Supplemental Information

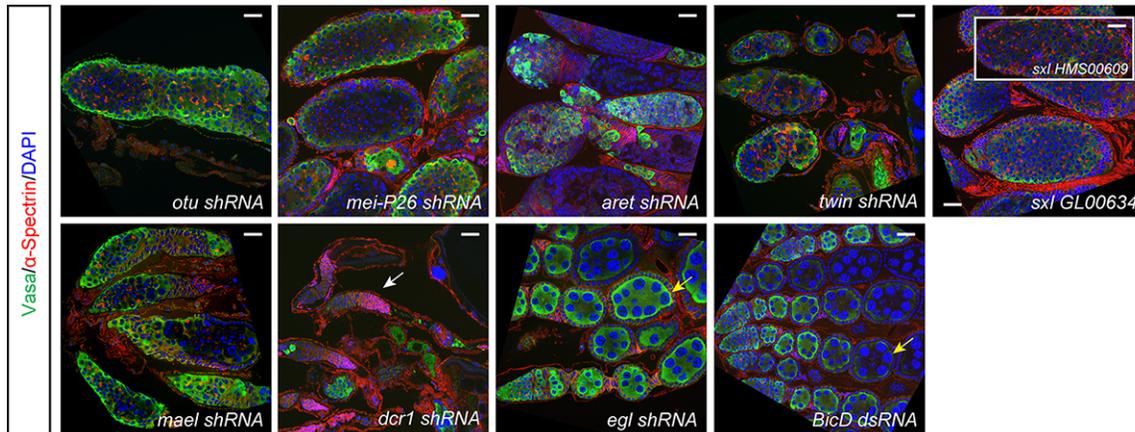# A Regulatory Network

# of *Drosophila* Germline

Stem Cell Self-Renewal

Dong Yan, Ralph A. Neumüller, Michael Buckner, Kathleen Ayers, Hua Li, Yanhui Hu, Donghui Yang-Zhou, Lei Pan, Xiaoxi Wang, Colleen Kelley, Arunachalam Vinayagam, Richard Binari, Sakara Randklev, Lizabeth A. Perkins, Ting Xie, Lynn Cooley, and Norbert Perrimon

Supplemental figures and table legends

Figure S1 (related to Figure 1). The germline screen identified many known essential regulators of GSC self-renewal and oogenesis

Ovaries expressing shRNAs targeting *otu, mei-P26, aret, twin, sxl, mael, dcr1* or *egl* by *MTD-Gal4* or *BicD* dsRNA by *UAS-dcr2; nanos-Gal4* stained for α-Spectrin, Vasa and DAPI. While *otu, mei-P26, aret, twin, sxl, mael* shRNAs block GSC differentiation, *dcr1* shRNA generates a stem-cell-loss phenotype (white arrow). In *egl* and *BicD* RNAi ovaries, oocytes do not form (yellow arrows) and there are 16 nurse cells instead. *Sxl* results are confirmed by two independent shRNAs; *MTD-Gal4/HMS00609* females are obtained by crossing *MTD-Gal4* males with *HMS00609* females. Scale bars: 20 μm.
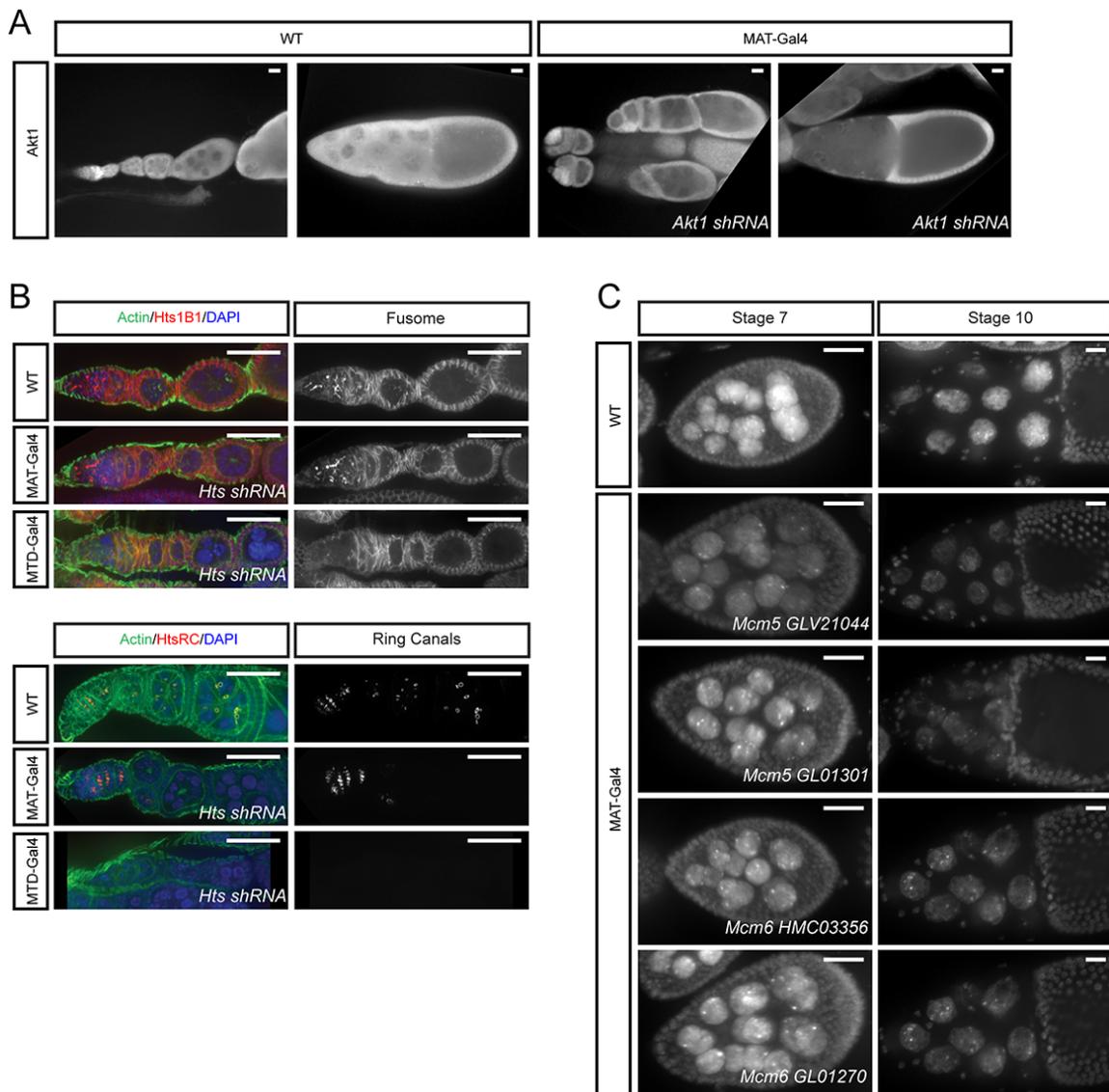
**Figure S2 (related to Figure 2). shRNA validation experiments**

(A) Expression of *Akt1 shRNA* by *MAT-Gal4* strongly reduces Akt1 protein in the germline after stage 1.

(B) Ovaries expressing *hts shRNA* by *MTD-Gal4* or *MAT-Gal4* were stained for Hts-1B1 or Hts-RC antibodies. In MAT ovaries, both the Hts-1B1 and Hts-RC proteins were present on fusomes and ring canals, respectively, in the germarium, but absent from egg chambers outside the germarium resulting in defective ring canals. In MTD ovaries, both

Hts-1B1 and Hts-RC were absent, and egg chambers showed the typical *hts* phenotype of too few nurse cells and defective ring canals.

(C) Knock down of Mcm5 and Mcm6 with two independent shRNAs driven by *MAT-Gal4* leads to ploidy defects in nurse cells of stage 7 and 10 egg chambers, as shown by DAPI staining.

Scale bars: 20 μm.
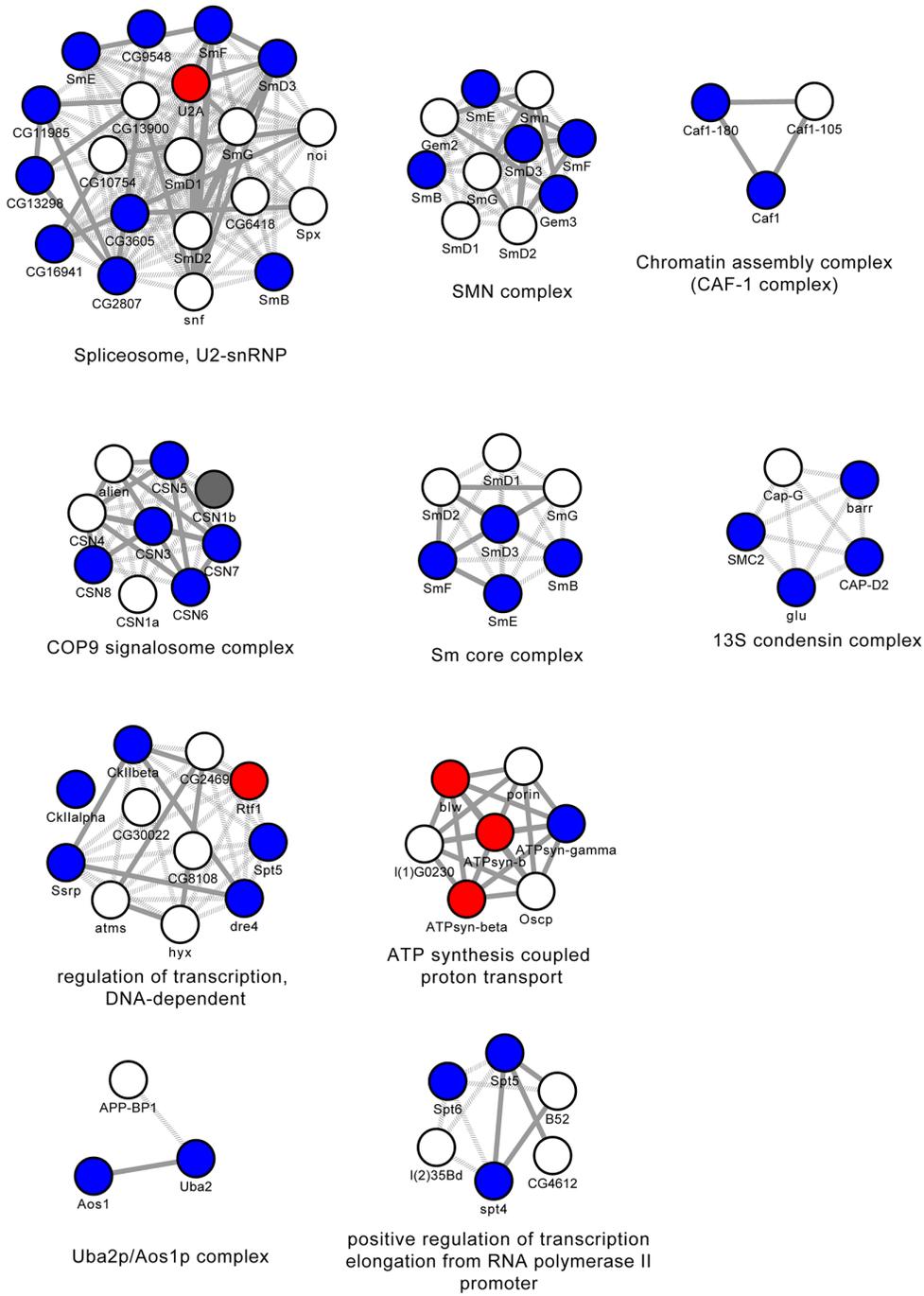
Figure S3 (related to Figure 3). Representative protein complexes identified using COMPLEAT

Red: differentiation defects; blue: GSC loss or agametic; white/grey: genes that were not screened or did not score. The full list of protein complexes are shown in Table S2.
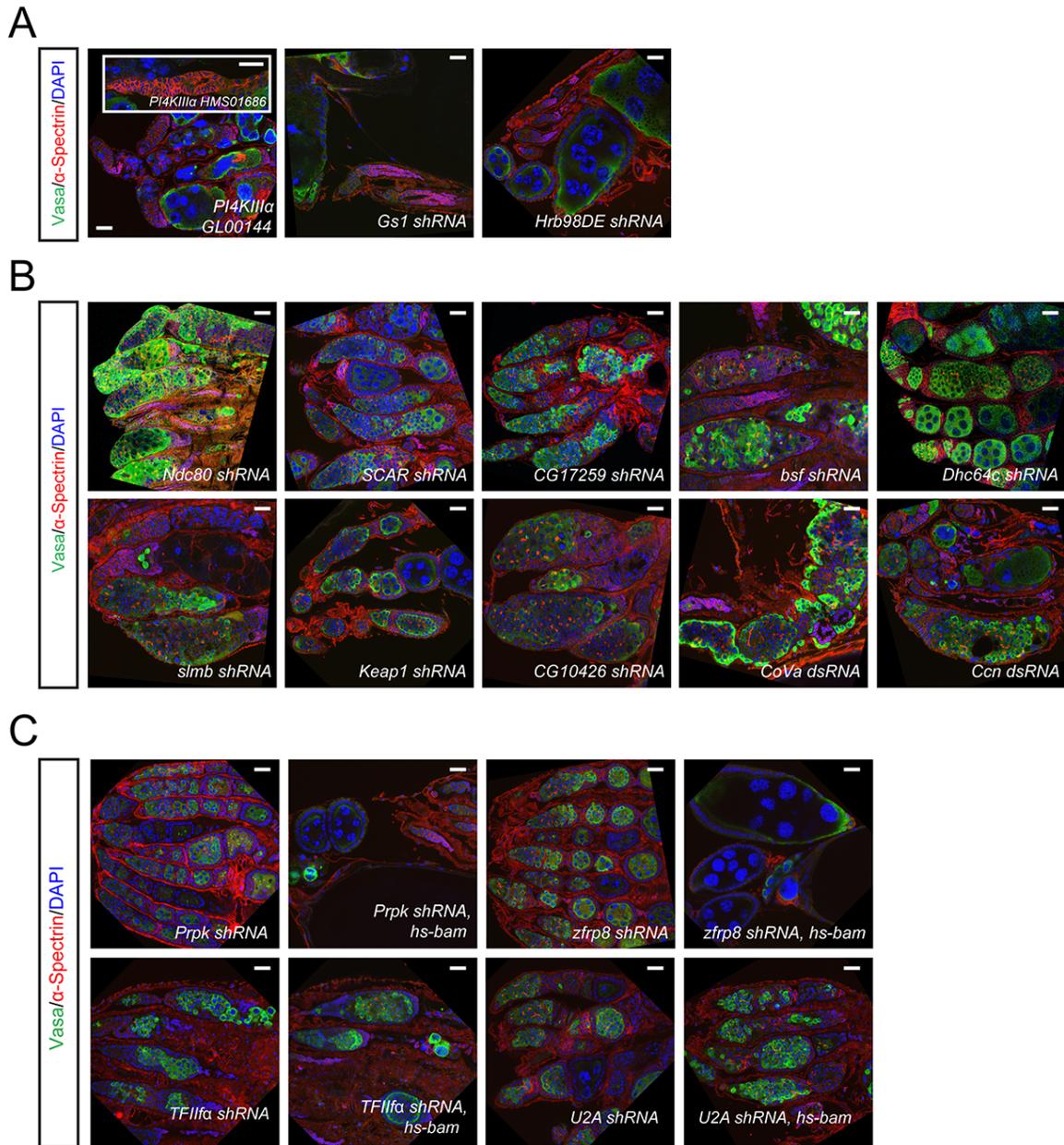
Figure S4 (related to Figure 4). Regulators for GSC differentiation or maintenance identified from the screen

(A-B) Ovaries expressing shRNAs or dsRNAs against indicated genes by *MTD-Gal4* or *UAS-dcr2; nanos-Gal4* were stained for α-Spectrin, Vasa and DAPI.

(A) *PI4KIIIα, Gs1* and *Hrb98DE* shRNAs generate an empty germarium and stem-cell-loss phenotype.

(B) *Ndc80, SCAR, CG17259, bsf, Dhc64c, slmb, Keap1, CG10426, CoVa* and *Ccn* RNAi block GSC differentiation.

(C) *bam* epistasis analysis. shRNAs against *Prpk, zfrp8, TFIIfα* or *U2A* expressed using *nanos-Gal4* with or without *hs-bam* expression. Ovaries are stained for α-Spectrin, Vasa and DAPI and quantification results are shown in the Figure 4E.

Scale bars: 20 μm.
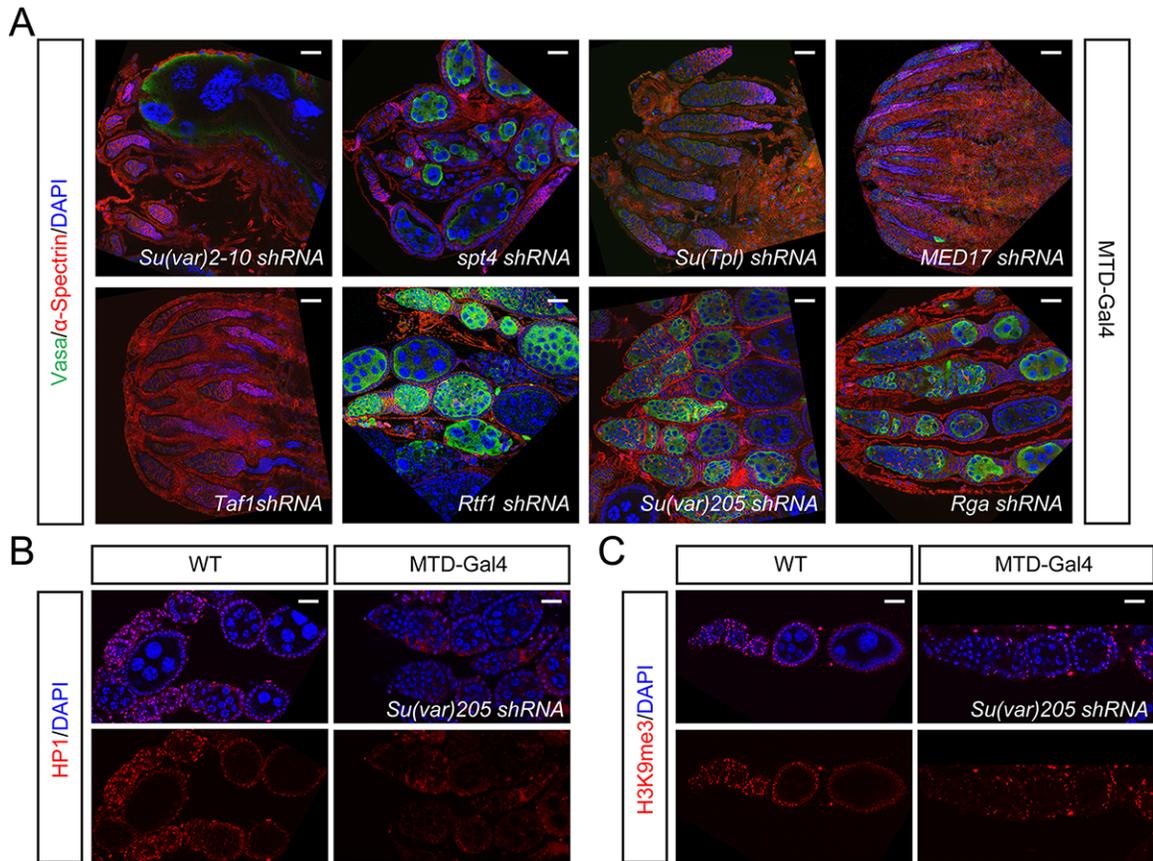
Figure S5 (related to Figure 5). Transcription factors required for GSC differentiation or maintenance

(A) Ovaries expressing shRNAs against indicated genes by *MTD-Gal4* were stained for α-Spectrin, Vasa and DAPI. *Su(var)2-10* and *spt4 shRNAs* generate an empty germarium and stem-cell-loss phenotype. *Su(Tpl), MED17* and *Taf1 shRNAs* result in an agametic phenotype. *Rtf1, Su(var)205* and *Rga* shRNAs lead to differentiation defects.

(B) Depletion of the Su(var)205 (HP1) protein in the germline upon *MTD-Gal4* driven expression of a *Su(var)205* targeting shRNA.

(C) Reduced levels of H3K9me3 in the germline upon depletion of *Su(var)205* by shRNA expression. Scale bars: 20 μm.
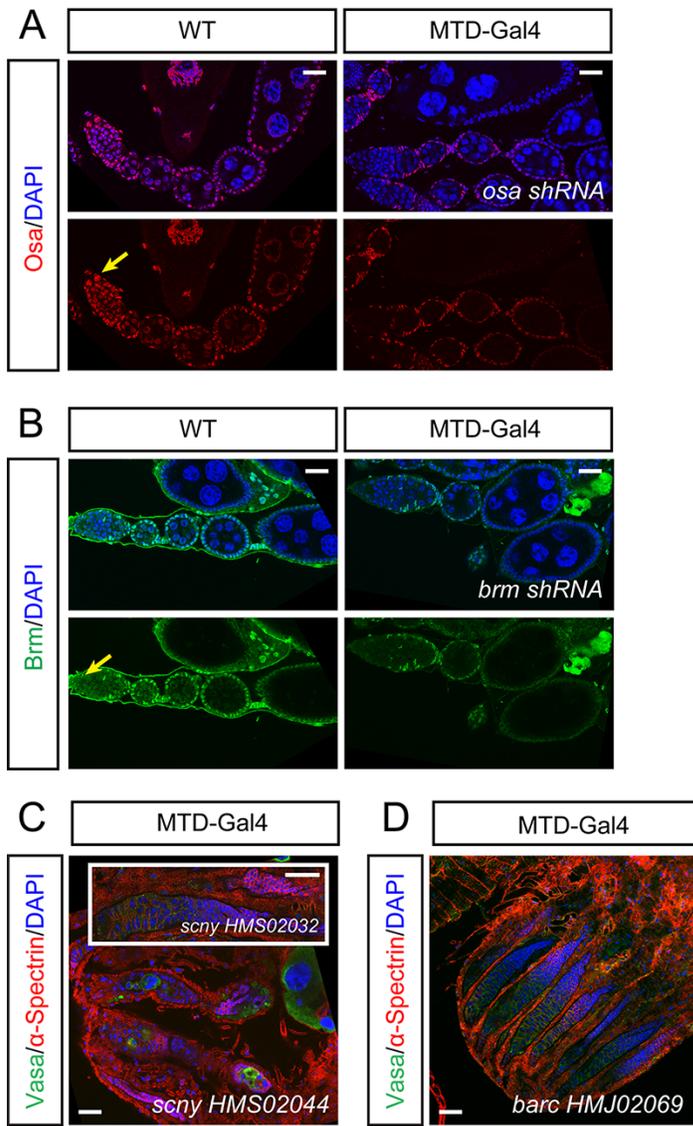
Figure S6 (related to Figure 6). Comparative analysis of self-renewal

(A,B) Depletion of the brahma complex members *osa* or *brm* does not induce a detectable phenotype in the germline. shRNAs are expressed by MTD-Gal4 and ovaries stained for Osa/DAPI (A) or Brm/DAPI (B). Yellow arrows indicate Osa and Brm expression in GSCs. (C) Expression of two independent shRNAs targeting *scny* by *MTD-Gal4* result in a depletion of germline cells. Ovaries are stained for α-Spectrin, Vasa and DAPI. (D) Knockdown of *barc* in the germline by *MTD-Gal4* results in an agametic phenotype. Scale bars: 20 μm.

A

Dm Set1   RRM, N-Set, Set, shRNA   1641

Hs Set1   1707

Sc Set1   1080

RRM
N-Set
Set
post-Set
shRNA

B

WT	MTD-Gal4

pH3/DAPI

Set1 HMS00581

C

MTD-Gal4

Vasa/α-Spectrin/DAPI

Set1 HMS01837   *Set1 HMS02179*

D

WT	MTD-Gal4

H3K4me1

Set1 HMS00581

H3K4me2

Set1 HMS00581

E

MTD-Gal4

H3K4me3/α-Spectrin/DAPI

Set1 HMS01837   *Set1 HMS02179*

F

bam

PolII

H3K4me3

21.069 Mb   21.0695 Mb   21.07 Mb   21.0705 Mb

dicer−1

PolII

H3K4me3

18.56 Mb   18.562 Mb   18.564 Mb   18.566 Mb

mei−P26

PolII

H3K4me3

9.058 Mb   9.06 Mb   9.062 Mb   9.064 Mb

spt6

PolII

H3K4me3

6.162 Mb   6.164 Mb   6.166 Mb   6.168 Mb

spt5

PolII

H3K4me3

15.325 Mb   15.326 Mb   15.327 Mb   15.328 Mb   15.329 Mb

put

PolII

H3K4me3

10.445 Mb   10.447 Mb   10.449 Mb   10.451 Mb

igru (CG11266)

PolII

H3K4me3

7.033 Mb   7.034 Mb   7.035 Mb   7.036 Mb   7.037 Mb

aos1

PolII

H3K4me3

8.2584 Mb   8.2588 Mb   8.2592 Mb   8.2596 Mb

mcm3

PolII

H3K4me3

5.218 Mb   5.2185 Mb   5.219 Mb   5.2195 Mb   5.22 Mb   5.2205 Mb

Figure S7 (related to Figure 7). Set1 is important for germline differentiation and GSC maintenence

(A) Schematic diagram showing domain structure of Set1 protein. RRM: RNA recognition motif; Dm: *Drosophila melanogaster*; Hs: *Homo sapiens*; Sc: *Saccharomyces cerevisiae*. The positions of three shRNAs are shown in the protein model.

(B) pH3 positive mitotic cells (yellow arrows) are restricted to the tip of the ovarioles in WT ovaries, but are detected throughout ovarioles in *MTD/Set1 shRNA* ovaries.

(C) Ovaries expressing *Set1 shRNA* (HMS01837 or HMS02179) by *MTD-Gal4* are labeled by α-Spectrin, Vasa and DAPI staining. White arrows: empty ovarioles, Yellow arrows: pseudo egg chambers filled with undifferentiated cells.

(D) H3K4me1 and H3K4me2 staining in WT and *Set1 shRNA/MTD-Gal4* ovaries.

(E) H3K4me3 depletion with *Set1 shRNA/MTD-Gal4* (HMS01837 and HMS02179).

(F) H3K4me3 and Pol II ChIP-seq experiment: Green (Pol II) and red (H3K4me3) tracks show enrichment on a select set of genes: those already known to affect the GSC lineage (*bam, dicer-1, mei-P26, spt6, put*) and additional regulators identified in this screen (*igru, spt5, aos1, mcm3*). Gene models are shown in black and grey represents whole cell extract.

Scale bars: 20 μm.

Table S1 (related to Figure 1). Germline screen results of 4608 TRiP transgenic lines and analysis of 366 genes that produce no/few eggs

Sheet 1: screen results of 4608 lines. RNAi lines that produce no/few eggs are labeled with blue color. The lines for which confocal images were recorded are labeled with "x". Each line is linked to the RSVP database, where the phenotypic data are stored. Sheet 2: screen results of 444 lines that produce no/few eggs. Among those lines, 174 agametic lines were re-screened with *MAT-Gal4* and the phenotypes are shown in the table. Those lines for which confocal images exist are marked with "x". According to the results of the two screens, all 444 lines were assigned to the respective phenotypic categories. dd: differentiation defects; sl: GSC loss; o: oogenesis defects; a: agametic; a, n: agametic (MAT normal); a, d: agametic (MAT defective). Sheet 3 contains the 'differentiation defective' hits with additional phenotypic information. Sheet 4: analysis of 366 genes that produce no/few eggs. For each gene the number of RNAi lines screened is shown. Complex analysis indicates whether this protein forms complexes with other hits in the screen. Based on these two criteria, a confidence score (from high to low) was given to each gene. Genes that also scored in the neuroblast screen are indicated. Finally, the DIOPT tool was used to find human orthologs of those genes.

Table S2 (related to Figure 3). Protein complexes containing genes identified from the GSC screen

Using the hits from the GSC screen, the COMPLEAT database identified 116 protein complexes. P-value, complex members and gene scored are shown in the table.

Supplemental Experimental Procedures:

RNAi screen and *Drosophila* strains

An important conclusion from our screen is that shRNA lines are more effective than dsRNA lines in the germline, although some dsRNA lines are also functional and generate expected phenotypes during early oogenesis. Only 4.6% of VALIUM1 and 5.0% of VALIUM10 lines resulted in detectable phenotypes, while 12.8% of VALIUM20 and 18.1% of VALIUM22 scored in our analysis (Figure 1C). VALIUM22 lines are slightly more effective than VALIUM20, consistent with its optimization for germline expression. A *pUASp-GFP* transgenic fly strain was established by injecting the *Drosophila* gateway vector 1077 in *white* mutant embryos using standard techniques. hs-bam experiment: larvae (48-72 hrs after egg laying) were incubated at 37 °C for 1 hr for three consecutive days. Females of the genotype *nanos-Gal4/+; shRNA/hs-bam* were dissected three days after eclosion. Siblings of *nanos-Gal4/+; shRNA/TM6B* were used as controls.

Bioinformatics analysis

1. Tissue expression heatmap (Figure 1F): The analysis was performed as described in Neumüller et al. (2011) with slight modifications: Microarray based expression data were retrieved from FlyAtlas (http://www.flyatlas.org/). Genes upregulated in each tissue/stage were analyzed for over- or under-representation with all germline hits and each phenotypic category respectively. The analysis result was visualized using a heatmap.

2. The GSC genetic network (Figure 2E): The network was built by assembling an interaction matrix consisting of protein-protein interactions (BioGrid, IntAct, MINT, DIP, DPiM and DroID (Sep 2012 version)), Genetic interactions (FlyBase, BioGrid and DroID (Sep 2012 version)) and Literature cocitation interactions: gene2pubmed association was retrieved from NCBI EntrezGene ftp site (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/) on Jan 8th 2013. Pairwise gene co-citation relationships were extracted from PubMed. Subsequently, an interaction matrix was established amongst all genes that scored in GSC screen and the resulting network was visualized using Cytoscape. As the network fulfills the purpose to graphically display the hits, we have predominantly focused on using different interaction data (not only PPI) to be able to display the vast majority of the identified factors. Distinct molecular complexes or groups of genes with the same molecular function are outlined in black. Note: Genes that are not part of the interaction matrices are not displayed in the respective figures.

3. Complex analysis: Complex analysis was done using COMPLEAT (http://www.flyrnai.org/compleat/), a tool that annotates protein complexes from both literature and predictions from protein-protein network, and does gene set enrichment analysis based on protein complexes. Using COMPLEAT, we identified 116 non-redundant protein complexes that are over-represented among the genes scored comparing to the experimental background with p value cut-off 0.05 (Table S2).

4. GSC and Nb comparison (Figure 6B) and GO enrichment heatmap (Figure 6A): To compare GSC and Nb self-renewal genes, Nb hits were downloaded from Neuroblasts

Screen online database (http://neuroblasts.imba.oeaw.ac.at/index.php) and results were consolidated. Genes that scored with any measurement e.g. cell size change, cell number change or cell death, were compared to the genes scored in this germline screen. GO term enrichment was performed with DAVID (http://david.abcc.ncifcrf.gov) for both, GSC and Nb screens.

5. Cell cycle and cell growth gene enrichment analysis (Figure 6C): Cell cycle genes were retrieved from studies using fission yeast (Hayles et al., 2013) and human cells (Kittler et al., 2007). A gene list associated with the regulation of nucleolar size was assembled from a comparative analysis in yeast and *Drosophila* (Neumuller et al., 2013) (all complex members that scored as hits were considered for this analysis). rRNA processing factors were described by (Tafforeau et al., 2013). The genes were then mapped to corresponding fly genes using DIOPT (Hu et al., 2011) with the most stringent filter. The probability of the enrichment was calculated based on a hypergeometric distribution using the experimental set as a background. The enrichment result was visually displayed using a heatmap by calculating the negative log 10 value of the probability score.

6. Peak calling for H3K4me3 and PolII (7G): The SICER software was used to define genomic regions with significant enrichment ("peaks") for Pol II binding or histone modifications. For H3K4me3 chip-seq data, we used G=0 (gap=0) because the initial analysis showed that K4 peaks are more likely to be sharp and contiguous. Peaks returned by SICER were further refined by taking the first positive score (IP minus WCE) and the

last positive score. Peaks were mapped to genes using the following procedures: First, all transcripts of a specific gene were taken, and the beginning of the gene is defined as the first base of all transcripts; while the end defined as the last base of all transcripts. Second, we use BED tools to pick out genes that have at least 1 bp overlap between the gene region and the peak region. Third, different filters were used to eliminate false positive genes. (1) For H3K4me3-associated genes, if the "distance to peak start (defined as peak start minus gene start)" < 0, the peak location is given as "upstream"; if the "distance to peak summit (defined as peak summit minus gene start)" > 0 and the distance is shorter than the gene length, the peak location is defined as "inside". Otherwise the peak location is given as "NA". (2) RNA polymerase 2-associated genes were identified using the same strategy as H3K4me3.

Supplemental References

Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics *12*, 357.