PNAS
www.pnas.org

Supplementary Information for

Mammalian germ cells are determined after PGC colonization of the nascent gonad

Peter K. Nicholls, Hubert Schorle, Sahin Naqvi, Yueh-Chiang Hu, Yuting Fan, Michelle A. Carmell, Ina Dobrinski, Adrienne L. Watson, Daniel F. Carlson, Scott C. Fahrenkrug and David C. Page*

* Correspondence to: dcpage@wi.mit.edu

This PDF file includes:

SI Discussion
SI Materials and Methods
Figs. S1 to S7
Tables S1 to S6
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

www.pnas.org/cgi/doi/10.1073/pnas.1910733116

SI Discussion

Comparison of *Dazl*-deficient mice on four genetic backgrounds

Given numerous reports that germline phenotypes vary according to genetic background, we considered whether the *Dazl*-dependent restriction of developmental potential is unique to the B6 strain. We approached this question amidst conflicting reports regarding *Dazl*'s role in germ cell viability. Specifically, in *Dazl*-deficient mice on an inbred B6 background, PGCs that colonize the nascent gonads subsequently die; no germ cells remain in the post-natal gonads of either sex (1). By contrast, on a mixed genetic background, some *Dazl*-deficient germline cells survive embryogenesis in males [although the animals are sterile (2), with defects in spermatogonial differentiation (3) and meiosis (4)]. These seemingly conflicting observations, albeit in different genetic backgrounds, raised the possibility that *Dazl* might be necessary to restrict developmental potential in B6 mice, but not in other inbred mouse strains.

To address whether our findings in B6 mice are unique to that strain, and to clarify the role of *Dazl* in germline survival, we backcrossed the $Dazl^{tm1Hjc}$ null allele to several *M. m. domesticus* strains of western European origin, and examined the testes of *Dazl*-deficient mice for the presence of germ cells. In juvenile males of the *M. m. domesticus* strains B6 and FVB/NTac (FVB), no germ cells were observed in *Dazl*-deficient testes at post-natal day 5, indicating that *Dazl* is necessary for embryonic germline survival in these strains (SI Appendix, Fig. S5A). However, in males of the *M. m. domesticus* strain 129S4, some *Dazl*-deficient germline cells survive embryogenesis, and a few commence but fail to complete spermatogenesis (SI Appendix, Fig. S5A and B). Analysis of 129.*Dazl*-deficient spermatocytes by meiotic spreads shows that many chromosomes remain un-synapsed, with no cells progressing beyond the zygotene stage (SI Appendix, Fig. S5C). Looking further afield, we backcrossed the $Dazl^{tm1Hjc}$ null allele to the *M. m. molossinus* strain (MSM/MsJ; MSM) derived from wild-trapped Japanese mice; the MSM and *M. m. domesticus* lineages diverged more than a million years ago. On an inbred MSM background, far fewer germline cells were observed in the post-natal testes of *Dazl*-deficient mice (SI Appendix, Fig. S5A). These findings demonstrate that expression of *Dazl*, beginning shortly after PGCs colonize the nascent gonads, is critical for germline survival in the testis, not only in multiple *M. m. domesticus* strains, but also in the distant outgroup *M. m. molossinus*. By contrast, no oocytes are observed in the post-natal ovary of *Dazl*-deficient females in any strain (SI Appendix, Fig. S5D).

We next considered the testicular phenotype of the *Dazl*-deficient germline prior to germline loss. In embryos, newly arrived PGCs rapidly proliferate, ceasing shortly after the commencement of male sexual differentiation. In B6.*Dazl* embryos, these newly arrived PGCs

remain proliferative for an extended period, until at least E15.5 (5). To test if the germline of *Dazl*-deficient embryos retains an extended period of mitosis in other genetic backgrounds, we injected pregnant mice with EdU, a thymidine analog that incorporates into the DNA of proliferative cells. As anticipated, EdU was detected in the proliferative somatic lineages of the testis in control and *Dazl*-deficient embryos at E17.5 (SI Appendix, Fig. S5E). By contrast, germline cells in control embryos did not incorporate EdU, whereas the germline of *Dazl*-deficient littermates was regularly EdU positive, indicating DNA synthesis and retained mitotic activity, regardless of genetic background.

Given the extended proliferation of the *Dazl*-deficient germline in the embryonic testes of all strains examined, we asked whether this ongoing proliferation is sufficient for teratoma formation in other strains (6). To test this, we examined post-natal testes for the occurrence of spontaneous testicular teratomas in each strain. While *Dazl*-deficient mice on a 129 strain background regularly produced teratomas (SI Appendix, Table S4), we did not observe testicular teratomas in any non-129 strain background (SI Appendix, Fig. S5F).

When these results are considered in conjunction with the retained *Nanog:GFP* positivity in the *Dazl*-deficient germline of B6 and 129S4 embryos (Fig. 2B), we conclude that the embryonic requirement for *Dazl* is consistent in all strains examined. Further, the extended proliferative activity and failure to restrict the expression of pluripotency markers in the *Dazl*-deficient germline reflects the retention of a PGC-like state on arrival at the gonad. In teratoma-resistant mouse strains, these PGC-like cells usually undertake apoptosis prior to birth. We note, however, that despite the abnormal germline development of the 129S4.*Dazl*-deficient germline, these cells may survive beyond birth, although mice are sterile with a failure to complete meiosis in both sexes.

SI Materials and Methods

Mouse models and maintenance. All experiments involving mice conformed to ethical principles and guidelines approved by the Committee on Animal Care at the Massachusetts Institute of Technology or the Cincinnati Children's Hospital Medical Center. Mice carrying the *129P2-Dazl*^{tm1Hjc} allele [*Dazl* mutant, RRID:IMSR_JAX:023802, (2)] were backcrossed to various *M. m. domesticus* strains, including C57BL/6NTac (B6, > 30 generations), FVB/NTac (FVB, > 30 generations), 129S4/SvJae (129S4, > 30 generations), 129S2/SvPasCrl (129S2, backcrossed from 129S4/SvJae for 10 generations), and the *M. m. molossinus* strain derived from wild-trapped Japanese mice, MSM/MsJ (MSM, backcrossed from B6 for 10 generations). The fluorescent Cre-reporter *Cg-Gt(ROSA)26Sor*$^{tm9(CAG-tdTomato)Hze}$ (*LSL-tdTomato*, RRID:IMSR_JAX:007909) allele (7) was maintained on a B6 background (Jackson Laboratory, Bar Harbor ME). The *CBA-Tg(Pou5f1-EGFP)*^{2Mnn} fluorescent reporter allele [*Oct4:EGFP*, RRID:IMSR_JAX:004654, (8)], present as a multi-copy transgene array near the telomere of chromosome 9 (9) was maintained on a B6 background. The *B6-Nanog*^{tm1Hoch} (*Nanog:GFP*, RRID:IMSR_JAX:016233) reporter allele (10) was backcrossed to B6 for seven generations, and to 129S4 for 10 generations. The Cre-recombinase *Mvh*$^{Cre\text{-}mOrange/+}$ [*Ddx4-Cre*, a null-allele (11)] was backcrossed to B6 and 129S4 for 15 generations. Mice carrying both *Sry*^{tm1} deletion (12) and *Tg(Sry)2Ei* transgene [referred to as *TgSry*, RRID:IMSR_JAX:010905, (13)], 129X1-*Bax*^{tm1Sjk} allele [RRID:IMSR_JAX:002994, (14)], and C57BL/6N-*Gcna*$^{tm1.1Dcp}$ allele [RRID:IMSR_JAX:031055, (15)] were each backcrossed to 129S4 for at least 10 generations. CD-1 outbred female mice were obtained from Charles River Laboratories (Wilmington MA).

Generation of Dazl reporter mice. We utilized a CRISPR/Cas9-mediated strategy to introduce a P2A-tdTomato sequence immediately 5' of the endogenous stop codon of *Dazl*. We designed a gRNA to a PAM site near the endogenous stop codon (in the 3'UTR of *Dazl*, target sequence: CTCTGCTAACTCATCTCAGG), and cloned this gRNA into the *BbsI* site of pX458 [also known as pSpCas9(BB)-2A-GFP, RRID:Addgene_48138]. We then performed pronuclear injection of this vector, together with an HDR template encoding *P2A-tdTomato* flanked by 2.89 kb upstream (including exon 10) and 1.98 kb downstream (including exon 11) of the endogenous *Dazl* locus, into one-cell embryos of an F1 B6D2 genetic background. Injected embryos were transferred to pseudopregnant CD-1 female mice. Correctly targeted mice carrying the *Dazl* reporter allele (B6D2-*Dazl*$^{em1(tdTomato)Huyc}$; referred to as *Dazl reporter*) were back-crossed for three generations to a B6 background.

Generation of Dazl conditional mice. We utilized a two-step CRISPR/Cas9-mediated strategy to insert *loxP* sites to introns flanking part of the RNA-recognition motif (RRM) of *Dazl*. We first designed gRNAs to target the intron between exons 3 and 4 (intron 3), and between exons 7 and 8 (intron 7; SI Appendix, Table S6). Targeting gRNA oligos were ligated into an expression plasmid pX330 (also known as pX330-U6-Chimeric_BB-CBh-hSpCas9, RRID:Addgene_42230). Cas9 mRNA and gRNAs were generated, and injected together with HDR templates containing *loxP* sequences into one-cell C57BL/6N embryos as previously described (16). After the first round of injections, a single mouse with a correctly targeted *loxP* in intron 7 was used to establish a homozygous colony. A second round of injections into one-cell embryos (homozygous for the *loxP* at intron 7) resulted in incorporation of an additional *loxP* site in intron 3 (the *Dazl* allele with two *loxP* insertions is named B6N-*Dazl*^{em1Dcp}, referred to as *Dazl-2L* in the text), confirmed by PCR and Sanger sequencing (SI Appendix, Fig. S7A). Mice were subsequently backcrossed five additional generations to a C57BL/6N background, and maintained as homozygotes (*Dazl-2L/2L*). We bred mice carrying this conditional allele with a germ cell-specific Cre [*Mvh*$^{Cre\text{-}mOrange}$, referred to as *Ddx4-Cre* in the text (11)] and observed germline recombination of the floxed allele in all Cre-carrying pups (B6N-*Dazl*$^{em1.1Dcp}$, referred to as *Dazl-1L* in the text). This recombined allele creates an out-of-frame mutation (*p.Tyr82Cys*), removing the 35 N-terminal amino acids of the RRM. To ensure that deletion of exons 4 through 7 phenocopied the established null allele [*129P2-Dazl*^{tm1Hjc} (2)], we intercrossed animals hemizygous for the recombined allele (*Dazl-1L/2L*) and analyzed the gonadal phenotype of homozygotes for germline survival on a B6 background (SI Appendix, Fig. S7B). We also back-crossed the *Dazl-1L* allele for five generations to the 129S4 background and assayed teratoma production (SI Appendix, Fig. S7D).

To generate conditional *Dazl* mice, we bred male mice carrying *Ddx4-Cre* and hemizygous for the recombined *Dazl* allele (B6.*Dazl-1L/+;Mvh*$^{+/Cre\text{-}mOrange}$; referred to as B6.*Dazl-1L/+;Cre* in the text) with female mice homozygous for the conditional allele (B6.*Dazl-2L/2L*). Resulting progeny heterozygous for the null allele and the conditional allele of *Dazl*, together with the Cre-recombinase (genotyped as B6.*Dazl-1L/2L*;*Cre*), were considered to be conditional for *Dazl* (referred to as B6.*Dazl* cKO in the text).

Generation of DAZL-deficient pigs. All experiments involving pigs conformed to the standard procedures and protocols approved by International Center for Biotechnology (Mt Horeb WI, formerly MOFA Global). *DAZL*-deficient males were generated as previously described using TALEN-mediated gene editing in porcine fibroblasts (17). In this study, *DAZL*-deficient sows

were generated by transfection of XX-bearing Landrace fetal fibroblasts with transcription activator-like effector nucleases (TALENs) and an oligonucleotide homology directed repair (HDR) template specific to the *DAZL* locus (listed in SI Appendix, Table S6). Briefly, 500,000 cells were transfected with 2 μg of TALENs mRNA and 2 μM oligonucleotide HDR template, and then electroporated using the Neon Transfection system (Thermo Fisher Scientific): input voltage 1800 V; pulse width 20 ms; and pulse number 1. Electroporated cells were cultured for three days at 30° C, followed by dilution cloning and screening of fibroblast clones by PCR amplification of DNA flanking the HDR target site, using primers listed in SI Appendix, Table S6, followed by BamHI digest of the DNA. Candidate cells with at least one HDR allele (BamHI positive) were further characterized by direct sequencing and/or TOPO cloned (Thermo Fisher Scientific) and verified by Sanger sequencing. Cell colonies with bi-allelic frame-shift mutations were pooled and cloned by chromatin transfer under contract with MOFA Global under Animal Welfare Assurance no. A4520/01. Sows were genotyped by the same methods used for colony screening (SI Appendix, Table S3).

Mouse RNA-seq analysis. To define the transcriptional program in the mouse germline as PGCs colonize the nascent gonads, we mapped raw RNA-seq reads from populations of sorted germline cells from GSE41908 (18) to the mouse transcriptome (Ensembl 84) using kallisto (19), with the following options: --bias --single, -l 200 -s 20. Transcript-level estimated counts and transcripts per million (TPM) values were summed to the gene level. We identified genes differentially expressed between E9.5, 11.5 and E13.5 using DESeq2 (20), requiring expression of $\geq$1 TPM at any age for inclusion in our analysis.

To assess germ cell-specificity in embryonic mouse gonads, raw RNA-seq reads from Kit^{+}/Kit^{+} (control) and Kit^{W}/Kit^{Wv} (germ cell-depleted) gonads (21) at E14.5, and from *Dazl*-deficient and *Stra8*-deficient ovaries at E14.5 (GSE70361) were mapped as above, and transcript-level estimated counts were summed to the gene level. For each comparison, we used the edgeR package (22) to obtain gene-level counts per million (CPM) values, normalized using the trimmed mean of M-values (TMM) method. Germ cell specificity was calculated by subtracting the CPM of germ cell-depleted (Kit^{W}/Kit^{Wv}) gonads from control gonads, divided by the CPM in control gonads. Where this ratio is < 0, as is the case for genes that increase in CPM in germ cell-depleted gonads compared with controls, the germ cell-specificity was set to 0 as none of the expression in control gonads could be assigned to germ cells. A ratio of 1 indicates germ cell-specific expression.

To assess whether germ cell factors were expressed in *Dazl*-deficient germline cells, we collected Oct4:EGFP-positive cells by FACS from control and *Dazl*-deficient embryonic mice at E10.3 (1-3 tail somites) and E11.5 (17-18 tail somites); *n* = three embryos at each time point. RNA was extracted using a Quick-RNA MiniPrep kit (Zymo Research, Irvine CA). Reverse transcription was performed using the SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View CA), and sequencing performed using a HiSeq2000 (Illumina). Transcripts were mapped to the mouse transcriptome using kallisto, and TPM values summed to the gene level (as above). Differential gene expression was assessed using DESeq2 (20).

Human single cell RNA-seq analysis. To define the transcriptional program in the human germline as PGCs colonize the nascent gonads, we analyzed raw RNA-seq reads from single cells from GSE86146 (23). Cell barcodes and unique molecular identifiers (UMIs) from read 2 were extracted using the umis python package (24) (https://github.com/vals/umis). We required that the cell barcode be an exact match to one of the 96 possibilities outlined (23). Both read 1 and read 2 were then mapped to the human transcriptome using kallisto (19) to obtain a pseudobam file; this was then passed to the tagcounts function to obtain gene-level UMI counts. The human transcriptome was defined as the subset of transcripts annotated in the GENCODE v24 annotation, comprising the union of "GENCODE Basic" transcripts and transcripts with a Consensus Coding Sequence (CCDS) ID.

We used the Seurat R package (http://satijalab.org/seurat/) to normalize, filter, and cluster single cells on the basis of their expression profiles (25). We first filtered out genes expressed in fewer than 20 single cells, and cells that expressed fewer than 2,000 genes or $1x10^5$ UMIs. After these filters, 1,882 single cells were considered for downstream analysis. We used the NormalizeData function to log-normalize the gene-level counts. We used the FindVariableGenes function with a minimum log-expression value and dispersion of 2 to identify 406 highly variable genes. Data were then scaled using the ScaleData function, and the number of UMIs was regressed out of the data. Principal components analysis (PCA) was performed on the 406 highly variable genes, and a jack-straw analysis with 200 replicates revealed the top 19 principal components (PCs) to be significant. The FindClusters function was used on these 19 PCs to assign single cells to clusters. For the visualization of clusters, tSNE was run on the top 19 PCs, with do.fast = TRUE. TPM values for each gene, in each cell, were calculated by normalizing gene UMI counts to cell library counts, and multiplying by $1x10^7$ (as most libraries had fewer than $1x10^6$ UMIs).

Human embryonic cells were sexed by summing the TPM of all Y-linked genes and defining cells with a total Y-linked gene TPM of less than 10 as XX, and greater than 10 as XY. Where appropriate, we excluded somatic cells on the basis of their expression of *WT1*, and meiotic cells on the basis of their expression of *STRA8*. With these filters applied, the following populations of germline cells were considered for further analysis: (group a) 37 'migratory' XY cells in cluster 5 from week 4, (b) 47 XX cells in cluster 5 from week 5; (c) 175 'early gonadal' XY cells in cluster 1, 11 and 6, (d) 281 XX cells in clusters 2 and 5; (e) 313 'late gonadal' XY cells in cluster 0 derived from weeks 19-25, and (f) 220 XX cells in clusters 7, 10 and 13, derived from weeks 11-26 (SI Appendix, Fig. S1A).

We used the SCDE R package (http://hms-dbmi.github.io/scde/) to analyze differential expression (26) between XY cells in group (a) and (c), and separately, between XX cells in (b) and (d). *P* values for 44 orthologous genes significantly upregulated in mouse were extracted and corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure.

To assess germ cell specificity in humans, we first defined four populations of cells: (group g) XY germ cells (clusters 1, 11 and 6, *WT1*-negative), (h) XY somatic cells (cluster 4, *WT1*-positive), (i) XX germ cells (clusters 2, 5, 7, 10, and 13, *WT1*-negative), and (j) XX somatic cells (cluster 3, *WT1*-positive). We used the scde.posteriors function from SCDE to obtain the maximum likelihood estimate (MLE) of gene expression in each cluster. For a given gene, XY germ cell specificity was defined as: (MLE in XY germ cells; group g) / (MLE in XY germ cells; group g, + MLE in XY somatic cells; group h), and XX germ cell specificity was defined as: (MLE in XX germ cells; group i) / (MLE in XX germ cells; group i, + MLE in XX somatic cells; group j). A ratio of 1 indicates germ cell-specific expression, a ratio of 0 indicates somatic cell-specific expression and/or not detected in germ cells. Custom scripts for scRNA-seq are available at https://github.com/snaqvi1990/pgc_scrnaseq/.

Cross-species single gene analysis of testis specificity. To determine gonad-specific gene expression in tetrapods, we analyzed published RNA-seq datasets from nine adult tissues (heart, muscle, kidney, liver, lung, colon, brain, spleen, and testis) isolated from seven species. For rhesus macaque, mouse, rat, bull, and chicken, raw reads were analyzed from GSE41637 (27), and for frog (*Xenopus laevis*), raw reads were analyzed from GSE73419 (28). For human, raw reads were obtained from the Genotype-Tissue Expression Consortium [GTEx, midpoint v6 release, dbGap accession phs000424.v6.p1, (29)]. For each of the corresponding nine adult tissues in human, we selected the 10 male samples with the highest RNA Integrity Number (RIN; sample IDs are provided in SI Appendix, Table S5). We used Ensembl 84 transcriptome

annotations for mouse, rat, and bull; Ensembl 88 annotation for rhesus (as they were derived from an improved genome assembly); GENCODE v24 annotation for human (as described above); and Xenbase v9.1 primary transcript assemblies for frog. Transcript-level TPM values were obtained using kallisto (19) in paired-end mode with the --bias option enabled, summed to the gene level, and summarized by the median TPM across all individuals in each tissue. Amniote orthology relationships were obtained from Ensembl Biomart, and human-frog orthologies were obtained from Xenbase. In all analyses, we required one-to-one orthology between human and mouse, but for comparisons of testis specificity, we allowed one-to-many orthology relationships, averaging testis specificity values in species where multiple orthologs were present.

Gonad specificity was calculated as previously described by dividing the testis TPM value by the sum of TPM values across all nine adult tissues, and repeated for each species (30). A ratio of 1 indicates gonad-specific expression; a ratio of 0 indicates no gonadal expression. Where no orthologous gene exists, or there is no annotation in the reference genome, the label 'no data' was applied (see Fig. 1D and SI Appendix, Fig. S2A). Violin plots were generated using the 'vioplot' package (https://CRAN.R-project.org/package=vioplot) and significance determined by Wilcoxon rank-sum test in R (https://www.r-project.org/).

Cross-species gene set analysis of testis specificity. The curated set of PGC factors (set iii, SI Appendix, Fig. S2A-D) was identified by survey of the literature. For each factor, we required that the gene be necessary for PGC or PGC-like cell development from pluripotent stem cell lines in either mouse or human, or has been reported as a marker of PGCs in any mammal *in vivo*.

To identify factors up-regulated on derivation of mouse PGC-like cells, we analyzed microarray data from GSE61924 (31). Gene expression was analyzed with NCBI's GEO2R tool, comparing day four cytokine-induced PGC-like cells to 2i + LIF-cultured mESCs. For human, raw reads from GSE60138 (32) were obtained and mapped to the above-described version of the human transcriptome using kallisto, with the following options: --bias -s 20 -l 200 --single. Transcript-level estimated counts and TPM values were summed to the gene level, and DESeq2 was used to perform differentially expression analysis between day four PGCLCs and day two pre-induced hESCs. Genes were required to be expressed >1 TPM in either condition to be considered for differential expression.

To identify genes commonly upregulated upon PGC-like cell induction in both human and mouse (set iv, SI Appendix, Fig. S2D), we used a procedure analogous to that used to identify genes induced on PGC colonization of the gonad. First, 101 genes upregulated in mouse (Benjamini-Hochberg-adjusted P value < 0.05, $\log_2$ fold-change > 1) were identified, of which 62

had one-to-one orthologs in human. Next, these 62 genes were tested for differential expression upon human PGC-like cell derivation as described above; of these 62 genes, 23 were also upregulated in human with adjusted *P* value < 0.05 (we did not require a $\log_2$ fold-change cutoff in humans as no such criteria were applied when delineating the gonadal PGC program).

Random sampling gene set analysis of testis specificity. To compare the testis specificity of the set of 13 genes commonly activated on PGC colonization to a cohort of unbiased gene sets (expressed at similar levels in both human and mouse PGCs), we sampled 500,000 random sets of 13 genes, obtained from the list of genes expressed at >1 TPM in both human and mouse migratory PGCs. To ensure equivalent expression levels in the germline, we recorded median testis specificity only if the set's expression levels in migratory PGCs did not differ significantly (two-sided Wilcoxon rank-sum test, *P* value > 0.1) from the set of 13 genes activated on PGC colonization (after arrival at the gonad). Using these criteria, 86,550 gene sets were identified for comparison of median testis specificity.

Analysis of pluripotency network expression in the germline. Human pluripotency factors were collated from the union of two analyses of naïve-like human ES cells (33, 34). Of this gene list, *CDHR1, DNMT3L*, *GCM1*, *LIPH, PRAP1*, *SOX2*, *ZFN57*, and *ZIC2* did not meet the criterion of TPM >10 in the human germline. Mouse 'naïve' and 'general' pluripotency factors were collated as previously described (35); *Klf4* and *Tbx3* did not meet the criterion of TPM >1 in sorted mouse germline cells; *Nr0b1* and *Klf2* were excluded due to expression in somatic lineages of the gonad.

Mouse genotyping. A small ear biopsy was taken prior to weaning. Genomic DNA was extracted in lysis buffer (100 mM Tris pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, and 100 μg/ml Proteinase K) at 65°C overnight. DNA was precipitated with an equal volume of isopropanol and centrifuged. The pellet was then washed in 70% v/v ethanol, centrifuged, and resuspended in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). Genotyping was performed using the PCR primers and conditions outlined in SI Appendix, Table S6, using Phusion DNA polymerase (New England Biolabs Inc, Ipswich MA).

Flow cytometry. Embryonic urogenital ridges or gonads (where appropriate) were dissected into phosphate buffered saline (PBS). A single-cell suspension was generated using trypsin, with 20 μg/ml DNase (Sigma, St Louis MO), in PBS. Cells were washed once with 20% v/v serum in

PBS, centrifuged at 500 *g*, resuspended in 1% v/v serum in PBS with DNase, and the suspension passed through a 40 μm filter. Oct4:EGFP-positive cells were collected using a FACSAria cell sorter (BD, Franklin Lakes NJ) into cell culture media. Expression from *Nanog:GFP*, *DAZL:tdTomato*, *LSL-tdTomato* or *Oct4:EGFP* reporters in germline cells was detected with a LSRFortessa cell analyzer (BD), and cytometry data analyzed with FlowJo software (v10.2, FlowJo, LLC., Ashland OR).

EG cell derivation. We used a modified EG derivation procedure (36). Briefly, Oct4:EGFP-positive cells were collected by flow cytometry and cultured in N2B27 medium, consisting of DMEM/F12 and Neurobasal, with N2 and B27 supplements, 0.1 mM nonessential amino acids, 2 mM L-Glutamine, 0.05% bovine albumin fraction V (Cat# 15260037), penicillin/streptomycin (each from Thermo Fisher Scientific, Waltham MA), and 0.1 mM 2-mercaptoethanol (Sigma-Aldrich). Cells were plated on 6-well culture dishes first coated with 10 μg/ml human plasma fibronectin for 1 h (Cat# FC010, EMD Millipore) on a monolayer of Sl220 feeder cells previously inactivated with mitomycin C (Cat# M4287, Sigma-Aldrich). For the first 48 h, cells were cultured in N2B27, with the addition of 20 ng/ml bFGF (Cat# PHG0261, Thermo Fisher Scientific), 50 ng/ml SCF (Cat# 455-MC, R&D systems), 1 μM ATRA (Cat# R2625, Sigma-Aldrich), the GSK inhibitor CHIR99021 at 3 μM (Cat# 4423, R&D systems) and 1,000 U/ml LIF (ESGRO, Cat# ESG1106, EMD Millipore) at 37°C in 5% CO_2 (EG derivation media). Cells were then subjected to half media changes every 48 h in N2B27, supplemented with 3 μM CHIR99021, the MEK inhibitor PD0325901 at 1 μM (Cat# 13034, Cayman Chemical, Ann Arbor MI) and 1,000 U/ml LIF (N2B27/2iLIF). Cultures were maintained in these conditions, and EG colony formation was monitored for 10 days, at which point the number of discrete colonies was counted and the efficiency of EG colony formation determined.

After 10 days, EG colonies were picked and cultured on CF6Neo MEF 7M MitoC feeder cells (Cat# GSC-6005M, MTI-GlobalStem, Gaithersburg MD) in DMEM/F12 supplemented with 10% fetal calf serum, 0.1 mM nonessential amino acids, 2mM L-Glutamine, 0.1 mM 2-mercaptoethanol and Penicillin/Streptomycin, with 3 μM CHIR99021, 1μM PD0325901 and 1,000 U/ml LIF. An early passage of EG cells were cryopreserved in 10% v/v DMSO in serum and stored under liquid nitrogen.

Histological analysis. Dissected tissues were fixed in 4% paraformaldehyde (PFA) or Bouin's solution for 16 h at 4°C. Tissues were washed in 70% ethanol, embedded in paraffin, and 5-μm sections cut for histology. Periodic acid-Schiff (PAS) staining was performed using a kit from

Sigma-Aldrich. Immunohistology was performed by dewaxing and dehydration of sections, followed by antigen retrieval in citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH6.0) by microwaving. Sections were then blocked in 5% donkey serum (Jackson ImmunoResearch Laboratories, West Grove PA), and incubated with primary antibodies (DAZL: #sc-27333, Santa Cruz Biotechnology, Dallas TX; DDX4: Cat# AF2030, RRID:AB_2277369, R&D Systems, Minneapolis MN; GCNA: a gift from George Enders; GFP: Cat# ab13970, RRID:AB_300798, Abcam, Cambridge, United Kingdom; RFP: Cat #600-401-379, RRID:AB_2209751, Rockland Immunochemicals Inc., Limerick PA; SOX9: Cat# AB5535, RRID:AB_2239761, EMD Millipore, Billerica MA). Sections were then washed in PBS, and later incubated with secondary antibodies (Jackson ImmunoResearch Laboratories). A glass coverslip was applied with ProLong Gold Antifade reagent with DAPI (Thermo Fisher Scientific), and immunohistology detected by confocal microscopy (Zeiss 700, Jena Germany). Meiotic spreads were prepared as previously described (37). 5-ethynyl-2'-deoxyuridine (EdU, Cat# A10044, Thermo Fisher Scientific) was diluted in sterile saline and injected into pregnant mice at 10 μg/g. After 3 h, embryos were dissected and histological sections prepared (as above), and EdU incorporation detected using a Click-iT imaging kit (Cat# C10337, Thermo Fisher Scientific).

Blastocyst injections. Chimeras were generated by injecting 129SB6F1.*Dazl*$^{tm1Hjc/tm1Hjc}$; *CBA-Tg(Pou5f1-EGFP)*^{2Mnn} EG cells derived from E15.5 testes (resulting from a 129S4 x B6 cross) into blastocysts from wildtype C57BL/6NTac embryos. Blastocysts were implanted into pseudopregnant CD-1 female mice and carried to term. To assess chimerism, various tissues were dissected from chimeric mice and genotyped for the *Dazl*^{tm1Hjc} and *CBA-Tg(Pou5f1-EGFP)*^{2Mnn} alleles (see SI Appendix, Table S6). To determine if EG cells contributed to the germline, gonads were dissected and prepared for flow cytometric detection of the *Oct4:EGFP* transgene (expressed by undifferentiated spermatogonia, Fig. S4H). *Dazl*'s essential role in spermatogenesis precludes an analysis of germline transmission (SI Appendix, SI Discussion).

Teratoma formation. To assess teratoma formation, animals were dissected at the age indicated in the text, and gonads isolated and fixed in either PFA or Bouin's fixative. Each gonad was prepared for histological analysis, and teratoma formation confirmed by the presence of derivatives from each somatic germ layer. A clinically trained pathologist examined 20 mouse gonadal tumors, and independently confirmed that each contained differentiated lineages of all three germ layers. Some teratomas in four-week-old animals were composed primarily of

neuroepithelium, consistent with the histology of immature teratomas. We did not identify tumors at extra-gonadal locations in any of our mice.

Detection of SNPs in 129S2 and 129S4 mice, and genotyping of teratomas. Genomic DNA was extracted from wildtype 129S2 and 129S4 mice and hybridized to the Mouse Diversity Genotyping Array (Cat# 901615, Thermo Fisher Scientific). We visually inspected normalized summary values for all variants to identify homozygous variants that differentiated the 129S2 and 129S4 substrains. We identified 411 SNPs that met our criteria, predominantly located on Chromosomes 1, 3 and 4 (204, 36 and 132 SNPs, respectively, Dataset S4). We then validated 11 of these SNPs that differentiate between 129S2 and 129S4 mice by Sanger sequencing (SI Appendix, Table S2).

To determine the cellular origin of spontaneous gonadal teratomas, genomic DNA was extracted from multiple pieces of each gonadal teratoma from 129SF1.*Dazl*-deficient mice, in parallel with host DNA (ear sample). For each sample, a PCR was performed at 11 discriminative loci (using the primers outlined in SI Appendix, Table S2), and analyzed for the presence of SNPs originating from 129S2 and 129S4 strains by PCR using Phusion DNA polymerase, followed by Sanger sequencing, and visualized using SnapGene Viewer software (v3.0.2, GSL Biotech, LLC., Chicago IL).

Calculation of mammalian evolutionary relationships. The evolutionary relationship between mammalian species was calculated using TimeTree (38), which estimates the timing of speciation based on a synthesis of the published literature.

Statistical analysis. Where applicable, the efficiency of EG cell derivation was compared using a t-test (Prism, v10.2 GraphPad Software Inc., La Jolla, CA). Where no EG cells were derived, a Fisher's exact test was employed (Prism). Statistical tests for the incidence of teratoma formation were performed using a Fisher's exact test using a two-tailed distribution (Prism). For all experiments, a P value of ≤ 0.05 was used as a measure of statistical significance. Statistical tests for RNA-seq data are described above.

Data availability. Data generated using the Mouse Diversity Genotyping Array (Thermo Fisher Scientific) for the SNP genotyping of 129S2 and 129S4 substrains has been deposited at the Gene Expression Omnibus under accession number GSE87771. Data generated from control and *Dazl*-deficient germline cells at E10.3 and E11.5 have been deposited under the SRA BioProject accession number PRJNA434733. Raw reads from RNA-seq experiments were downloaded from

publically available datasets (GSE41637, GSE41908, GSE60138, GSE70361, GSE86146, GSE73419 and phs000424.v6.p1). Array data (GSE61924) was analyzed with NCBI's GEO2R tool. Subsequent re-analyses of RNAseq of the mouse and human germline can be found in Datasets S1-S3.

SI Figures

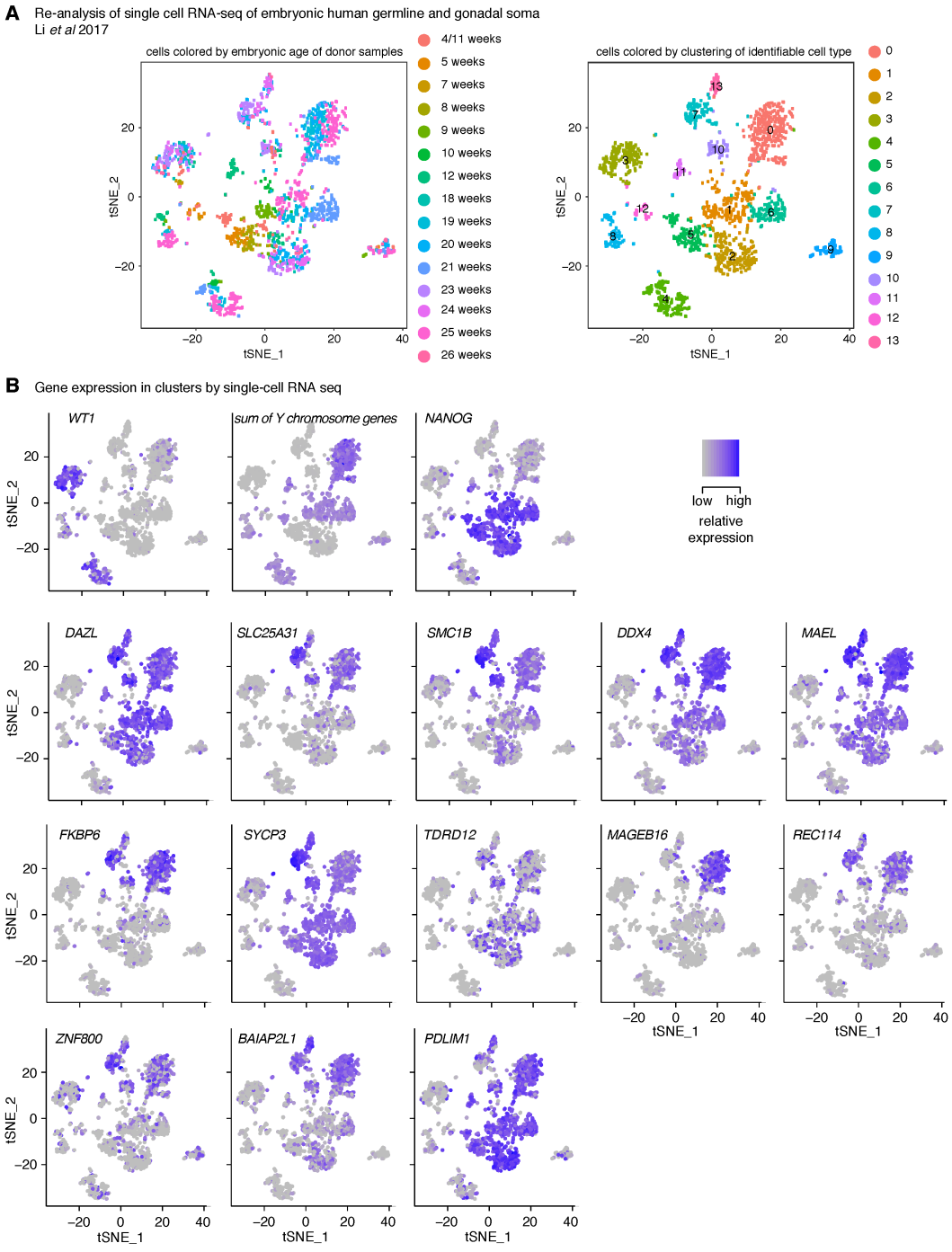


Fig. S1. Re-analysis of single cell RNA-seq from human embryos.

(*A*) tSNE plots of all cells collected from human embryos. Left: cells colored according to embryonic origin. Right: cells colored according to un-supervised clustering. Migratory

germline cells are found in cluster 5 (green). Male gonadal germline cells correspond to clusters 1, 11, and 6; and female gonadal germline cells to clusters 5 and 2. (*B*) Expression of cell type-specific marker genes (top panels, $n = 3$), and commonly induced genes (lower panels, $n = 13$), in human single-cell RNA-seq, according to clustering. Somatic cells visualized by *WT1* expression, male cells by summed expression of all Y-chromosome genes, and the pluripotency network expressed by migratory and newly gonadal germline cells illustrated by expression of *NANOG*.

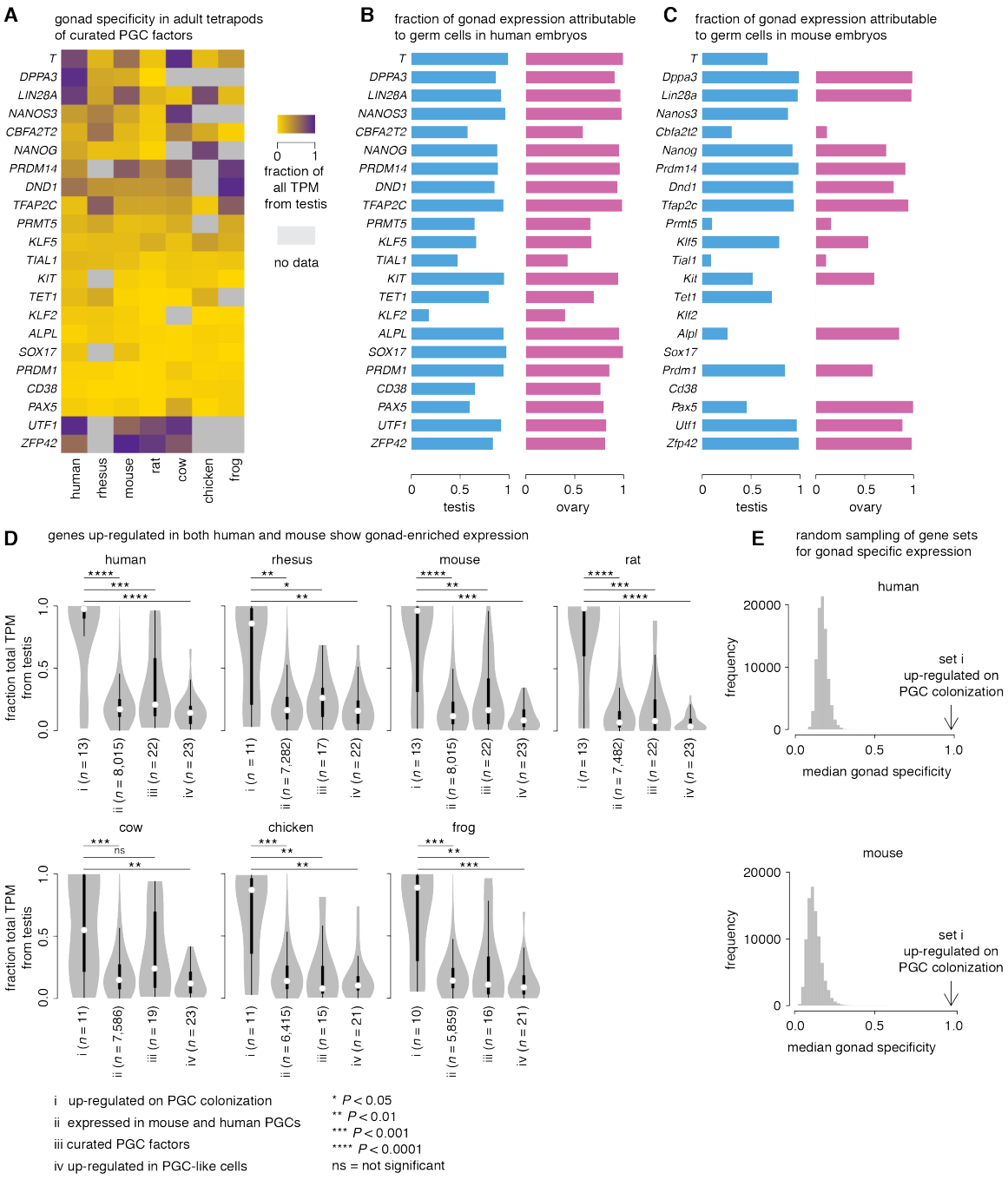


Fig. S2. The induced gonadal program distinguishes germ cells from other lineages.

(*A*) Heatmap; a curated set of PGC factors does not show significant enrichment for gonadal expression in tissues of adult tetrapods (n = 22). Specificity fraction is determined by calculating testis expression, divided by the sum of expression in all analyzed tissues, for each species. Genes with no annotated ortholog are shown in gray. (*B*) Germ cell expression of PGC factors in embryonic human testis and ovary, and (*C*) in E14.5 mouse testis and ovary by RNA-seq (see SI Materials and Methods). A ratio of 1 indicates germ cell-specific expression; 0 indicates somatic

cell expression. The curated set of PGC factors does not show clear germ cell specificity in the embryonic mouse. (*D*) Violin plots; black bar, interquartile range; circle, median value. Gonad-specific expression of the set of genes commonly upregulated in mouse and human in seven tetrapod species (left violin plot in each panel, set i, $n \leq 13$ orthologs), all genes commonly expressed in mouse and human migratory PGCs (set ii, $n \leq 8{,}015$), a curated set of PGCs factors (set iii, $n \leq 22$, see panel (*A*)), and factors expressed on PGC-like cell derivation (right, set iv, $n < 23$). * *P* value < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001, ns = not significant by Wilcoxon rank-sum test. (*E*) Mean gonad specificity of 500,000 randomly sampled sets of 13 genes expressed in human and mouse PGCs, of which 86,500 sets met expression criteria similar to set i. Of these, the greatest specificity achieved in human was 0.44 (compared with 0.98 for set i) and the greatest specificity achieved in mouse was 0.57 (compared with 0.96 for set i).

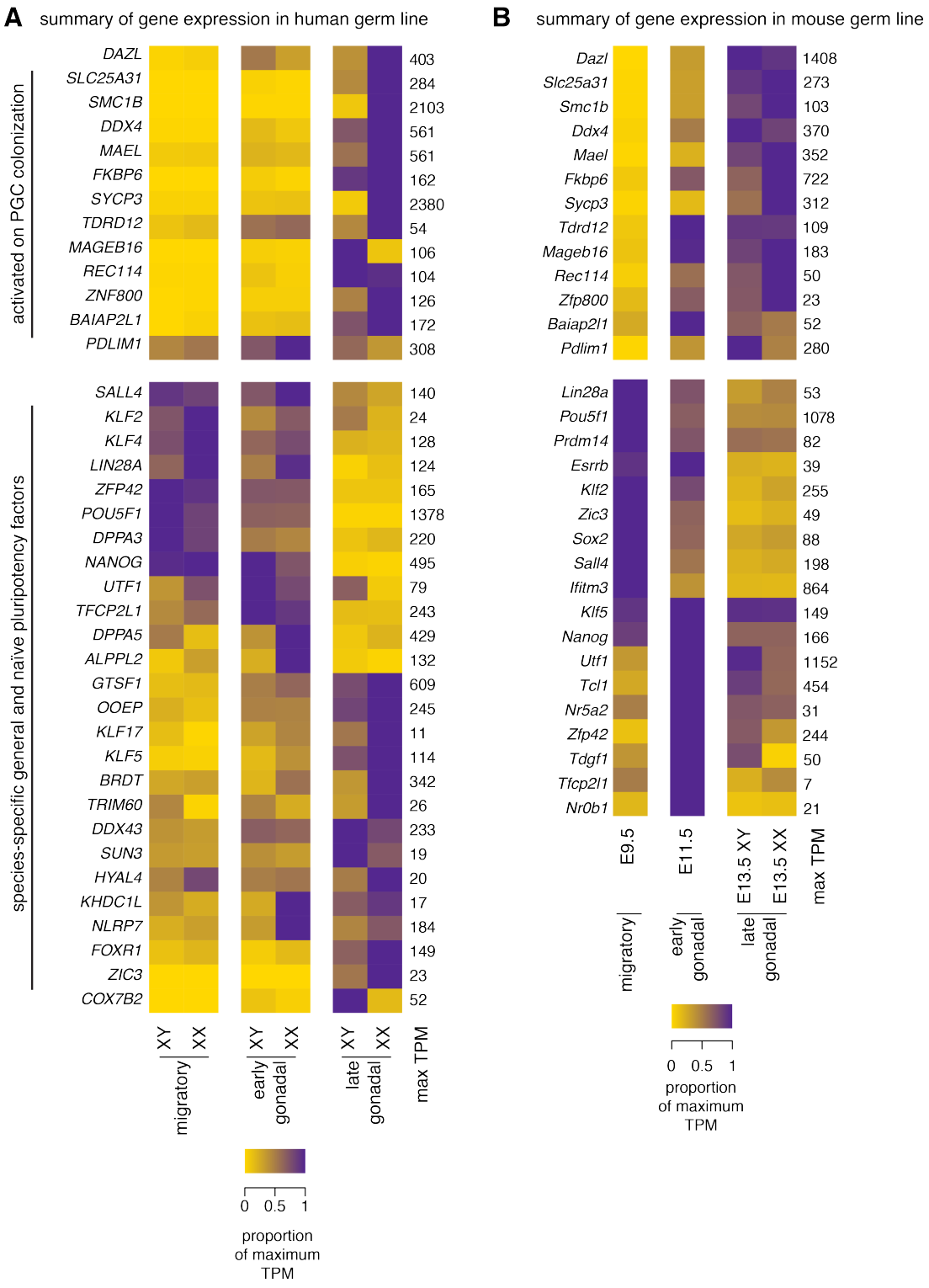


Fig. S3. Expression of genes activated on PGC colonization, and of general and naïve pluripotency factors in the mammalian germline.

(*A*, *B*) Heatmap; expression of factors commonly induced on PGC colonization (upper panel) and species-specific naïve and general pluripotency factors (lower panel) in migratory, early and late gonadal germline cells, derived from single-cell RNA-seq of human embryos (left, *A*) and RNA-seq of mouse embryos (right, *B*). Each gene's expression is shown as a proportion of its maximum expression. Gene order in upper panel as displayed in Fig. 1D, and in lower panel by hierarchical clustering for each species. TPM, transcripts per million.

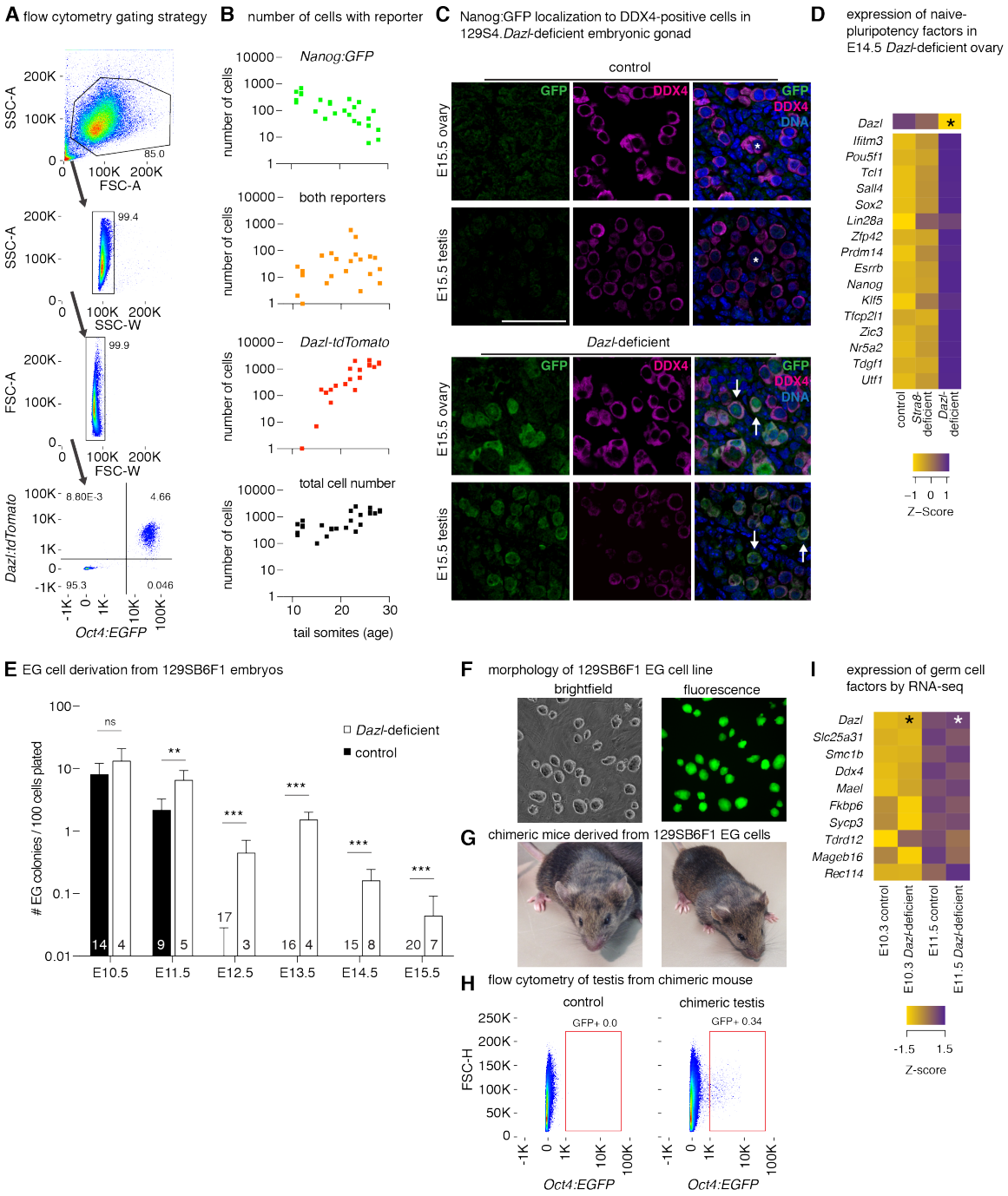
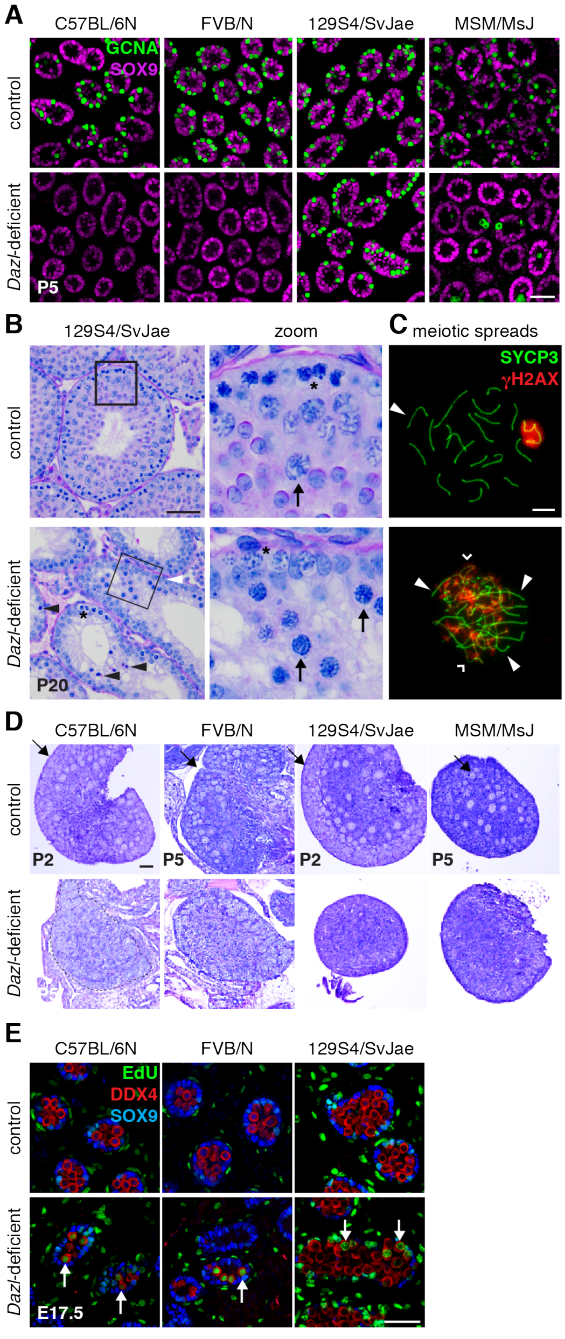


Fig. S4. *Dazl* is necessary for germ cell commitment in diverse strains of mice.

(*A*) Gating strategy for flow cytometric analysis and collection of germline cells from an embryo carrying both *Oct4:EGFP* and *Dazl:tdTomato* reporter alleles, collected at E12.5. Germline expression of DAZL:tdTomato in Oct4:EGFP-positive cells is presented in lower panel. (*B*) Flow cytometric counting of cells expressing *Nanog:GFP*, *Dazl-tdTomato*, or both reporters, from isolated urogenital ridges. Each point represents one embryo. Data was then averaged across

intervals of three tail somites, and presented in Fig. 2A. (*C*) Immunofluorescence of E15.5 ovary and testis in *Nanog:GFP* controls (upper panels) and *Nanog:GFP*; *Dazl*-deficient gonad (lower panels) in 129S4 embryo. Nanog:GFP-reporter (stained with an antibody to GFP, green) was expressed in germline cells of *Dazl*-deficient gonad (stained with DDX4, magenta, marked by arrows). Asterisks show GFP-negative germline cells. DNA stained with DAPI (blue). (Scale bar, 50 μm.) (*D*) Heatmap showing expression of naïve and general pluripotency factors in control, *Stra8*-deficient, and *Dazl*-deficient ovaries at E14.5. * denotes expression from exons retained in *Dazl* mutant allele. (*E*) Derivation of EG cells from 129SB6F1 control and *Dazl*-deficient embryos. Cells were collected by FACS at the embryonic age indicated on x-axis, and cultured under defined conditions. After 10 days, the number of EG cell colonies per 100 cells plated was counted, and the rate of EG cell derivation determined. The number of embryos tested is listed in each column, mean + SD, ** *P* value < 0.01, *** < 0.001, ns = not significant, using t-test or Fishers exact test as appropriate. (*F*) Morphology of EG cell line derived from E15.5 F1 *Dazl*-deficient testis cultured on feeder cells (left). Cells derived from PGCs carry *Oct4:EGFP* transgene, which is expressed in EG cell colonies (right). (*G*) Two chimeric mice at 30 days of age, derived from injection of E15.5 F1 *Dazl*-deficient EG cells (which contain the white-bellied Agouti coat color allele from the maternal 129S4 background) into B6 blastocysts. High contribution of EG cells to the epidermis can be observed by the brown (Agouti) coat color. (*H*) Flow cytometric analysis of control and chimeric testes for EG cell contribution to the germline, revealed by expression from the *Oct4:EGFP* transgene. (*I*) Heatmap of expression of germ cell factors (identified in Fig. 1D) from sorted control and *Dazl*-deficient migratory (E10.3, left) and gonadal (E11.5, right) germlines; $n = 3$ embryos at each time point. * denotes expression from exons retained in *Dazl* mutant allele.



F Teratoma incidence in *Dazl*-deficient males

	C57BL/6N	FVB/N	129S	MSM/MsJ
control	0 / 37	0 / 20	6 / 747 (0.80%)	0 / 46
Dazl-deficient	0 / 20	0 / 7	87 / 324 (26.8%)	0 / 23

Fig. S5: Requirement for *Dazl* in diverse strains of mice.

(*A*) Immunofluorescence of post-natal day five testes from control and *Dazl*-deficient mice of indicated strains. Germ cells stained with GCNA (green), and Sertoli cells with SOX9 (magenta). (Scale bar, 50 μm.) (*B*) Histology of control and 129S4.*Dazl*-deficient testis stained with periodic acid-Schiff at 20 days of age. Germ cells in control animals (upper panel) synchronously undertake spermatogenesis. In 129S4.*Dazl*-deficient mice, spermatogenesis is disrupted (lower panel): in these testes, spermatogonia at the basement membrane proliferate (asterisk) and commence spermatogenesis, but fail to progress beyond the zygotene stage (arrow). Germ cell death is widely observed in the 129S4.*Dazl*-deficient testis (open arrow-head). (Scale bar, 50 μm.) (*C*) Meiotic spreads of spermatocytes show asynapsis of some chromosomes in *Dazl*-deficient mice; these cells do not progress beyond the zygotene stage on a 129S4 background. (Scale bar, 5 μm.) (*D*) Histology of control and *Dazl*-deficient ovaries stained with periodic acid-Schiff. Primordial follicles in control ovaries are marked (arrow). (Scale bar, 50 μm.) (*E*) Immunofluorescence of embryonic day 17.5 testes from control and *Dazl*-deficient mice of indicated strains. Germline cells stained with DDX4 (red) and Sertoli cells with SOX9 (blue). EdU (green) marks DNA synthesis, a feature of mitotic activity. Extended proliferation of the *Dazl*-deficient germline is marked (arrow). (Scale bar, 50 μm.) (*F*) Incidence of testicular teratomas in control and *Dazl*-deficient mice in each strain background.

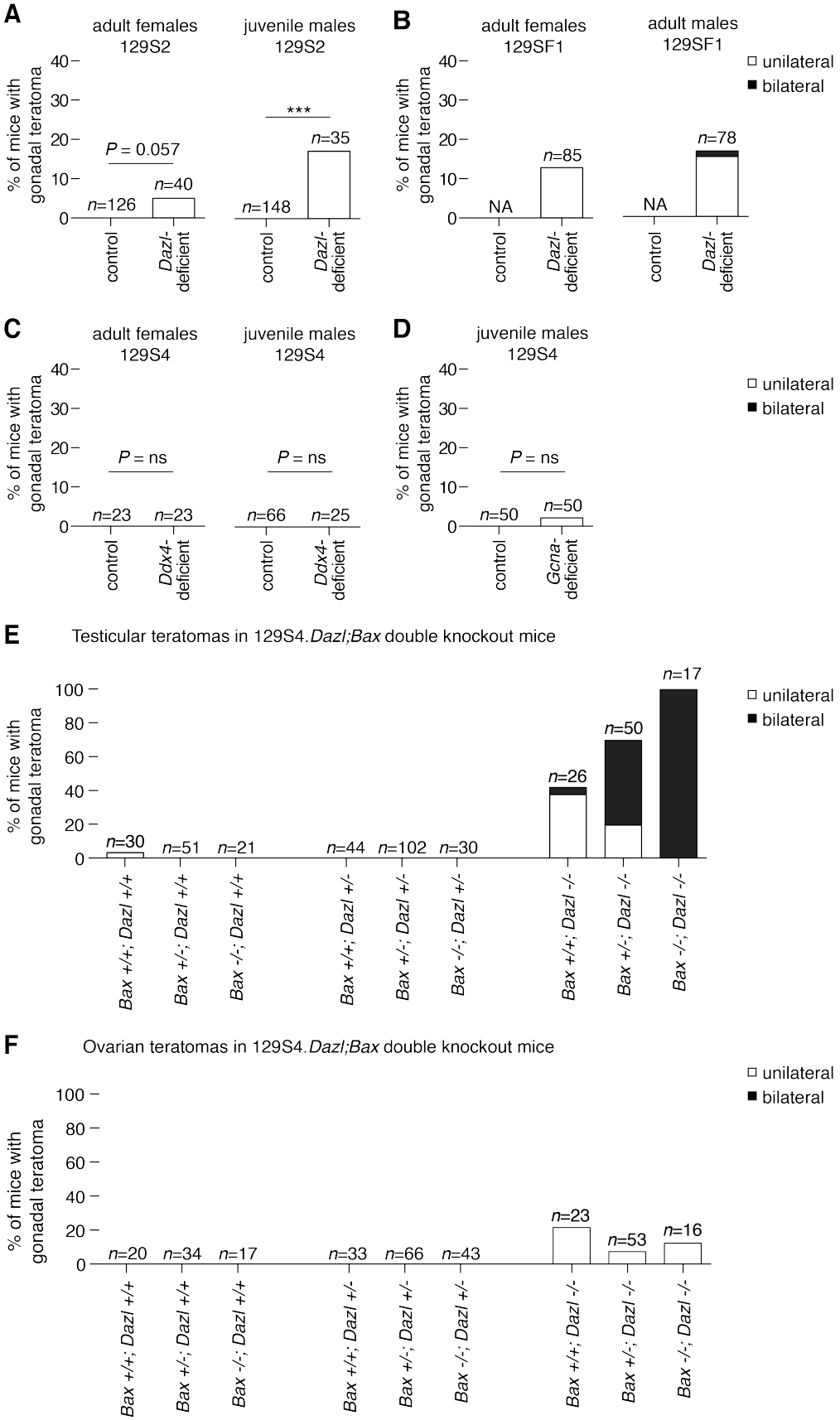

A
adult females 129S2
juvenile males 129S2
% of mice with gonadal teratoma
P = 0.057
n=126
n=40
***
n=148
n=35
control
Dazl-deficient
B
adult females 129SF1
adult males 129SF1
NA
n=85
NA
n=78
unilateral
bilateral
C
adult females 129S4
juvenile males 129S4
P = ns
n=23
n=23
n=66
n=25
Ddx4-deficient
D
juvenile males 129S4
n=50
n=50
Gcna-deficient
E
Testicular teratomas in 129S4.Dazl;Bax double knockout mice
n=30
n=51
n=21
n=44
n=102
n=30
n=26
n=50
n=17
Bax +/+; Dazl +/+
Bax +/-; Dazl +/+
Bax -/-; Dazl +/+
Bax +/+; Dazl +/-
Bax +/-; Dazl +/-
Bax -/-; Dazl +/-
Bax +/+; Dazl -/-
Bax +/-; Dazl -/-
Bax -/-; Dazl -/-
F
Ovarian teratomas in 129S4.Dazl;Bax double knockout mice
n=20
n=34
n=17
n=33
n=66
n=43
n=23
n=53
n=16

Fig. S6. Rate of spontaneous teratoma formation in mice.

(*A*) Incidence of gonadal teratomas in control and 129S2.*Dazl*-deficient mice. Females were dissected at two months of age, and males at four weeks of age. (*B*) Incidence of gonadal teratomas in 129SF1.*Dazl*-deficient mice. Females and males were dissected at three months of age. (*C*) Incidence of teratomas in control and 129S4.*Ddx4*-deficient mice. Females dissected at two months of age, and males at four weeks of age. (*D*) Incidence of testicular teratomas in *Gcna*-deficient mice. Males were dissected at 28 days of age. (*E*) Incidence of testicular teratomas in control, 129S4.*Dazl*- and *Bax*-deficient male mice, dissected at 28 days of age. (*F*) Incidence of ovarian teratomas in control, 129S4.*Dazl*- and *Bax*-deficient female mice, dissected at two months of age. *n* = number of animals examined, NA = not assessed, *** < 0.001 using Fisher's exact test.

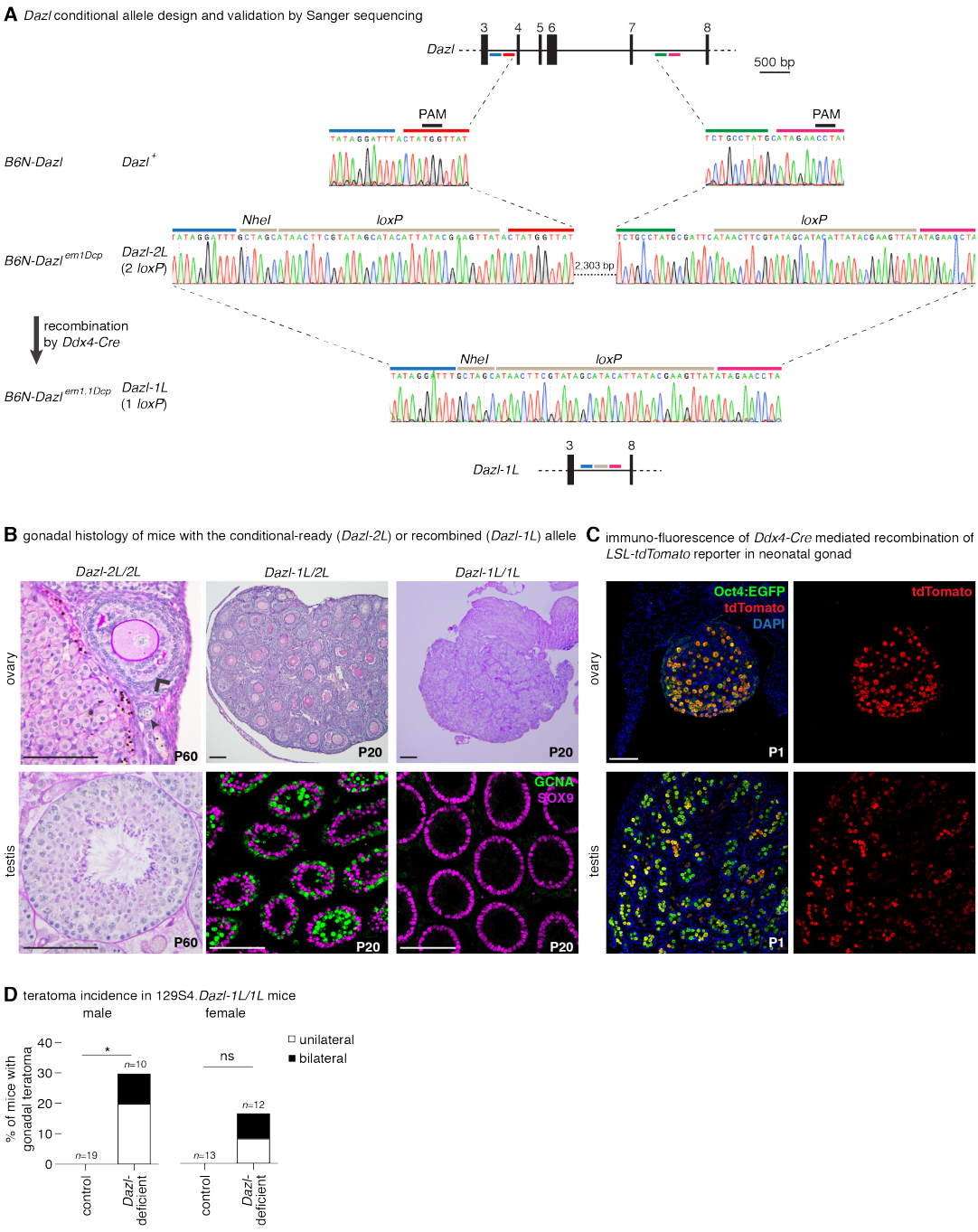


Fig. S7. Generation of conditional *Dazl* allele in mice.

(*A*) CRISPR/Cas9-mediated strategy to insert *loxP* sites to intron 3 and intron 7 of *B6N-Dazl* (upper panels). Correct targeting on a single allele confirmed by Sanger sequencing, and designated as B6N-*Dazl*^{em1Dcp} (carrying 2 *loxP* sites, referred to as *Dazl-2L*, central panels). This conditional allele was recombined by *Ddx4-Cre*, and designated as B6N-*Dazl*$^{em1.1Dcp}$ (containing a single *loxP* site, referred to as *Dazl-1L*, lower panel). When recombined, the *Dazl-1L* is predicted to introduce a frame-shift mutation (*p.Tyr82Cys*) resulting in a premature stop codon. (*B*)

Histology of mice homozygous for the conditional allele (B6.*Dazl-2L/2L*, left panels), show complete gametogenesis at two months of age in both sexes. The histology of mice zygotically homozygous for the recombined allele (B6.*Dazl-1L/1L*) show complete germline loss in both sexes (right panels), phenocopying the established allele (GCNA; green, SOX9; magenta). Primordial follicle (arrowhead) and secondary follicle (chevron) marked in ovary (***C***) Immunofluorescence of P1 ovary (left) and testis (right) in mice carrying *Ddx4-Cre;Oct4:EGFP*; and *LSL-tdTomato* alleles. tdTomato expression (stained with an antibody to tdTomato protein, red) is restricted to Oct4:EGFP-positive cells (stained with an antibody to EGFP protein, green). DNA stained with DAPI (blue). (Scale bar, 100 μm.) (***D***) Incidence of gonadal teratomas in control mice, or mice zygotically homozygous for the recombined *Dazl* allele (129S4.*Dazl-1L/1L*). Females were dissected at two months of age, and males between one and two months of age. *n* = number of animals examined, * *P* value < 0.05, ns = not significant, using Fishers exact test.

SI TABLES

Table S1: Efficiency of EG cell derivation, analyzed by sex.

Data from Fig. 2C: C57BL/6 background

	Control			*Dazl*-deficient		
Age	XX	XY	P value	XX	XY	P value
E10.5	15.2 n = 1	9.3 ± 2.8 n=7	NA	10.8 ± 2.2 n=2	10.9 n=1	NA
E11.5	0.92 ± 0.61 n = 7	1.8 ± 1.2 n=8	0.20	2.2 ± 1.1 n=3	7.0 ± 4.7 n=4	0.15
E12.5	0.02 ± 0.04 n=10	0.03 ± 0.06 n=8	0.82	0.34 ± 0.20 n=5	0.56 ± 0.40 n=3	0.32
E13.5	0.0 ± 0.0 n=13	0.0 ± 0.0 n=9	1#	0.19 ± 0.11 n=4	0.62 n=1	NA
E14.5	0.0 ± 0.0 n=2	0.0 ± 0.0 n=3	NA	0.13 n=1	0.12 n=1	NA
E15.5	0.0 ± 0.0 n=12	0.0 ± 0.0 n=5	1#	0.20 ± 0.09 n=2	0.20 n=1	NA

Data from Appendix SI, Fig. S3E: F1 genetic background

	Control			*Dazl*-deficient		
Age	XX	XY	P value	XX	XY	P value
E10.5	8.7 ± 4.6 n = 8	6.5 ± 2.6 n=6	0.29	11.4 n=1	13.7 ± 9.2 n=3	NA
E11.5	2.5 ± 1.3 n = 5	1.7 ± 0.61 n=4	0.28	7.6 ± 3.3 n=3	4.9 ± 0.19 n=2	NA
E12.5	0.08 ± 0.12 n=9	0.00 ± 0.00 n=8	0.08#	0.43 n=1	0.45 ± 0.37 n=2	NA
E13.5	0.0 ± 0.0 1.0 n=7	0.0 ± 0.0 n=9	1#	1.5 ± 0.50 n=2	1.6 ± 0.69 n=2	NA
E14.5	0.0 ± 0.0 1.0 n=7	0.0 ± 0.0 n=8	1#	0.08 + 0.10 n=4	0.20 ± 0.07 n=4	0.08
E15.5	0.0 ± 0.0 1.0 n=11	0.0 ± 0.0 1.0 n=9	1#	0.02 ± 0.02 n=4	0.06 ± 0.06 n=3	0.30

Mean ± SD
n = number of embryos tested
P value: t test, or Fisher's exact (# where appropriate).
NA: not assessed.

Table S2: SNP genotyping of teratomas in 129SF1.*Dazl*-deficient mice.

probeset_id	129S2	129S4	dbSNP_RS_ID	Chromosome	Position	Chrosmosome size (Mb)	Forward PCR primer 5'-3'	Reverse PCR primer 5'-3'
JAX00013233	AA	GG	rs31655561	chr1	180,384,510	195.47	ACACCACGTAGCGGCCATAA	TATCGCAAAGCGCAAAGCGT
JAX00113289	AA	CC	rs31338493	chr3	131,841,253	160.04	GAAGGACACGGGGTGAGAGA	CAAGGTGAACACATGCCCCA
JAX00118220	CC	TT	rs28269710	chr4	41,558,197	156.51	GGTCCTGCATGAAAGCTGGC	TAAAGCTGTCCAGAGGCCCA
JAX00577572	GG	AA	rs29507845	chr5	29,683,821	151.84	ACAGCAAAACTGCAGGGGTC	GTGGGAAGCAGGTTTCTGGG
JAX00616625	TT	GG	rs37217772	chr6	86,927,101	149.74	TAGCCTTAGAGAACCCAATGCCT	TCTCCAGAAGTCTCACTTTGCTCT
JAX00161035	CC	TT	rs3723238	chr8	41,323,781	129.4	TCCAGATTAGAAATTACAAGGGAGC	GCCCAAGGAGGCCTGAAAAA
JAX00171247	GG	AA	rs33739868	chr9	52,251,828	124.6	ACCACTTCGTGCTCTAACTCCA	CCGCAGAGGCAAGTGACTATT
JAX00287403	GG	CC	rs45770550	chr10	38,057,450	130.7	TTTTCTTATCCTCCCAGGGTAATGT	ATGCCTGTCAATGGGGGTAAC
JAX00316931	TT	CC	rs28193762	chr11	84,261,858	122.08	GCAGCAGGGTATGAACCGC	AAGCTGCAGCCTAAACACCAAA
JAX00352868	GG	AA	rs29955428	chr13	24,980,135	120.42	GTCATCAGCTCCCCTTTCGC	CCAGGTGGCCTGATTTGGTG
JAX00395455	TT	GG	rs32523635	chr15	22,469,829	104.04	TGGAGCAAGAAATTTGGAGTGCT	GCCAATTCCAATCTCACATCCCT

Table S3: Genotyping of TALEN-mediated deletions in pigs.

Animal ID	Mutation*	DAZL premature termination codon alleles	Left Ovary	Right Ovary
78	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	~20 lbs	atrophied
79	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	~30 lbs	atrophied
81	2 bp insertion//3 bp deletion, 2 bp insertion	67PTC/63PTC	atrophied	atrophied
82	Heterozygous RFLP, 10 bp deletion	63PTC/63PTC	~20 lbs	~3 inch diameter
83	Homozygous RFLP	63PTC/63PTC	atrophied	atrophied
84	Homozygous RFLP	63PTC/63PTC	atrophied	atrophied
85	Homozygous RFLP	63PTC/63PTC	atrophied	atrophied
89	Homozygous RFLP	63PTC/63PTC	atrophied	atrophied
90	2 bp insertion//3 bp deletion, 2 bp insertion	67PTC/63PTC	~20 lbs	atrophied
91	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	~3 inch diameter	~3 inch diameter
92	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	atrophied	atrophied
93	8 bp deletion//1 bp insertion	59PTC/61PTC	atrophied	~5 inch diameter
94	Heterozygous RFLP, 10 bp deletion	63PTC/63PTC	atrophied	~10 inch diameter
97	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	atrophied	~5 inch diameter
98	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	~5 inch diameter	atrophied
99	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	15.4 kg	atrophied
100	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	atrophied	~20 lbs

101	Heterozygous RFLP, 8 bp deletion	63PTC/59PTC	~5 inch diameter	~20 lbs
102	Heterozygous RFLP, 10 bp deletion	63PTC/63PTC	atrophied	atrophied
103	Heterzygous RFLP, 1bp insertion//2bp insertion	57PTC/65PTC	~20 lbs	atrophied

* Mutation to *DAZL* generated by TALEN-mediated homology-directed repair (HDR). The HDR template generates a novel BAM-HI restriction site, which can be assayed by restriction fragment length polymorphism (RFLP). Where mass is listed in pounds, or diameter listed in inches, are estimates. Also see W. Tan, *et al*., PNAS **110**, 16526—16531 (2013). PTC: Premature termination codon

Table S4: Aggregated numbers of mice with spontaneous gonadal teratomas.

Testicular teratomas		control mice		*Dazl*-deficient mice	
experiment	**related figure**	**# mice**	**# with teratoma**	**# mice**	**# with teratoma**
129S4 adult	Figure 3B	127	2	69	19
129S4 juvenile	Figure 3B	188	1	65	20
129S2	Figure S6A	148	0	35	6
129F1	Figure S6B	NA	NA	78	12
Sry^{tm1}*;SryTG;* XY male	Figure 4A	75	2	41	16
Ddx4	Figure S6C	66	0	NA	NA
Gcna	Figure S6D	50	0	NA	NA
Bax	Figure 4B Figure S6E	74	1	26	11
Dazl-1L	Figure S7D	19	0	10	3
total # animals		747	6	324	87
% of animals with teratoma			0.80		26.8

ovarian teratomas		control mice		*Dazl*-deficient mice	
experiment	**related figure**	**# mice**	**# with teratoma**	**# mice**	**# with teratoma**
129S4	Figure 3B	131	0	108	11
129S2	Figure S6A	126	0	40	2
129F1	Figure S6B	NA	NA	85	11
XX female	Figure 4A	80	0	32	4
Ddx4	Figure S6C	23	0	NA	NA
Bax	Figure 4B Figure S6F	53	0	23	5
Dazl-1L	Figure S7D	13	0	12	2
total # animals		426	0	300	35
% of animals with teratoma			0.00		11.7

Table S5: Sample IDs for human GTEx data.

Sample ID	Run accession #
GTEX-11DXY-3226-SM-5GIDE	SRR1373858
GTEX-11EQ9-0426-SM-5A5JY	SRR1324062
GTEX-11EQ9-0526-SM-5A5JZ	SRR1414185
GTEX-11EQ9-1926-SM-5PNVV	SRR1357367
GTEX-11EQ9-2126-SM-5PNVW	SRR1324390
GTEX-11GS4-2326-SM-5A5KS	SRR1469746
GTEX-11GS4-2526-SM-5A5KT	SRR1310735
GTEX-11GSO-2526-SM-5PNVX	SRR1400180
GTEX-11LCK-0126-SM-5A5M5	SRR1402109
GTEX-11NSD-0326-SM-5A5LS	SRR1442862
GTEX-11NUK-2926-SM-5A5MD	SRR1310008
GTEX-11OC5-0326-SM-5PNW5	SRR1385101
GTEX-11ONC-2426-SM-5GU7H	SRR1388832
GTEX-11TT1-1026-SM-5PNW7	SRR1404461
GTEX-11TT1-1626-SM-5EQL7	SRR1467500
GTEX-11TT1-2226-SM-5GU6B	SRR1455236
GTEX-12584-0526-SM-5FQTS	SRR1312406
GTEX-12696-0926-SM-5FQTV	SRR1433066
GTEX-12BJ1-1026-SM-5EGJA	SRR1337624
GTEX-13111-0126-SM-5EGHO	SRR1352639
GTEX-13112-2126-SM-5GCO4	SRR1329154
GTEX-132NY-0926-SM-5P9G3	SRR1407179
GTEX-132QS-1226-SM-5P9GD	SRR1476784
GTEX-1399R-1926-SM-5K7X8	SRR1358582
GTEX-139T6-0226-SM-5IFGZ	SRR1405245
GTEX-13FTW-1126-SM-5J2NV	SRR1412969
GTEX-13FTW-1326-SM-5LZZD	SRR1419518
GTEX-13O61-0526-SM-5J2M1	SRR1468595
GTEX-13O61-0726-SM-5J2MD	SRR1434671
GTEX-13OVL-1826-SM-5KLZR	SRR1340662
GTEX-145MG-3026-SM-5RQJA	SRR1486097
GTEX-145MO-0726-SM-5NQB9	SRR1397218
GTEX-147F4-2626-SM-5Q5CS	SRR1317086
GTEX-N7MS-0926-SM-2HMIZ	SRR607839
GTEX-N7MS-1626-SM-3LK5F	SRR809943
GTEX-N7MS-2425-SM-26GMD	SRR627425
GTEX-N7MS-2425-SM-26GMU	SRR627455
GTEX-NFK9-0126-SM-3LK5H	SRR820839
GTEX-NFK9-0626-SM-2HMIV	SRR598044

GTEX-NFK9-0926-SM-2HMJU	SRR602106
GTEX-NFK9-2026-SM-3LK5K	SRR808638
GTEX-NPJ8-0426-SM-2HMK6	SRR598148
GTEX-NPJ8-1226-SM-3MJHM	SRR810899
GTEX-OXRL-0526-SM-2I3EZ	SRR607586
GTEX-PVOW-2526-SM-2XCF7	SRR663753
GTEX-Q2AI-0526-SM-2I3EJ	SRR600417
GTEX-Q2AI-0926-SM-48U1F	SRR1099354
GTEX-Q2AI-1226-SM-48U14	SRR1078299
GTEX-QDVN-0526-SM-48TZ4	SRR1082447
GTEX-QDVN-1326-SM-48TZ3	SRR1085235
GTEX-QEG4-1826-SM-4R1JN	SRR1375721
GTEX-QESD-2026-SM-447BI	SRR1102152
GTEX-QMRM-0526-SM-2I5GA	SRR599025
GTEX-S33H-0126-SM-4AD62	SRR1101348
GTEX-S3XE-0426-SM-3K2AC	SRR659283
GTEX-S3XE-1126-SM-4AD4N	SRR1105033
GTEX-S3XE-1526-SM-4AD5A	SRR1094287
GTEX-S4Z8-0226-SM-4AD5K	SRR1073507
GTEX-S4Z8-0526-SM-4AD4T	SRR1087007
GTEX-SIU7-0526-SM-3NM8I	SRR809065
GTEX-TKQ2-1726-SM-4DXUP	SRR1120939
GTEX-U4B1-0326-SM-3DB8K	SRR656899
GTEX-UTHO-2426-SM-4JBHD	SRR1075102
GTEX-UTHO-3026-SM-3GAFB	SRR656745
GTEX-VJYA-0126-SM-4KL1P	SRR1079782
GTEX-VUSG-0126-SM-4KL1X	SRR1095913
GTEX-WH7G-0426-SM-3NMBJ	SRR815661
GTEX-WHWD-0126-SM-4OORS	SRR1370495
GTEX-WHWD-0426-SM-3LK83	SRR816064
GTEX-WHWD-1226-SM-4OOS1	SRR1323446
GTEX-WZTO-2926-SM-3NM9I	SRR820078
GTEX-XAJ8-0126-SM-47JYG	SRR1093196
GTEX-XAJ8-0426-SM-47JYJ	SRR1096924
GTEX-XBED-0826-SM-47JYC	SRR1098785
GTEX-XBED-1526-SM-4AT5W	SRR1081116
GTEX-XBED-2626-SM-4E3J5	SRR1096662
GTEX-XMK1-0126-SM-4B65F	SRR1075198
GTEX-XMK1-1726-SM-4B64Z	SRR1073217
GTEX-XMK1-2026-SM-4B65K	SRR1081663
GTEX-XOTO-0526-SM-4B662	SRR1068855

GTEX-XOTO-3026-SM-4B65M	SRR1083632
GTEX-XXEK-1926-SM-4BRVD	SRR1077066
GTEX-Y5V6-2026-SM-5IFHO	SRR1490658
GTEX-YEC4-0326-SM-4W216	SRR1481012
GTEX-YF7O-0626-SM-4W21R	SRR1472462
GTEX-ZPU1-0826-SM-57WG2	SRR1446203
GTEX-ZUA1-3026-SM-59HJC	SRR1334564

Table S6: Oligo and TALEN sequences used.

gene/allele	primers (5'-3')	product size (bp)
129P2-Dazl^{tm1Hjc}	S: AGCACTTAATTGGGCTTGCT	WT: 600
	AS: AAGCAGTGAAAGAATGCATGCTGA	
	S: GCTTCCTCTTGCAAAACCACACTGC	KO: 300
	AS: TGAAAGTACTCGTGCTGACTCAATTTC	
CBA-Tg(Pou5f1EGFP)^{2Mnn} *Oct4:EGFP*	TG-S: GGTCACCATCACTCAACTCC	WT: 456 TG: 217
	WT-S: TGAGGAAACATTACCAAAGGCACC	
	AS: GCCTGTAATCCCAGTCCTGA	
B6-Nanog^{tm1Hoch} *Nanog:GFP*	S: GAGAATAGGGGGTGGGTAGG	~500
	AS: CACCCCGGTGAACAGCTC	
Ube1x/Ube1y	S: TGGTCTGGACCCAAACGCTGTCCACA	*Ube1x:* 217 *Ube1y:* 198
	AS: GGCAGCAGCCATCACATAATCCAGATG	
Sry^{tm1} Presence of Y	S: CTGGAGCTCTACAGTGATGA	350
	AS: CAGTTACCAATCAACACATCAC	
Tg(Sry)2Ei Presence of Sry	S: AGCCCTACAGCCACATGATA	420
	AS: GTCTTGCCTGTATGTGATGG	
129X1-Bax^{tm1Sjk}	S: GAGCTGATCAGAACCATCATG	WT: 304
	AS: GTTGACCAGAGTGGCGTAGG	
	S: CCGCTTCCATTGCTCAGCGG	KO: 507
	AS: GTTGACCAGAGTGGCGTAGG	
Mvh$^{cre-mOrange}$	S: AGAGCACAGCCCCATAGTTGAAAGATAG	WT: 158 KI: 294
	WT AS: GAGGTTTGAGTATTTCTGCCTCCCAATC	
	KI AS: TTGCGAACCTCATCACTCGTTGC	
B6.Cg-Gt(ROSA)26Sor$^{tm9(CAG-tdTomato)Hze}$ *LSL-tdTomato*	S: AAGGGAGCTGCAGTGGAGTA	WT: 297
	AS: CCGAAAATCTGTGGGAAGTC	
	S: GGCATTAAAGCAGCGTATCC	KI: 196
	AS: CTGTTCCTGTACGGCATGG	
Dazl reporter		
sgRNA exon 11 oligos for pX458	S: CACCGCTCTGCTAACTCATCTCAGG AS: AAACCCTGAGATGAGTTAGCAGAGC	
B6D2-Dazl$^{em1(tdTomato)Huyc}$ *genotyping*	S: GGTGCATTTATTGTTGCAGAGTTC	WT: 516 KI: 240
	WT AS: GCCAATTTCAACTAAAAGTCCTTACGG	
	KI AS: TCAGCAGGCTGAAGTTAGTAGC	
Dazl conditional		
Intron 3-flox genotyping	S: TGGTGTGTCGAAGGGGTATG	WT: 581 *loxP*: 621
	AS: TCCAGGACTTACTCAACACGAA	
Intron 7-flox genotyping *Dazl*^{em1Dcp} *(Dazl-2L)*	S: AGCACTTAATTGGGCTTGCT	WT: 600 *loxP* (*2L*): 640
	AS: AAGCAGTGAAAGAATGCATGCTGA	
Dazl$^{em1.1Dcp}$ *(Dazl-1L)*	S: TCTTAAATTGTAGCTGTCTTGCCT	*1L*: 621 WT: 2,914
	AS: AAGCAGTGAAAGAATGCATGCTGA	
sgRNA intron 3 oligos for pX330	S: CACCGTGTTAATATAGGATTTACTA AS: TTTGATCATTTAGGATATAATTGTG	
sgRNA intron 7 oligos for pX330	S: CACCGATGCACTAGGTTCTATGCAT AS: TTTGTACGTATCTTGGATCACGTAG	
gRNA production intron 3	ttaatacgactcactatagGTGTTAATATAGGATTTACTA	
gRNA production intron 7	ttaatacgactcactatagGATGCACTAGGTTCTATGCAT	
HDR intron 3 *loxP*	TCAGTATTTTTATCACAATATATTGATCTTTCATTATCATTTTCTGTTAATATAGGATTTG CTAGCataacttcgtataGCATACATtatacgaagttatACTATGGTTATGATTTTATATTTCATTTGTTT TTCTAGCTATGGATTTGTCTCATTTTAT	
HDR intron 7 *loxP*	TTCACAAACTCATAAATCTCATAAAAATACTAAATTGGAAGCCATAATGTCTGCCTATG CGAATTCataacttcgtataGCATACATtatacgaagttatATAGAACCTAGTGCATGTTTGTGCAGGCC CCGTGTATGCTGCATCCATCCCTGTGAATTC	
Pig studies		
DAZL Genotyping of pigs	S: ATTTGGGCCCTGTTGAAAAC	314
	AS: ACTCACCCTTTGGACACACC	
HDR template Pig DAZL	AATTCTTCTCCATAGACGGATGAAACCGAAATTAGAAGTTGGATCCTTTGCTAGATATG GTTCAGTAAAAGGAGTGAAGATATTCACAGA	
TALENs	Left RVD: NN NN NI NG NN NI NI NI HD HD NN NI NI NI NG NG Right RVD: HD NG NG NG NG NI HD NG NN NI NI HD HD NI NG NI NG	

SI DATASETS

Dataset S1: RNA-seq analysis of mouse germline from E9.5 to E11.5.

Dataset S2: Genes up-regulated following PGC colonization of the nascent gonads in mouse, and human one-to-one orthologs.

Dataset S3: Single-cell RNA-seq analysis of human germline as PGCs colonize the nascent gonads.

Dataset S4: Identification of SNPs that distinguish between 129S2 and 129S4 substrains.

SI References

1. Lin Y, Page DC (2005) *Dazl* deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. *Dev Biol* 288(2):309–316.

2. Ruggiu M, et al. (1997) The mouse *Dazla* gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* 389(6646):73–77.

3. Schrans-Stassen BH, Saunders PT, Cooke HJ, de Rooij DG (2001) Nature of the spermatogenic arrest in *Dazl* -/- mice. *Biol Reprod* 65(3):771–776.

4. Saunders PTK, et al. (2003) Absence of *mDazl* produces a final block on germ cell development at meiosis. *Reproduction* 126(5):589–597.

5. Gill ME, Hu Y-C, Lin Y, Page DC (2011) Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. *Proc Natl Acad Sci USA* 108(18):7443–7448.

6. Salz HK, Dawson EP, Heaney JD (2017) Germ cell tumors: Insights from the *Drosophila* ovary and the mouse testis. *Mol Reprod Dev*. doi:10.1002/mrd.22779.

7. Madisen L, et al. (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* 13(1):133–140.

8. Szabó PE, Hübner K, Schöler H, Mann JR (2002) Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech Dev* 115(1-2):157–160.

9. Nicholls PK, Bellott DW, Cho T-J, Pyntikova T, Page DC (2019) Locating and Characterizing a Transgene Integration Site by Nanopore Sequencing. *G3 (Bethesda)* 9(5):1481–1486.

10. Maherali N, et al. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1(1):55–70.

11. Hu Y-C, de Rooij DG, Page DC (2013) Tumor suppressor gene *Rb* is required for self-renewal of spermatogonial stem cells in mice. *Proc Natl Acad Sci USA* 110(31):12685–12690.

12. Wang H, et al. (2013) TALEN-mediated editing of the mouse Y chromosome. *Nat Biotechnol* 31(6):530–532.

13. Washburn LL, Albrecht KH, Eicher EM (2001) C57BL/6J-*T*-associated sex reversal in mice is caused by reduced expression of a *Mus domesticus Sry* allele. *Genetics* 158(4):1675–1681.

14. Knudson CM, Tung KS, Tourtellotte WG, Brown GA, Korsmeyer SJ (1995) *Bax*-deficient mice with lymphoid hyperplasia and male germ cell death. *Science* 270(5233):96–99.

15. Carmell MA, et al. (2016) A widely employed germ cell marker is an ancient disordered protein with reproductive functions in diverse eukaryotes. *Elife* 5. doi:10.7554/eLife.19993.

16. Wang H, et al. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153(4):910–918.

17. Tan W, et al. (2013) Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proc Natl Acad Sci USA* 110(41):16526–16531.

18. Yamaguchi S, et al. (2013) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res* 23:329–339.

19. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34(5):525–527.

20. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

21. Soh YQS, et al. (2015) A Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet* 11(9):e1005531.

22. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

23. Li L, et al. (2017) Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell*. doi:10.1016/j.stem.2017.03.007.

24. Svensson V, et al. (2017) Power analysis of single-cell RNA-sequencing experiments. *Nat Methods* 14(4):381–387.

25. Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33(5):495–502.

26. Kharchenko PV, Silberstein L, Scadden DT (2014) Bayesian approach to single-cell differential expression analysis. *Nat Methods* 11(7):740–742.

27. Merkin J, Russell C, Chen P, Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338(6114):1593–1599.

28. Session AM, et al. (2016) Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* 538(7625):336–343.

29. GTEx Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235):648–660.

30. Julien P, et al. (2012) Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol* 10(5):e1001328.

31. Murakami K, et al. (2016) NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature* 529(7586):403–407.

32. Irie N, et al. (2014) *SOX17* Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell* 160:253–268.

33. Theunissen TW, et al. (2014) Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell* 15(4):471–487.

34. Guo G, et al. (2017) Epigenetic resetting of human pluripotency. *Development* 144(15):2748–2763.

35. Kalkan T, et al. (2017) Tracking the embryonic stem cell transition from ground state pluripotency. *Development*. doi:10.1242/dev.142711.

36. Leitch HG, et al. (2013) Rebuilding pluripotency from primordial germ cells. *Stem Cell Reports* 1(1):66–78.

37. Soh YQS, et al. (2017) *Meioc* maintains an extended meiotic prophase I in mice. *PLoS Genet* 13(4):e1006704.

38. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol* 32(4):835–845.