cells bearing the transgenic construct indicated. Scale bar: 500 µm.

element (OBE) nullified the effect of transactivation (Fig. S4E). In contrast to the similar and higher levels of endogenous expression of *Nanog* and *T*, respectively, compared with that of *Blimp1*, in TKO (Figs 1G and 2E), the basal levels of luciferase activity of *T-luc* and *Nanog-luc* without Dox were much lower than that of *Blimp1-luc*. This could be due to the genomic region in the construct being insufficient for the transactivation, raising the possibility that these genes are regulated by OVOL2B. Nevertheless, this analysis demonstrates that transcription of *Blimp1* is promoted by OVOL2B through direct binding to its enhancer region.

Destination of PGC specification by repression of *Zeb1*

As *Zeb1* was identified as a direct target of OVOL2A during PGC specification, we investigated the functional consequence of repression of *Zeb1*. For this purpose, we deleted *Zeb1* in TKO ES cells and then induced PGCLCs (Fig. S5A). Surprisingly, BV expression was almost restored in *Zeb1*$^{-/-}$ aggregates (Fig. 5A). In sharp contrast, deletion of other EMT-related genes, such as *Snai1* and *Snai2*, did not restore BV expression in TKO (Fig. 5A, Fig. S5A), demonstrating that *Zeb1* plays a unique role in counteracting PGC specification. Q-PCR analysis revealed that not only BV expression but also the expression of PGC-related genes (*Blimp1*, *Prdm14* and *Tfap2c*) and pluripotency-associated genes (*Pou5f1*, *Nanog* and *Sox2*) were restored in *Zeb1*$^{-/-}$ cells to levels comparable with those in wild type (Fig. 5B). In addition, a substantial restoration of *Cdh1* expression was observed in *Zeb1*$^{-/-}$ but not in *Snai1*$^{-/-}$ and *Snai2*$^{-/-}$ cells (Fig. S5B). Compared with the PGC-related genes, expression of pluripotency-associated genes was partially restored in *Snai1*$^{-/-}$ and *Snai2*$^{-/-}$ cells, suggesting that suppression of EMT promotes the maintenance of pluripotency, as previously reported (Li et al., 2010; Samavarchi-Tehrani et al., 2010). The unique potential of *Zeb1* in PGC specification was confirmed by enforced expression of *Zeb1*, *Snai1* and *Snai2* in wild-type ES cells, followed by PGCLC induction. Among these ES cell lines, BV expression at day 2 of PGCLC induction was severely disturbed in *Zeb1*-Tg line but not in *Snai1*-Tg and *Snai2*-Tg lines (Fig. 5C, Fig. S5C). We noticed that *Zeb1*-Tg aggregates were ruffled and fragile seemingly because of a deficit in their cell adhesion. These results demonstrated that *Zeb1*-mediated progression of EMT hampered PGCs through downregulation of pluripotency-associated genes and PGC-related genes.

Progression of EMT during PGC specification in *Ovol2*$^{-/-}$ embryos

Finally, we verified whether the defective PGC specification in *Ovol2*$^{-/-}$ embryos was attributed to advanced EMT progression, as observed in the *in vitro* system. As a marker of EMT, we examined HMGA2 protein, a representative factor promoting EMT (Dong et al., 2017), because the commercially available ZEB1 antibodies failed to detect endogenous ZEB1 in gastrulating embryos. *Hmga2* was upregulated in *Ovol2*$^{-/-}$ and TKO aggregates in the same manner as *Zeb1*, *Snai1* and *Snai2* (Fig. 2E, Fig. S5D). Immunofluorescence analysis revealed that BV-positive PGCs in E7.5 *Ovol2*$^{+/-}$ embryos showed a negligible level of HMGA2 expression and a specific level of CDH1 (Fig. 6A); the expression level of CDH1 was comparable with that in visceral endoderm, as reported previously (Okamura et al., 2003). In sharp contrast, HMGA2 was clearly visible and CDH1 became faint in BV-positive cells of E7.5 *Ovol2*$^{-/-}$ embryos (Fig. 6A). These results demonstrated that EMT occurred in parallel with PGC specification in BV-positive cells in the *Ovol2*$^{-/-}$ embryos. The level of HMGA2 expression was also elevated in embryonic and extra-embryonic mesodermal cells, suggesting that *Ovol2* modulates EMT in not only PGCs but also surrounding somatic cells.

Based on this series of results, we propose a role for *Ovol2* in the fate determination of PGCs and surrounding somatic cells during gastrulation as follows (Fig. 6B). As gastrulation occurs, OVOL2A

Development (2022) 149, dev200319. doi:10.1242/dev.200319

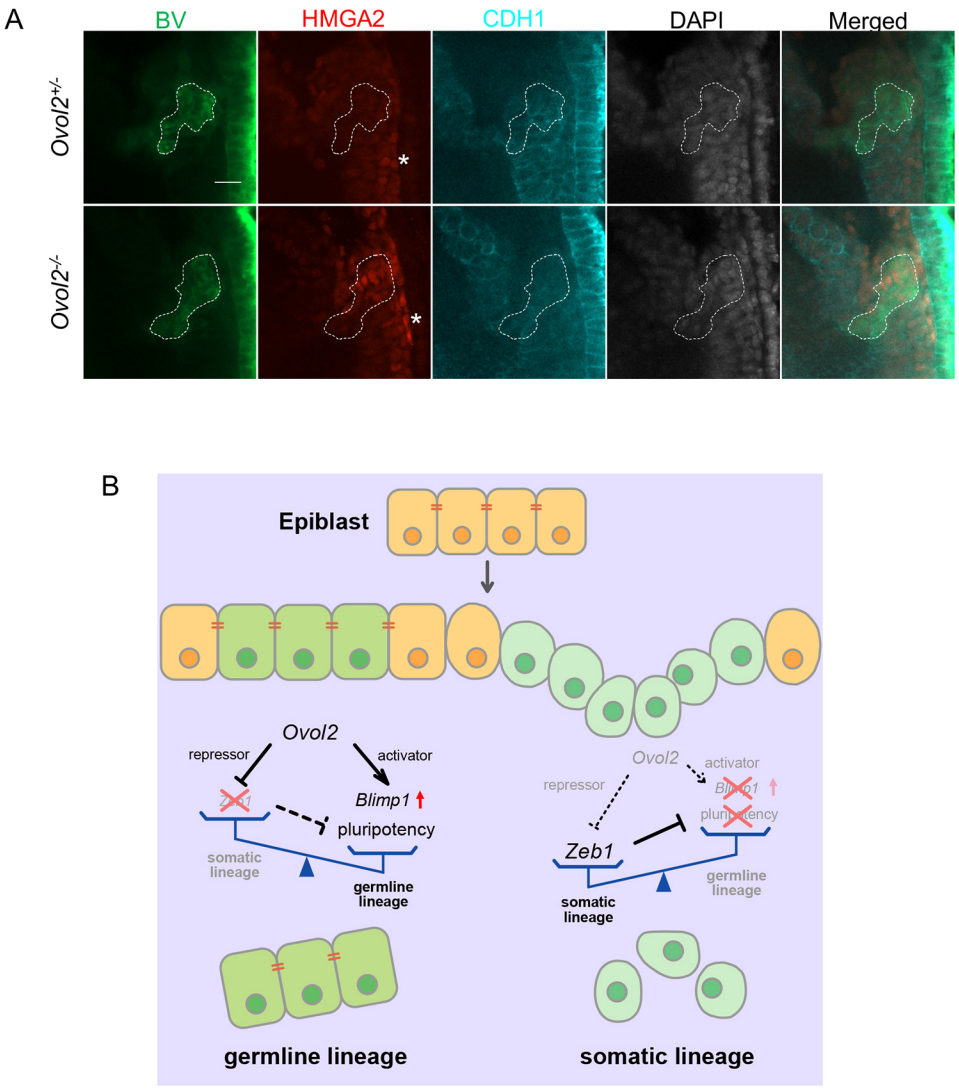

Fig. 6. *Ovol2*$^{-/-}$ PGCs fail to prevent EMT. (A) Impaired EMT repression in *Ovol2*$^{-/-}$ PGCs. BV (green), HMGA2 (red), CDH1 (cyan), DAPI (white) and their merged images in *Ovol2*$^{+/-}$ and *Ovol2*$^{-/-}$ at the EB/MB stage are shown. Dotted lines delineate PGCs recognized by BV expression. Asterisks indicate HMGA2 expression in the embryonic mesoderm. Scale bar: 20 µm. (B) A proposed model for functional requirement of *Ovol2* for PGC specification. During gastrulation, PGC specification and nascent mesoderm differentiation, which are mutually exclusive events, occur simultaneously. *Ovol2* plays dual roles in PGC specification, i.e. promotion of *Blimp1* expression through direct binding to the promoter/enhancer region and repression of *Zeb1*, a key gene promoting EMT at this stage.

directly binds to the TSS of *Zeb1* gene and then represses the gene expression, thereby maintaining an epithelium state with expression of cell adhesion molecules, including *Cdh1*. A continuous epithelium state in part contributes to the maintenance of pluripotency in the epiblast. Simultaneously, OVOL2B directly binds to the promoter/enhancer of *Blimp1* gene and then activates gene expression, which elicits the downstream gene expression program for PGC specification. This double-edged function ensures allocation of the PGC population during gastrulation.

DISCUSSION

Here, we revealed that *Ovol2* balances segregation through the repression of EMT-related genes and the activation of PGC-related genes. Using an *in vitro* culture system, we have found that *Ovol2*, but not *Ovol1* or *Ovol3*, plays a crucial role in PGC specification through the inhibition of EMT-related genes and activation of *Blimp1*. There are distinct role(s) in PGC specification for different gene family member and also for splice variants, as enforced expression of *Ovol2a* or *Ovol2b* in TKO resulted in distinct downstream gene expression (Fig. 2H). Of note, it was revealed that OVOL2B promotes *Blimp1* expression through direct binding to an enhancer region enriched with H3K27ac (Fig. 4F,H). The unique role of OVOL2B may be due to its domain structure, which has only a transactivation domain, as opposed to OVOL2A, which has both transactivation and repression domains. These structural and functional differences resemble the *Drosophila ovo* gene that encodes at least two variants, OVO-A and OVO-B, which have both transactivation and repression domains, and only a transcription activation domain, respectively. These splicing variants have distinctive, nearly opposite, roles in *Drosophila* oogenesis (Andrews et al., 2000). As in the case of *Drosophila* OVO-A and OVO-B, OVOL2A and OVOL2B possess an identical zinc-finger domain, and indeed almost all OVOL2A-binding peaks correspond to OVOL2B-binding peaks (Fig. 4A,B). What makes the difference in the accessibility of each variant to the binding sites in the genome is currently unclear. However, different functions of the variants could be regulated by the distinctive expression dynamics of *Ovol2a* and *Ovol2b*, in which the expression peak of *Ovol2a* is earlier than that of *Ovol2b* (Fig. 1D). Considering the dynamics, it is possible that repression of EMT precedes activation of *Blimp1* during mouse PGC specification.

One of the major outcomes of *Ovol2*-mediated EMT inhibition is that *Cdh1* expression in the epithelium is sustained. Indeed, *Cdh1* expression sharply dropped in *Ovol2*$^{-/-}$ and TKO cells (Fig. 2E, Fig. S2C), and was partially restored by enforced expression of *Ovol2a* and/or *Ovol2b* (Fig. 2H). On the other hand, our results

Development (2022) 149, dev200319. doi:10.1242/dev.200319

revealed that enforced expression of *Cdh1* in TKO rescued only PGC-related gene expression but not pluripotency-associated gene or EMT-related gene expression (Fig. 3D). This suggests that functional requirement of *Cdh1* in PGC specification is to promote the expression of the PGC-related genes, but not the pluripotency-associated genes. Interestingly, deletion of EMT-related genes rather enforced expression of *Cdh1* restored pluripotency-associated gene expression in TKO cells (Figs 3D and 5B), suggesting that EMT-related genes play a dominant role in disrupting pluripotency during gastrulation. Among the EMT-related genes, *Zeb1* was the crucial factor for limiting PGC differentiation, as evidenced by the observation that PGCLCs were induced from TKO ES cells by disruption of *Zeb1* and that enforced expression of *Zeb1* nullified PGCLC induction from wild-type ES cells (Fig. 5A,C). These characteristics were not observed for other EMT-related genes, such as *Snai1* and *Snai2*, emphasizing the unique role of *Zeb1*. It is generally accepted that EMT is not simply a binary switch between the epithelium and mesenchyme, but a gradual transition with multiple intermediate states exhibiting, for example, co-expression of epithelial and mesenchymal markers (Bakir et al., 2020; Hamidi et al., 2020). A system biology approach has revealed that the balance between *Ovol2* and *Zeb1* governs the intermediate states in the MCF10A cell line (Hong et al., 2015). According to the general concept of EMT, there should be epiblast cells in the intermediate states during gastrulation. Considering the opposing roles of *Ovol2* and *Zeb1* in PGC specification, it is plausible that the balance of these genes determines the intermediate state, thereby generating heterogeneity in PGC competence in the embryonic region. Supporting this idea, earlier single cell analysis indicated that *Ovol2* and *Zeb1* are highly heterogenous in epiblast cells at E6.5 (Nakamura et al., 2016).

Given that *Ovol2* and *Zeb1* determine the balance between PGCs and somatic cells, it is still puzzling why enforced expression of *Ovol2* did not induce the entire cell population into PGCLCs. This may be due to the functional threshold of *Ovol2*, as previous studies showed that overexpression of *Ovol2*, by as much as 12,000-fold, did not completely nullify the expression of EMT-associated genes (Roca et al., 2013; Kitazawa et al., 2016; Ye et al., 2016). Viewed from the other side, we should also consider why deletion of *Ovol2* did not result in a complete loss of PGC(LC)s. A clue could be served from our result that $Zeb1^{-/-}$ in TKO restored not only the PGC-related genes but also pluripotency-associated genes (Fig. 5B). Considering that EMT is a multistep process with intermediate states, it is possible that a small proportion of PGCs is specified without active repression of EMT by *Ovol2*. In this regard, *Ovol2* safeguards a specific population of PGCs from EMT during gastrulation.

MATERIALS AND METHODS

Animals and cells

All animal experiments were performed in accordance with the guidelines established by Kyushu University (A20-26-3 and 1-15). $Ovol2^{+/-}$ mice (RBRC02891) were provided by RIKEN BioResource Research Center (Unezaki et al., 2007; Hayashi et al., 2017). BVSC R26rtTA [reverse tetracycline transactivator (rtTA) under the Rosa26 locus] ES cells were provided by Prof. Saitou (Kyoto University, Japan) (Nakaki et al., 2013) and BVSC H18 ES cells (Hikabe et al., 2016) were used in this study. These ES cell lines were maintained under a 2i plus LIF condition without feeders (Ying et al., 2008). HEK293T cells were from American Type Culture Collection (ATCC; CRL-11268).

Vector construction

The CRISPR/Cas9 constructs were generated using pX330 vectors (Addgene 42230) expressing hCas9 and gRNAs against Ovol family genes. Guide RNAs for Ovol genes were designed to delete exon 1 of *Ovol1*, exon 2 of *Ovol2* and exon 3 of *Ovol3*. Guide RNAs for EMT-related genes were designed to delete exons 1 and 2 of *Snail1*, exons 2 and 3 of *Snail2*, and exon 6 of *Zeb1*. Oligos were inserted into BbsI-digested pX330 vector. The *Ovol2* variants and *Cdh1*-coding sequences for forced expression were amplified by PCR flanked with SfiI/NheI and NotI/EcoRI sites from cDNAs derived from ES cells, respectively. To construct these plasmids, cDNAs encoding *Ovol2* variants and *Cdh1* were cloned into PB-TET and PB-CAG destination vectors. To construct pX459-GFP, a GFP fragment from pCAG-Cre:GFP (Addgene 13776) amplified by PCR flanked with an EcoRI site was inserted into EcoRI-digested pX459 (Addgene 62988) vector. Two or four guide RNAs were designed for each gene. Oligos were inserted into BbsI-digested pX459-GFP vector.

Genomic regions containing regulatory elements of *Zeb1*, *Blimp1*, *Nanog* and *T* were amplified from mouse C57BL/6J genomic DNA. These regions were cloned into a pGL4.26-based (Promega, E8441) or a PL-sin-C(3+)A-based (Addgene 21313) luciferase reporter plasmid upstream of a minimal promoter. The primers used in this study are listed in Table S1.

Generation of Ovol-deficient BVSC R26rtTA ES cells

For transfection, the pX330 vectors were transfected into BVSC R26rtTA ES cells with Lipofectamine 3000 (Invitrogen) together with pPB-CAG-rtTA-IRES-Hygro vectors (Addgene 102423) on feeders. The total amount of vector DNA was 2.5 µg. Transfectants were selected with hygromycin B for 2 days (150 µg/ml). Three days after the transfection, the cells were sorted for DsRed2 expression and seeded as a single cell. Single colonies were picked up and cultured on mitomycin C-treated mouse embryonic fibroblasts.

Generation of transgenic ES cells

The PBTET-*Ovol2a* or -*Ovol2b* was transfected into TKO ES cells together with PBase vectors and pGG131 vectors using 4D-Nucleofector (LONZA) in a 60 mm dish under a 2i plus LIF condition. The total amount of DNA was 2 µg. Transfectants were selected with hygromycin B for 4 days. PBCAG-*Cdh1*, PBCAGDD-*Snai1*, -*Snai2* or -*Zeb1*-IRESneo were transfected into TKO and wild-type ES cells together with PBase vectors using Lipofectamine 2000 (Invitrogen) in 12-well plates. The total amount of DNA was 4 µg. Transfectants were selected with G418 for 4 days. After drug selection, cells were seeded singly and single colonies were picked up.

Generation of *Ovol2b*-Tg ES cells with the stable luciferase reporter construct

HEK293T cells were seeded in one well of a 12-well plate. On the next day, PL-sin-C(3+)A-*Blimp1*, HPV275, P633, HPV17 and pHCMV-VSV-G plasmids were transfected into HEK293T cells using Lipofectamine 2000. After 24 h, the medium was replaced with 2i plus LIF medium. The next day, *Ovol2b*-Tg ES cells were seeded in one well of a 24-well plate with virus-containing supernatants from the HEK293T cultures for 24 h. The infected cells were selected with zeocin for 4 days (4 µg/ml).

PGCLC induction

5×10^4 ES cells were cultured in one well of a 24-well plate coated with human plasma fibronectin (Merck Millipore) (16.7 µg/ml) in N2B27 medium containing activin A (20 ng/ml; Preprotech), bFGF (12 ng/ml; Wako) and KSR (1%). The medium was changed 24 h later. The EpiLCs were then cultured under a floating condition by plating 2×10^3 cells in one well of a low-cell-binding U-bottom 96-well plate (Greiner) in GK15+BMP4 (500 ng/ml; R&D Systems), LIF (1000 u/ml; Nacalai), EGF (50 ng/ml; R&D Systems) and SCF (100 ng/ml; R&D Systems). For activation of *Ovol2a*- or *Ovol2b*-Tg, Dox (1.5 µg/ml) was added to the medium.

Q-PCR analysis

Total RNAs from ES cells, EpiLCs, aggregates and sorted BV-positive cells were extracted and purified using an RNeasy Micro Kit (QIAGEN), and reverse transcribed by PrimeScript (Takara). The first-strand cDNAs were used for Q-PCR analysis with Power SYBR Green (ABI).

Development (2022) 149, dev200319. doi:10.1242/dev.200319

Transcriptome analysis

Total RNAs were extracted and purified using an RNeasy Micro Kit, and mRNAs were isolated with the NEBNext poly(A) mRNA magnetic isolation module (NEB). Biologically duplicated samples were prepared at each stage. Purified RNAs were subjected to library construction using a NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB). Adaptor-ligated cDNA libraries were amplified by 12-cycle PCR. Sequencing of the libraries was performed with Hiseq 2500 and Nextseq 550 (Illumina). Obtained reads were mapped to the mouse GRCm38/mm10 genome using Hisat2. Mapped reads were counted by featureCounts. Principal component analysis (PCA) was performed using R software with FactoMineR. For DEG analysis, the false discovery rate (FDR) and $\log_2$ fold-change were calculated using edgeR (Robinson et al., 2010). The DAVID database was used for gene ontology (GO) analysis (Huang et al., 2009).

ChIP-seq analysis

Whole aggregates at day 2 (equivalent to 2×10^6 cells) were trypsinized, washed and collected by centrifugation at 270 ***g*** for 5 min. For crosslinking, the pellets were resuspended in PBS containing 1% formaldehyde, incubated for 10 min and quenched with 125 mM glycine. The fixed cells were resuspended in 1 ml of LB1 [50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% TritonX-100] and pelleted by centrifugation at 1500 ***g*** for 5 min. The pellets were resuspended in 1 ml of LB2 [20 mM Tris (pH 7.5), 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA] and pelleted by centrifugation at 1500 ***g*** for 5 min. The pellets were resuspended in 1 ml of LB3 [20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% TritonX-100, 0.1% sodium deoxycholate and 0.1% SDS] and pelleted by centrifugation at 1500 ***g*** for 5 min. The nuclei were lysed in 200 µl of SDS buffer [20 mM Tris (pH 7.5), 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium-deoxycholate, 1% SDS and protease inhibitor cocktail]. The lysed nuclei were sonicated using a sonicator (Branson) for 10 cycles. Protein-DNA complexes were immunoprecipitated at 4°C overnight using 2 µg of antibodies bound to 50 µl of Dynabeads Protein G (Invitrogen). Immunoprecipitates were washed with 1 ml of LB3 twice, high-salt buffer [20 mM Tris (pH 7.5), 500 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% TritonX-100, 0.1% sodium-deoxycholate and 0.1%SDS], RIPA buffer [50 mM HEPES-KOH (pH 7.4), 0.25 M LiCl, 1 mM EDTA, 0.5% sodium-deoxycholate and 1% NP-40] and TE buffer [50 mM Tris (pH 8.0) and 10 mM EDTA]. For collection of the protein-DNA complexes, beads were resuspended with 100 µl of elution buffer [50 mM Tris (pH 8.0), 10 mM EDTA and 1% SDS]. The immunoprecipitated and input DNA were reverse crosslinked by incubating at 65°C overnight. The mixtures were supplemented with 20 µg of RNaseA and incubated at 37°C for 1 h. After Proteinase K digestion, the DNA was purified using a PCR purification kit (Fastgene, FG-91302) and dissolved with distilled water.

The ChIP and input DNAs were sheared to an average size of ∼150 bp by ultra-sonication (Covaris, S220). Sonicated DNA fragments were end-repaired, ligated to sequencing adapters and amplified according to the manufacturer's instructions (NEB, E7645). Libraries were sequenced using NextSeq 550 with single-end 75 nucleotide read lengths.

For data analysis, ChIP-seq reads were aligned to the mouse reference genome (GRCm38/mm10) using Bowtie v2.3.5 using default parameters (Langmead and Salzberg, 2012). Peaks were called with MACS version 2.2.6 (Zhang et al., 2008) with default settings and visualized using an Integrative Genomics Viewer (Thorvaldsdottir et al., 2013). The consensus sequences of OVOL2A and OVOL2B were identified by HOMER (Heinz et al., 2010). Genomic annotation of the peaks identified from the ChIP-seq data was performed using bedtools (Quinlan and Hall, 2010). Unsupervised hierarchical clustering (UHC) was performed using the hclust function with Pearson correlation distances and Ward's method (ward.D2). The normalized IP/input ratios were determined as peak density divided by input within 1 kb of the TSSs. Previously published H3K4me3, H3K27me3 and H3K27ac ChIP-seq datasets (Kurimoto et al., 2015) were aligned to the mouse reference genome with Bowtie v2.3.5 using the '-N 1 -3 5 –local' options. The reads were processed according to the experimental application, as described above, except that the normalized IP/input ratios were determined as the peak density within 1 kb of the TSSs divided by input within 5 kb. Bivalent genes have been listed previously (Kurimoto et al., 2015).

Luciferase assay

HEK293T cells were used to test the regulatory elements for *Zeb1* and were transfected with the reporter plasmids (80 ng/well) together with the *Ovol2a* or *Ovol2b* expression plasmids (or a mock plasmid) (120 ng/well). The transfections were performed with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Transfected cells were seeded at a density of 7.5×10^4 cells in one well of a 96-well plate and were lysed 24 h after the transfection for analyses using the Dual-Glow Luciferase Assay System (E2920; Promega).

Whole aggregates derived from stable luciferase reporter *Ovol2b*-Tg ES cells at day 1±Dox (equivalent to 5×10^4 cells) were trypsinized and collected by centrifugation. Luciferase assays were performed with a ONE-Glo Luciferase Assay system (Promega, E6110) using Ensight (PerkinElmer).

Immunofluorescence analysis

For whole-mount immunofluorescence analysis of aggregations, aggregates were fixed at day 2 in 4% paraformaldehyde (PFA) in PBS for 1 h at room temperature, washed with PBST (0.2% Tween20), soaked in blocking buffer (PBS containing 0.1% BSA and 0.3% Triton X-100) overnight at 4°C and incubated with primary antibodies diluted with blocking buffer for 2 days at 4°C. The samples were washed with washing buffer (PBS containing 0.3% Triton X-100) and then incubated with secondary antibodies and DAPI overnight at 4°C. Finally, the samples were washed and mounted in Fluoro-KEEPER antifade reagent (Nacalai Tesque, 12593-64). For whole-mount immunofluorescence analysis of embryos, isolated embryos were fixed in 4% PFA in PBS for 1 h at 4°C, washed with PBST (0.2% Tween20) and incubated in blocking solution 1 (PBS containing 1% FBS and 0.2% Tween20) overnight. Embryos were incubated with primary antibodies in blocking solution 1 for 3 days at 4°C, washed with PBST, incubated with secondary antibodies and DAPI for 2 days at 4°C, and then washed and mounted in Fluoro-KEEPER antifade. For immunofluorescence analysis of PGCLCs at day 4, aggregates were trypsinized and then spread onto MAS-coated glass slides (Matsunami, MAS-04). The slides were fixed in 4% PFA in PBS for 15 min at room temperature, washed with PBST and permeabilized with 0.2% Triton X-100 in PBS for 15 min at room temperature. Next, the slides were incubated in blocking solution 2 (PBS containing 5% FBS and 0.2% Tween20) for 1 h at room temperature followed by incubation with primary antibodies in blocking solution 2 overnight at 4°C. After washing with PBST, the slides were incubated with secondary antibodies and DAPI in blocking solution 2 for 1 h at room temperature, washed and mounted in Fluoro-KEEPER antifade. The antibodies used in this study are listed in Table S2.

Acknowledgements

We are grateful to Drs K. Nakashima, T. Matsuda, Y. Ohkawa, T. Ito, F. Miura, and S. Okada for technical support and to Dr M. Saitou for providing BVSC R26rtTA ES cells. We thank the Research Support Center, Kyushu University Graduate School of Medical Sciences for their technical assistance.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: K.H.; Validation: N.H., K.S., Makoto Hayashi, S.K.; Formal analysis: Y.N., G.N., N.H., K.S., Masafumi Hayashi, Makoto Hayashi, S.K., K.H.; Investigation: Y.N., G.N., N.H., Masafumi Hayashi, K.H.; Writing - original draft: Y.N.; Writing - review & editing: K.H.; Supervision: K.H.; Project administration: K.H.; Funding acquisition: K.H.

Funding

This research was funded by KAKENHI Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan (18H05544 and 18H05545 to K.H.; 18H05552 to S.K.); by a Research Fellowship from the Japan Society for the

Promotion of Science (Y.N. and M.H.); by the Takeda Science Foundation (K.H.); by the Luca Bella Foundation (K.H.); and by The Open Philanthropy Project (K.H.).

Data availability

The RNA-seq and ChIP-seq data have been deposited in GEO under accession number GSE184651.

Peer review history

The peer review history is available online at https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.200319.

References

Andrews, J., Garcia-Estefania, D., Delon, I., Lu, J., Mevel-Ninio, M., Spierer, A., Payre, F., Pauli, D. and Oliver, B. (2000). OVO transcription factors function antagonistically in the Drosophila female germline. *Development* **127**, 881-892. doi:10.1242/dev.127.4.881

Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y., Iwanari, H., Mochizuki, Y., Hamakubo, T., Kato, Y., Shirahige, K. et al. (2013). A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell* **27**, 516-529. doi:10.1016/j.devcel.2013.11.001

Aramaki, S., Kagiwada, S., Wu, G., Obridge, D., Adachi, K., Kutejova, E., Lickert, H., Hübner, K. and Schöler, H. R. (2021). Residual pluripotency is required for inductive germ cell segregation. *EMBO Rep.* **22**, e52553. doi:10.15252/embr.202152553

Bakir, B., Chiarella, A. M., Pitarresi, J. R. and Rustgi, A. K. (2020). EMT, MET, plasticity, and tumor metastasis. *Trends Cell Biol.* **30**, 764-776. doi:10.1016/j.tcb.2020.07.003

Chakrabarti, R., Hwang, J., Andres Blanco, M., Wei, Y., Lukačišin, M., Romano, R.-A., Smalley, K., Liu, S., Yang, Q., Ibrahim, T. et al. (2012). Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nat. Cell Biol.* **14**, 1212-1222. doi:10.1038/ncb2607

Chidambaram, A., Allikmets, R., Chandrasekarappa, S., Guru, S. C., Modi, W., Gerrard, B. and Dean, M. (1997). Characterization of a human homolog (OVOL1) of the Drosophila ovo gene, which maps to chromosome 11q13. *Mamm. Genome* **8**, 950-951. doi:10.1007/s003359900620

Cieply, B., Riley, P., Pifer, P. M., Widmeyer, J., Addison, J. B., Ivanov, A. V., Denvir, J. and Frisch, S. M. (2012). Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res.* **72**, 2440-2453. doi:10.1158/0008-5472.CAN-11-4038

Dai, X., Schonbaum, C., Degenstein, L., Bai, W., Mahowald, A. and Fuchs, E. (1998). The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes Dev.* **12**, 3452-3463. doi:10.1101/gad.12.21.3452

De Craene, B. and Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. *Nat. Rev. Cancer* **13**, 97-110. doi:10.1038/nrc3447

Dong, J., Wang, R., Ren, G., Li, X., Wang, J., Sun, Y., Liang, J., Nie, Y., Wu, K., Feng, B. et al. (2017). HMGA2-FOXL2 axis regulates metastases and epithelial-to-mesenchymal transition of chemoresistant gastric cancer. *Clin. Cancer Res.* **23**, 3461-3473. doi:10.1158/1078-0432.CCR-16-2180

Dongre, A. and Weinberg, R. A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69-84. doi:10.1038/s41580-018-0080-4

Hamidi, S., Nakaya, Y., Nagai, H., Alev, C., Kasukawa, T., Chhabra, S., Lee, R., Niwa, H., Warmflash, A., Shibata, T. et al. (2020). Mesenchymal-epithelial transition regulates initiation of pluripotency exit before gastrulation. *Development* **147**, dev184960. doi:10.1242/dev.184960

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532. doi:10.1016/j.cell.2011.06.052

Hayashi, M., Shinozuka, Y., Shigenobu, S., Sato, M., Sugimoto, M., Ito, S., Abe, K. and Kobayashi, S. (2017). Conserved role of Ovo in germline development in mouse and Drosophila. *Sci. Rep.* **7**, 40056. doi:10.1038/srep40056

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589. doi:10.1016/j.molcel.2010.05.004

Hikabe, O., Hamazaki, N., Nagamatsu, G., Obata, Y., Hirao, Y., Hamada, N., Shimamoto, S., Imamura, T., Nakashima, K., Saitou, M. et al. (2016). Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature* **539**, 299-303. doi:10.1038/nature20104

Hong, T., Watanabe, K., Ta, C. H., Villarreal-Ponce, A., Nie, Q. and Dai, X. (2015). An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* **11**, e1004569. doi:10.1371/journal.pcbi.1004569

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57. doi:10.1038/nprot.2008.211

Jägle, S., Busch, H., Freihen, V., Beyes, S., Schrempp, M., Boerries, M. and Hecht, A. (2017). SNAIL1-mediated downregulation of FOXA proteins facilitates the inactivation of transcriptional enhancer elements at key epithelial genes in colorectal cancer cells. *PLoS Genet.* **13**, e1007109. doi:10.1371/journal.pgen.1007109

Kitazawa, K., Hikichi, T., Nakamura, T., Mitsunaga, K., Tanaka, A., Nakamura, M., Yamakawa, T., Furukawa, S., Takasaka, M., Goshima, N. et al. (2016). OVOL2 maintains the transcriptional program of human corneal epithelium by suppressing epithelial-to-mesenchymal transition. *Cell Rep.* **15**, 1359-1368. doi:10.1016/j.celrep.2016.04.020

Kobayashi, T., Zhang, H., Tang, W. W. C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D. A., Webb, R. et al. (2017). Principles of early human development and germ cell program from conserved model systems. *Nature* **546**, 416-420. doi:10.1038/nature22812

Kobayashi, T., Castillo-Venzor, A., Penfold, C. A., Morgan, M., Mizuno, N., Tang, W. W. C., Osada, Y., Hirao, M., Yoshida, F., Sato, H. et al. (2021). Tracing the emergence of primordial germ cells from bilaminar disc rabbit embryos and pluripotent stem cells. *Cell Rep* **37**, 109812. doi:10.1016/j.celrep.2021.109812

Kojima, Y., Sasaki, K., Yokobayashi, S., Sakai, Y., Nakamura, T., Yabuta, Y., Nakaki, F., Nagaoka, S., Woltjen, K., Hotta, A. et al. (2017). Evolutionarily distinctive transcriptional and signaling programs drive human germ cell lineage specification from pluripotent stem cells. *Cell Stem Cell* **21**, 517-532.e15. doi:10.1016/j.stem.2017.09.005

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K. and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* **22**, 1617-1635. doi:10.1101/gad.1649908

Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T. et al. (2015). Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell* **16**, 517-532. doi:10.1016/j.stem.2015.03.002

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923

Lawson, K. A., Dunn, N. R., Roelen, B. A. J., Zeinstra, L. M., Davis, A. M., Wright, C. V. E., Korving, J. P. W. F. M. and Hogan, B. L. M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424-436. doi:10.1101/gad.13.4.424

Lee, B., Villarreal-Ponce, A., Fallahi, M., Ovadia, J., Sun, P., Yu, Q.-C., Ito, S., Sinha, S., Nie, Q. and Dai, X. (2014). Transcriptional mechanisms link epithelial plasticity to adhesion and differentiation of epidermal progenitor cells. *Dev. Cell* **29**, 47-58. doi:10.1016/j.devcel.2014.03.005

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q. et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51-63. doi:10.1016/j.stem.2010.04.014

Liu, P., Wakamiya, M., Shea, M. J., Albrecht, U., Behringer, R. R. and Bradley, A. (1999). Requirement for Wnt3 in vertebrate axis formation. *Nat. Genet.* **22**, 361-365. doi:10.1038/11932

Lü, J., Andrews, J., Pauli, D. and Oliver, B. (1998). Drosophila OVO zinc-finger protein regulates ovo and ovarian tumor target promoters. *Dev. Genes Evol.* **208**, 213-222. doi:10.1007/s004270050175

Mevel-Ninio, M., Terracol, R. and Kafatos, F. C. (1991). The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development. *EMBO J.* **10**, 2259-2266. doi:10.1002/j.1460-2075.1991.tb07762.x

Nair, M., Bilanchone, V., Ortt, K., Sinha, S. and Dai, X. (2007). Ovol1 represses its own transcription by competing with transcription activator c-Myb and by recruiting histone deacetylase activity. *Nucleic Acids Res.* **35**, 1687-1697. doi:10.1093/nar/gkl1141

Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y. and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222-226. doi:10.1038/nature12417

Nakamura, T., Yabuta, Y., Okamoto, I., Aramaki, S., Yokobayashi, S., Kurimoto, K., Sekiguchi, K., Nakagawa, M., Yamamoto, T. and Saitou, M. (2015). SC3-seq: a method for highly parallel and quantitative measurement of single-cell gene expression. *Nucleic Acids Res.* **43**, e60. doi:10.1093/nar/gkv134

Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Yamamoto, T. and Saitou, M. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57-62. doi:10.1038/nature19096

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A. et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213. doi:10.1038/nature03813

Ohinata, Y., Sano, M., Shigeta, M., Yamanaka, K. and Saitou, M. (2008). A comprehensive, non-invasive visualization of primordial germ cell development in

mice by the Prdm1-mVenus and Dppa3-ECFP double transgenic reporter. *Reproduction* **136**, 503-514. doi:10.1530/REP-08-0053

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T. and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571-584. doi:10.1016/j.cell.2009.03.014

Okamura, D., Kimura, T., Nakano, T. and Matsui, Y. (2003). Cadherin-mediated cell interaction regulates germ cell determination in mice. *Development* **130**, 6423-6430. doi:10.1242/dev.00870

Oliver, B., Perrimon, N. and Mahowald, A. P. (1987). The ovo locus is required for sex-specific germ line maintenance in Drosophila. *Genes Dev.* **1**, 913-923. doi:10.1101/gad.1.9.913

Oliver, B., Pauli, D. and Mahowald, A. P. (1990). Genetic evidence that the ovo locus is involved in Drosophila germ line sex determination. *Genetics* **125**, 535-550. doi:10.1093/genetics/125.3.535

Oliver, B., Singer, J., Laget, V., Pennetta, G. and Pauli, D. (1994). Function of Drosophila ovo+ in germ-line sex determination depends on X-chromosome number. *Development* **120**, 3185-3195. doi:10.1242/dev.120.11.3185

Payre, F., Vincent, A. and Carreno, S. (1999). ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature* **400**, 271-275. doi:10.1038/22330

Peinado, H., Olmeda, D. and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat. Rev. Cancer* **7**, 415-428. doi:10.1038/nrc2131

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842. doi:10.1093/bioinformatics/btq033

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140. doi:10.1093/bioinformatics/btp616

Roca, H., Hernandez, J., Weidner, S., McEachin, R. C., Fuller, D., Sud, S., Schumann, T., Wilkinson, J. E., Zaslavsky, A., Li, H. et al. (2013). Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PLoS ONE* **8**, e76773. doi:10.1371/journal.pone.0076773

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T. A., Datti, A., Woltjen, K., Nagy, A. and Wrana, J. L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64-77. doi:10.1016/j.stem.2010.04.015

Senft, A. D., Bikoff, E. K., Robertson, E. J. and Costello, I. (2019). Genetic dissection of Nodal and Bmp signalling requirements during primordial germ cell development in mouse. *Nat. Commun.* **10**, 1089. doi:10.1038/s41467-019-09052-w

Takaku, M., Grimm, S. A., Shimbo, T., Perera, L., Menafra, R., Stunnenberg, H. G., Archer, T. K., Machida, S., Kurumizaka, H. and Wade, P. A. (2016). GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.* **17**, 36. doi:10.1186/s13059-016-0897-0

Thorvaldsdottir, H., Robinson, J. T. and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178-192. doi:10.1093/bib/bbs017

Unezaki, S., Horai, R., Sudo, K., Iwakura, Y. and Ito, S. (2007). Ovol2/Movo, a homologue of Drosophila ovo, is required for angiogenesis, heart formation and placental development in mice. *Genes Cells* **12**, 773-785. doi:10.1111/j.1365-2443.2007.01084.x

Watanabe, K., Villarreal-Ponce, A., Sun, P., Salmans, M. L., Fallahi, M., Andersen, B. and Dai, X. (2014). Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. *Dev. Cell* **29**, 59-74. doi:10.1016/j.devcel.2014.03.006

Weber, S., Eckert, D., Nettersheim, D., Gillis, A. J. M., Schäfer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L. H. J. et al. (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biol. Reprod.* **82**, 214-223. doi:10.1095/biolreprod.109.078717

Wu, R.-S., Hong, J.-J., Wu, J.-F., Yan, S., Wu, D., Liu, N., Liu, Q.-F., Wu, Q.-W., Xie, Y.-Y., Liu, Y.-J. et al. (2017). OVOL2 antagonizes TGF-beta signaling to regulate epithelial to mesenchymal transition during mammary tumor metastasis. *Oncotarget* **8**, 39401-39416. doi:10.18632/oncotarget.17031

Ye, G. D., Sun, G. B., Jiao, P., Chen, C., Liu, Q. F., Huang, X. L., Zhang, R., Cai, W. Y., Li, S. N., Wu, J. F. et al. (2016). OVOL2, an inhibitor of WNT signaling, reduces invasive activities of human and mouse cancer cells and is down-regulated in human colorectal tumors. *Gastroenterology* **150**, 659-671.e16. doi:10.1053/j.gastro.2015.11.041

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523. doi:10.1038/nature06968

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. doi:10.1186/gb-2008-9-9-r137

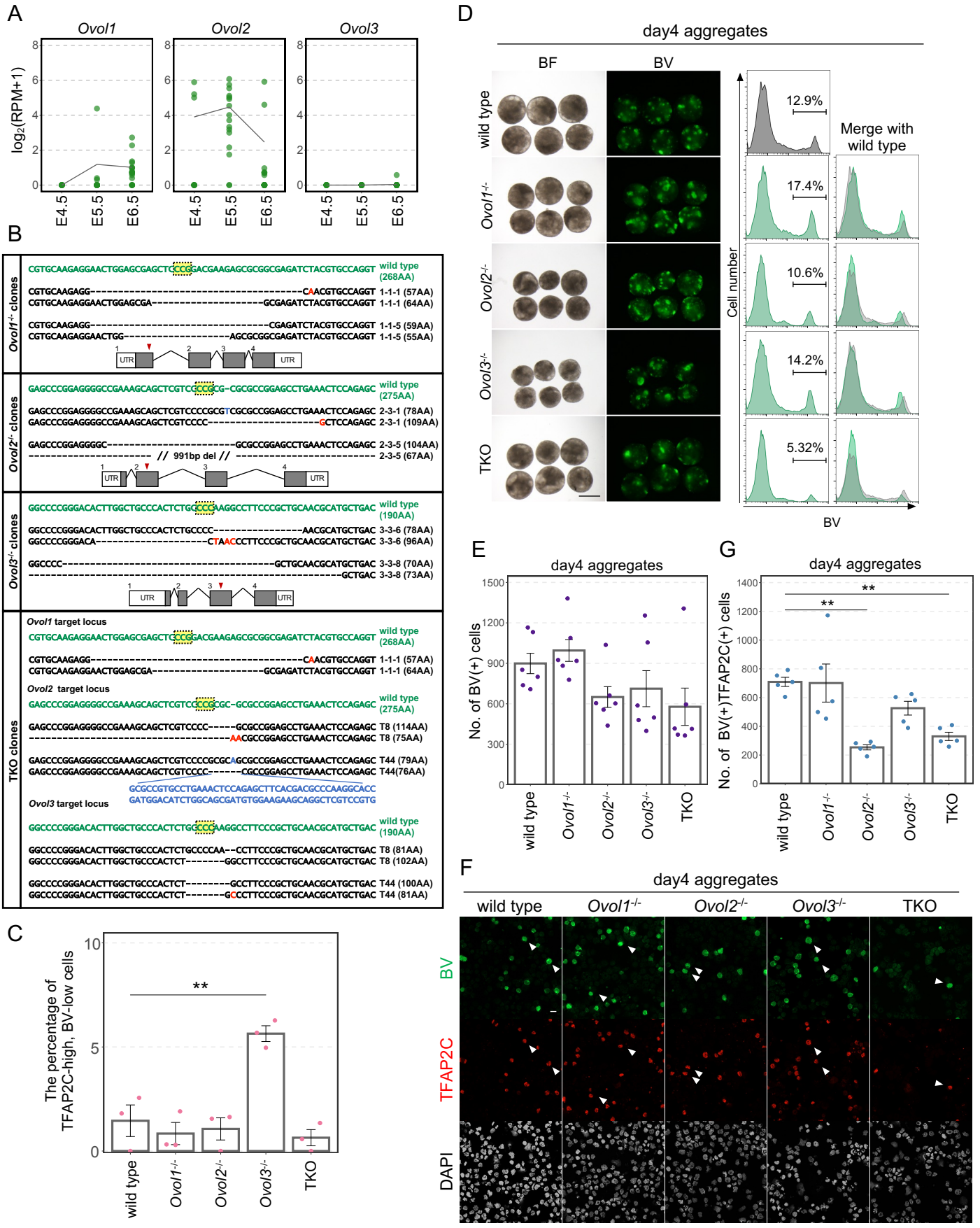

Fig. S1. Ovol genes contribute to PGC specification.

(A) Expression of Ovol genes in epiblast cells in vivo. The data are from Nakamura et al (Nakamura et al., 2016). Mean values are connected by the lines. (B) Disruption of Ovol genes by CRISPR/Cas9. The genomic structures and the CRISPR target sites (red arrowheads) are shown. The DNA sequence in green shows the sequence around the PAM sequence (yellow). The DNA sequences in black show both alleles in the KO line. Red and blue characters indicate point mutations and insertions, respectively. (C) The percentage of TFAP2C-high and BV-low expressing cells. **p<0.01, by Tukey's test. (D) Aggregates from Ovol gene-KO ES cells at day 4 of PGCLC induction. Images of BF and BV, and FACS analysis of BV expression in Ovol gene-KO and wild type day 4 aggregates are shown. Scale bar, 500 µm. (E) The number of BV-positive cells in day 4 aggregates. The mean values with SDs are from three independent experiments. (F) The immunofluorescence analysis of BV and TFAP2C in day 4 aggregates. Arrowheads indicate BV- and TFAP2C-positive cells. Scale bar, 20 µm. (G) The number of BV- and TFAP2C-positive cells in day 4 aggregates. **p<0.01, by Tukey's test.

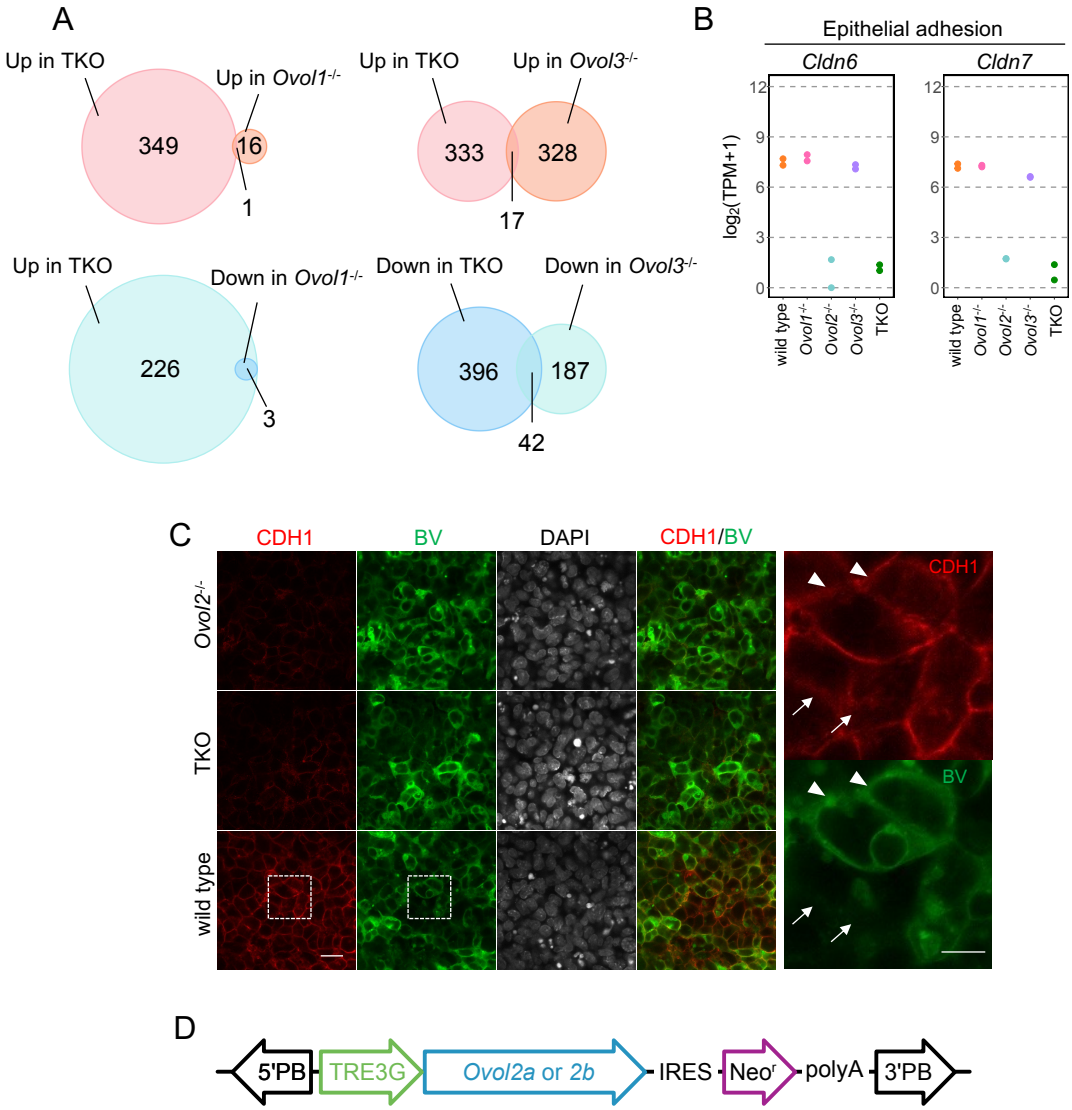

Fig. S2. Epithelial adhesion is impaired in Ovol2-/- and TKO aggregates.

(A) Venn diagrams representing the overlaps of DEGs. Comparisons of DEGs upregulated (top) or downregulated (bottom) in Ovol1-/- or Ovol3-/- and TKO compared to wild type at day 2 of PGCLC induction are shown. (B) Expression levels of genes involved in epithelial adhesion in day 2 aggregates. The log2 (TPM+1) values from two independent RNA-seq analyses are shown. (C) Immunofluorescence analysis of CDH1 and BV in Ovol2-/-, TKO, and wild type day 2 aggregates. Arrowheads indicate cells with high expression of BV and CDH1, and arrows indicate cells with low expression of BV and CDH1. Scale bar, 20 µm. (D) Dox-inducible transgenic construct for enforced expression of Ovol2a and Ovol2b.

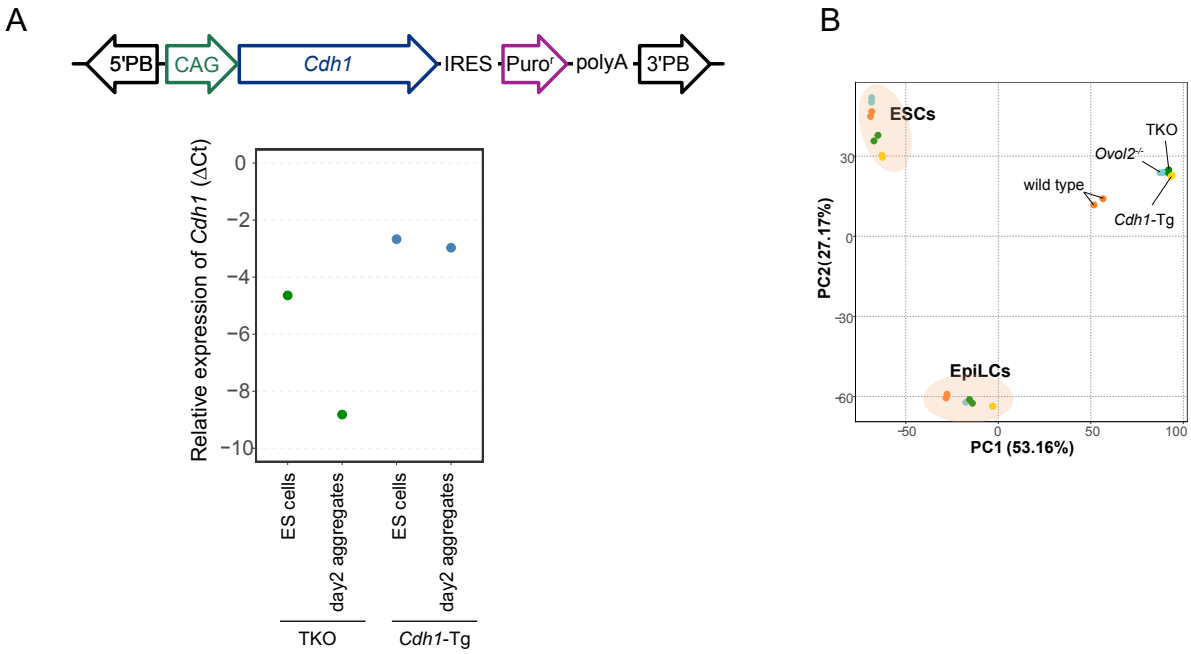

Fig. S3. Cell adhesion up-regulates genes involved in PGCs.

(A) Generation of Cdh1-Tg ES cells. The PiggyBac transposon-based construct for enforced expression of Cdh1 is shown (top). Q-PCR analysis to estimate the expression levels of Cdh1 (bottom). ΔCt values were determined by comparison to the house-keeping genes, Arbp and Ppia. The average ΔCt values from two technical duplicates are shown. (B) PCA of transcriptomes of Cdh1-Tg ES cells, EpiLCs, and day 2 aggregates. Each color corresponds to the genotype indicated.

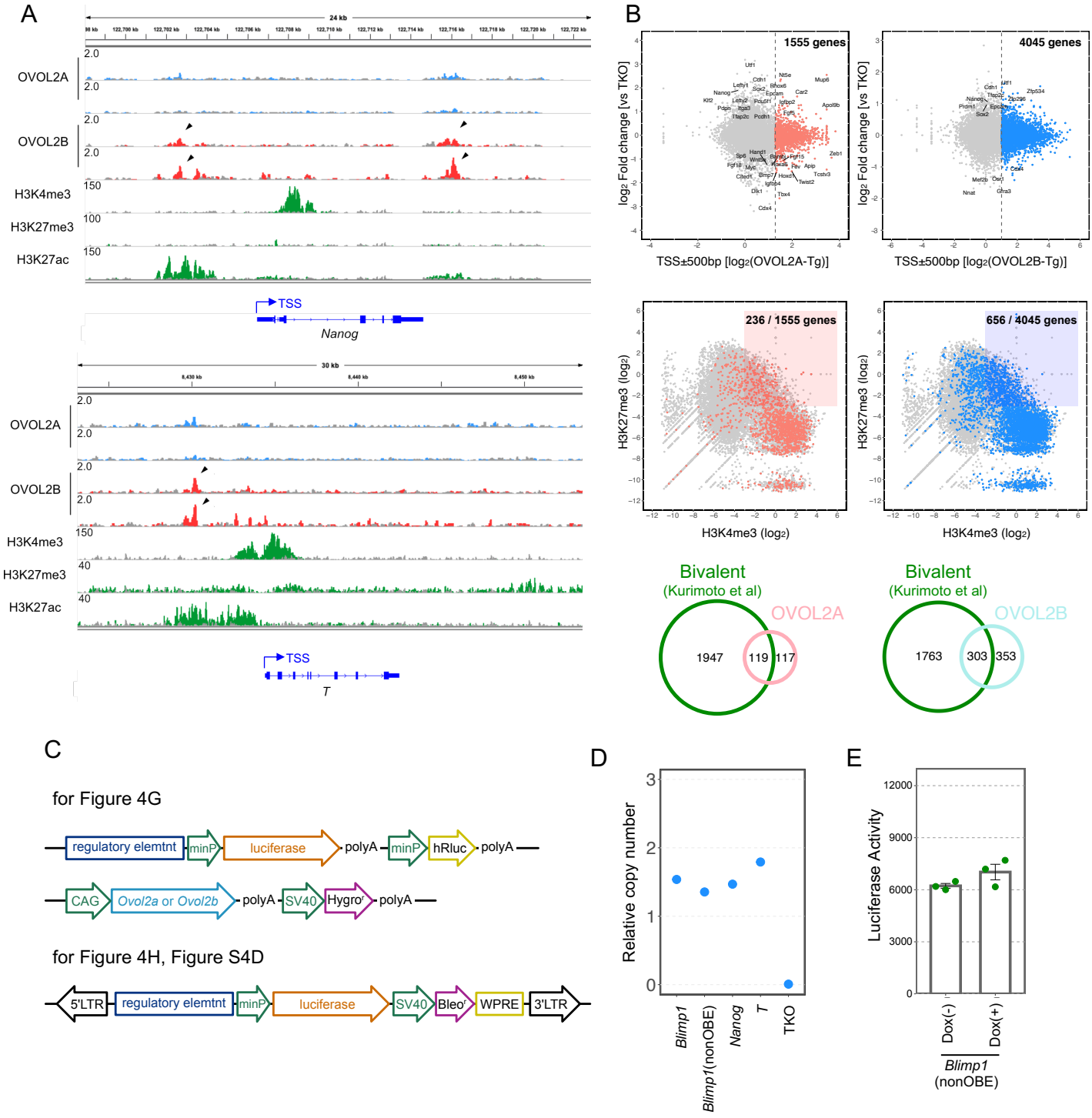

Fig. S4. Ovol2b does not regulate the expression of Nanog or T.

(A) OVOL2A- or OVOL2B-binding in Nanog and T loci. ChIP-seq tracks of OVOL2A- (blue) or OVOL2B (red)-binding region, H3K4me3, H3K27me3 and H3K27ac around Nanog and T loci are shown as CPM. Grey in the track shows CPM from input. (B) Bivalent histone modifications at OVOL2A- or OVOL2B-binding gene loci. The top plots show the log2-fold change of gene expression in Ovol2a- (left) or Ovol2b- (right) Tg, compared to TKO, at the Y-axis, and the enrichment of OVOL2A- (left) or OVOL2B- (right) binding sites around all TSSs at the X-axis. The cut-off lines of the enrichment are 1.28 for OVOL2A and 0.99 for OVOL2B. The middle plots show the log2 levels of H3K4me3 (X-axis) and H3K27me3 (Y-axis) enrichment. Red or blue dots are the highly enriched genes in OVOL2A (left) or OVOL2B (right) shown in the top plots. The bottom Venn diagrams show the overlap of the bivalent histone modifications between OVOL2A- or OVOL2B-binding genes and all genes in day 2 PGCLCs reported previously (Kurimoto et al., 2015). (C) Constructs for the reporter assay. Constructs of a luciferase-expressing vector (top), Ovol2a- or Ovol2b-expressing vector (middle), and a lentiviral vector expressing the luciferase reporter (bottom) are shown. (D) Estimation of the relative copy number of the luciferase constructs. ΔCt value of each reporter construct was determined by comparison to Il2 locus. The average ΔCt values from two technical duplicates are shown. (E) Luciferase analysis using the Blimp1/Prdm1-luc reporter without OBE.

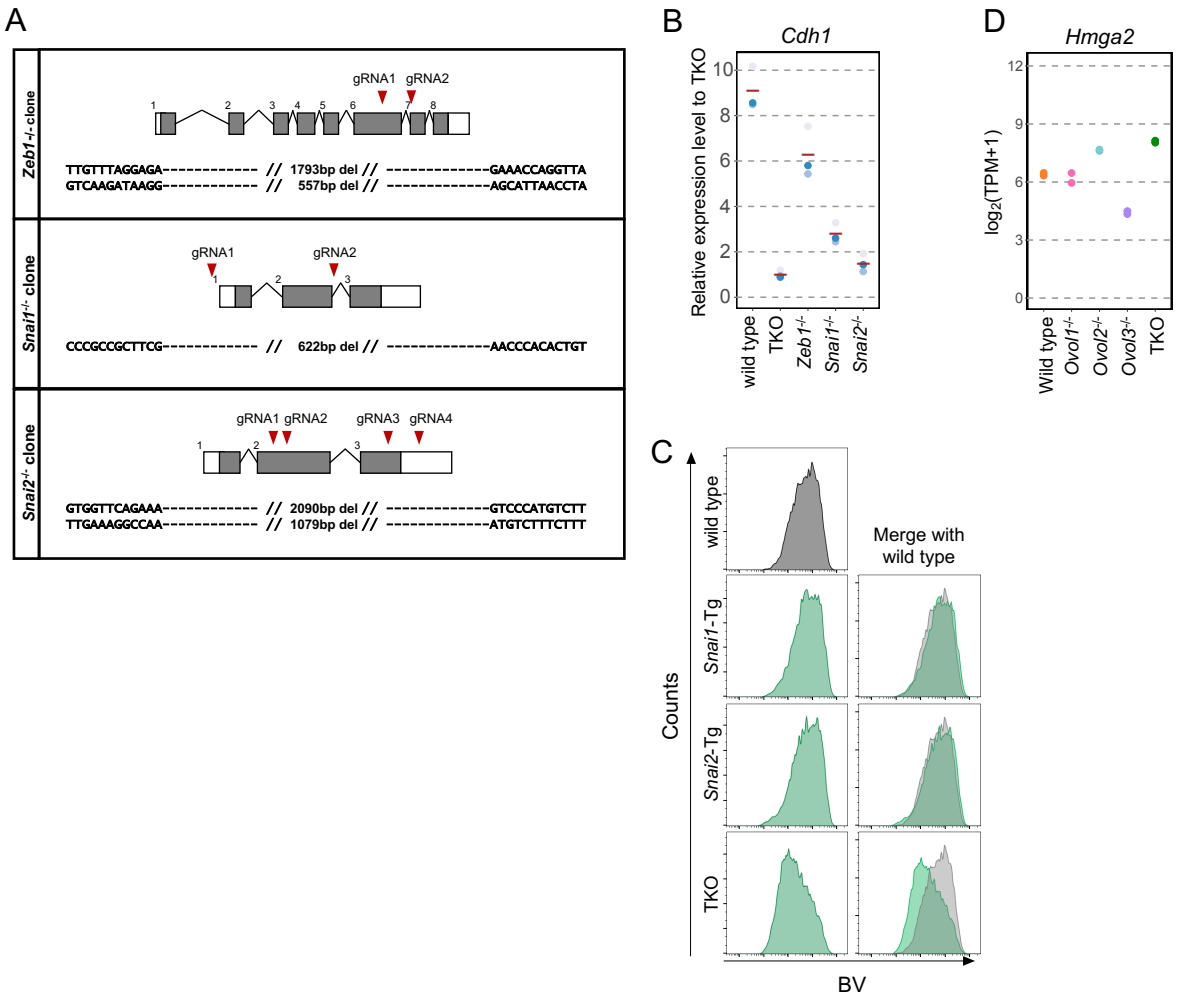

Fig. S5. The repression of EMT promotes PGC specification.

(A) Disruption of Zeb1, Snai1 or Snai2 by CRISPR/Cas9. The CRISPR target sites (red arrowheads) are depicted at each panel. DNA sequences show both alleles in the KO line. The types of deletion in the Snai1-/- clone were all the same, possibly due to a large deletion or the same deletion in both alleles. The gRNA and primer sequences are listed in Table 1. (B) Q-PCR analysis of Cdh1 gene. The relative ΔCt values in day 2 aggregations are shown. The value for TKO was set as 1. The plots were compiled from three independent experiments. The mean values are indicated as lines. (C) FACS analysis of BV expression in Snai1-Tg, Snai2-Tg, wild type, and TKO aggregates at day 2. It was not possible to analyze Zeb1-Tg aggregates, as they were fragile and less viable. (D) Expression levels of Hmga2 in Ovol gene-KO aggregates. The log2 (TPM+1) values from two independent RNA-seq analyses are shown.

Table S1. Oligonucleotide sequences for primers and gRNAs used in this study.

(CRISPR)	
Name	**Sequence**
*Ovol1*_gRNA_Fw	CACCGATCTCGCCGCGCTCTTCGTC
*Ovol1*_gRNA_Rv	AAACGACGAAGAGCGCGGCGAGATC
*Ovol2*_gRNA_Fw	CACCGTTTCAGGCTCCGGCGCGCG
*Ovol2*_gRNA_Rv	AAACCGCGCGCCGGAGCCTGAAAC
*Ovol3*_gRNA_Fw	CACCGCGTTGCAGCGGGAAGGCCTT
*Ovol3*_gRNA_Rv	AAACAAGGCCTTCCCGCTGCAACGC
*Snai1*_gRNA1_Fw	CACCCTCGAGGTGGGGGCGTGACCGTACTGT
*Snai1*_gRNA1_Rv	AAACACAGTACGGTCACGCCCCCACCTCGAG
*Snai1*_gRNA2_Fw	CACCCTCGAGGGTGTGGGTTGAGCCCGGATA
*Snai1*_gRNA2_Rv	AAACTATCCGGGCTCAACCCACACCCTCGAG
*Snai2*_gRNA1_Fw	CACCCTCGAGGAATAGGGCTGTATGCTCCCG
*Snai2*_gRNA1_Rv	AAACCGGGAGCATACAGCCCTATTCCTCGAG
*Snai2*_gRNA2_Fw	CACCCTCGAGGCATTCTGTTTGAGTAAACAC
*Snai2*_gRNA2_Rv	AAACATTAGTGACGAAGAGGAGAG
*Snai2*_gRNA3_Fw	CACCCTCGAGGTTTATGCAGAAGCGACATTC
*Snai2*_gRNA3_Rv	AAACGAATGTCGCTTCTGCATAAA
*Snai2*_gRNA4_Fw	CACCCTCGAGGAAAGACATGGGACACGCACC
*Snai2*_gRNA4_Rv	AAACGGTGCGTGTCCCATGTCTTT
*Zeb1*_gRNA1_Fw	CACCCTCGAGGTTGCAGTTCGGGCATTCGTA
*Zeb1*_gRNA1_Rv	AAACTACGAATGCCCGAACTGCAACCTCGAG
*Zeb1*_gRNA2_Fw	CACCCTCGAGGTTACTAACCTGGTTTCCGTT
*Zeb1*_gRNA2_Rv	AAACAACGGAAACCAGGTTAGTAACCTCGAG
px459_eGFP_Fw	AAAGAAAAAGGAATTCGGCAGTGGAGCTAGCGCCACTAACTTCTCCCTGTTG
px459_eGFP_Rv	GCTCTAGTTAGAATTCTTACTTGTACAGCTCGTCC

(Genotyping)	
Name	**Sequence**
*Ovol1*_type_Fw	GTACACTCTGTCCTAAGAACGTGG
*Ovol1*_type_Rv	CGGGTTTTCTGGATTTACGACACG
*Ovol2*_type_Fw	GGCCTGAGAACCCGTCTCCCGACA
*Ovol2*_type_Rv	AGGTCACTAGGAGCCTGGCTGCGG
*Ovol3*_type_Fw	AGTGTGTGAGGCCACTTCTGTATG
*Ovol3*_type_Rv	CTGGAAACCCCATACCCCATGCGT
*Snai1*_type_Fw	TCACACCTTTCCTAAGCGGC
*Snai1*_type_Rv	GCAAGTGTGAAATCGGCACC
*Snai2*_type_Fw	CCAGTGCTCAAGAACCGAGA
*Snai2*_type_Rv	TAGGCGTGGCTATTAACCGT
*Zeb1*_type_Fw	ACTGCTGAGAAAGACAGGGC
*Zeb1*_type_Rv	CCTGGAGGCCTTTGTACCAG
*β*geo_type_Fw	GCTTGCCGAATATCATGGTG
*β*geo_type_Rv	CTTCAGCAATATCACGGGTAGC
Blimp1-mVenus_Fw	ACTCATCTCAGAAGAGGATCTG
Blimp1-mVenus_Rv	CACAGTCGAGGCTGATCTCC
*Ovol2*KO_type_Fw	CATAGCCCATGTGTGGCTGCTG
*Ovol2*KO_type_Rv	GCCGGCCTTAAAACATCCCAC

(Cloning)	
Name	**Sequence**
*Ovol2a*_CDS_Fw	AAGCTTATGCCCAAAGTCTTTTTG
*Ovol2a*_CDS_Rv	CTCGAGTCACTTTTTTTCCTCCTC
*Ovol2b*_CDS_Fw	AGCCTGAAACTCCAGAGCTTC
*Ovol2c*_CDS_Fw	TGTGACCTTTGTGGCAAGAG
*Ovol2b2c*_CDS_Rv	TTAACTCGAGGGATCCAAGC
*Cdh1*_CDS_Fw	AAGCTTATGGGAGCCCGGTGCCGC
*Cdh1*_CDS_Rv	GAATTCCTAGTCGTCCTCACCACCGCCGTAC
*Snai1*_CDS_Fw	CATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGCTTATGCCGCGCTCCTTCCTG
*Snai1*_CDS_Rv	CCTGCGGTCGCGGCCGCTCAGCGAGGGCCTCCGGAGCA
*Snai2*_CDS_Fw	CATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGCTTATGCCGCGCTCCTTCCTGG
*Snai2*_CDS_Rv	CCTGCGGTCGCGGCCGCTCAGTGTGCCACACAGCAG
*Zeb1*_CDS_Fw	CATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGCTTATGGCGGATGGCCCCAGG
*Zeb1*_CDS_Rv	CCTGCGGTCGCGGCCGCCTAAGCTTCATTTGTCTTC

(Luciferase assay)	
Name	**Sequence**
*Zeb1*_RE_Fw	GCTCGCTAGCCTCGAGTGTCTGTGCTATTTCCCGGC
*Zeb1*_RE_Rv	TCTAGTGTCTAAGCTTTCGTAAAGCCTCGAGTGTCG
*Prdm1*_RE_Fw	GCTCGCTAGCCTCGAGCTTACAACTTGGTCCCCATG
*Prdm1*_RE_Rv	TCTAGTGTCTAAGCTTAAGCAGCTCTCTGCATTGCC
*Nanog*_RE_Fw	GCTCGCTAGCCTCGAGGGAGAGCAATGAATAAGAAATG
*Nanog*_RE_Rv	TCTAGTGTCTAAGCTTTCACTTTGTAGACCAGGCTG
*T*_RE_Fw	GCTCGCTAGCCTCGAGAGCTTCGGGTGCCTTATCC
*T*_RE_Rv	TCTAGTGTCTAAGCTTGTCCCCACTCCCCCTAGTTTC

(Copy number estimation)	
Name	**Sequence**
*IL2*_qFw	CTAGGCCACAGAATTGAAAGATCT
*IL2*_qRv	GTAGGTGGAAATTCTAGCATCATCC
Luc_qFw	AACACCCCAACATCTTCGAC
Luc_qRv	GATCTCCTTCTCGGTCATGG

(Q-PCR)	
Name	**Sequence**
*Arbp*_qFw	CAAAGCTGAAGCAAAGGAAGAG
*Arbp*_qRv	AATTAAGCAGGCTGACTTGGTTG
*Ppia*_qFw	TTACCCATCAAACCATTCCTTCTG
*Ppia*_qRv	AACCCAAAGAACTTCAGTGAGAGC
*Ovol1*_qFw	GGAGACCTTTTTACCTGCCAC
*Ovol1*_qRv	GCCCTTCCCACAGTAAGTGC
*Ovol2*total_qFw	GGACCTGTATCTGCATGTGAAC
*Ovol2*total_qRv	GGCCGCCAACTTTTTGGAAG
*Ovol2a*_qFw	GAAGAGAGGATCCCACCATG
*Ovol2a*_qRv	GTCGCGGAGCAGACAGCCCAGGCTC
*Ovol2b*_qFw	GCCTGGAAGTGGGTGGAA
*Ovol2b*_qRv	TGTGCCCTCGTCTTGCCTC
*Ovol2c*_qFw	GGATGGCTAGGGTAGACTCG
*Ovol2c*_qRv	ACCCCGACACTCAACCACAC
*Ovol3*_qFw	GCTGACTCGGCACCTCAAAT
*Ovol3*_qRv	CGGAATGGCCGTATTCCAGT
*Pou5f1*_qFw	GATGCTGTGAGCCAAGGCAAG
*Pou5f1*_qRv	GGCTCCTGATCAACAGCATCAC
*Nanog*_qFw	CTTTCACCTATTAAGGTGCTTGC
*Nanog*_qRv	TGGCATCGGTTCATCATGGTAC
*Sox2*_qFw	CATGAGAGCAAGTACTGGCAAG
*Sox2*_qRv	CCAACGATATCAACCTGCATGG
*Blimp1*_qFw	AGCATGACCTGACATTGACACC
*Blimp1*_qRv	CTCAACACTCTCATGTAAGAGGC
*Prdm14*_qFw	ACAGCCAAGCAATTTGCACTAC
*Prdm14*_qRv	TTACCTGGCATTTTCATTGCTC
*Tfap2c*_qFw	GGGCTTTTCTCTCTTGGCTGGT
*Tfap2c*_qRv	TCCACACGTCACCCACACAA
*Snai1*_qFw	AGCAGGGTGGTTACTGGACAC
*Snai1*_qRv	CCATTATTCATGGTCCCTTCTG
*Snai2*_qFw	GATGTGCCCTCAGGTTTGAT
*Snai2*_qRv	ACACATTGCCTTGTGTCTGC
*Zeb1*_qFw	ACCGCCGTCATTTATCCTGAG
*Zeb1*_qRv	CATCTGGTGTTCCGTTTTCATCA
*Cdh1*_qFw	GAGGCCAAGCAGCAATAC
*Cdh1*_qRv	CACGTCTACCACGTCCACAG

Table S2. Antibodies used in this study.

Name	Description	Company	Cat.No.	Dilution
Anti-GFP	Chicken polyclonal	Abcam	ab13970	1:500
Anti-CDH1	Rabbit monoclonal	CST	3195	1:200
Anti-CDH1	Rat monoclonal	Santa Cruz	sc-59778	1:100
Anti-HMGA2	Rabbit monoclonal	CST	8179	1:200
Anti-AP2γ	Mouse monoclonal	Santa Cruz	sc-12762	1:100

Name	Description	Company	Cat.No.	Dilution
Anti-Chick IgY	Donkey polyclona / Alexa488	Jackson ImmunoResearch	703-545-155	1:500
Anti-Rabbit IgG	Donkey polyclona / Alexa568	Life Technologies	A10042	1:500
Anti-Rabbit IgG	Goat polyclona / AlexaPlus555	Life Technologies	A32732	1:500
Anti-Rat IgG	Goat polyclonal / Alexa633	Life Technologies	A11077	1:500
Anti-Mouse IgG	Goat polyclona / AlexaPlus555	Life Technologies	A32727	1:500