

Molecular characterization of embryonic gonads by gene expression profiling in *Drosophila melanogaster*

Shuji Shigenobu*^{†‡}, Yu Kitadate*^{†‡}, Chiyo Noda*, and Satoru Kobayashi*^{†§¶}

*Okazaki Institute for Integrative Bioscience, National Institute for Basic Biology, National Institutes of Natural Sciences, Higashiyama, Myodaiji, Okazaki 444-8787, Japan; [†]Department of Basic Biology, School of Life Science, Graduate University for Advanced Studies, Nishigonaka, Myodaiji, Okazaki 444-8585, Japan; and [§]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Honcho, Kawaguchi 332-0012, Japan

Edited by Allan C. Spradling, Carnegie Institution of Washington, Baltimore, MD, and approved July 13, 2006 (received for review May 10, 2006)

In many animal species, germ-line progenitors associate with gonadal somatic cells to form the embryonic gonads (EGs) that later develop into functional organ producing gametes. To explore the genetic regulation of the germ-line development, we initiated a comprehensive identification and functional analysis of the genes expressed within the EGs. First, we generated a cDNA library from gonads purified from *Drosophila* embryos by FACS. Using this library, we catalogued the genes expressed in the gonad by EST analysis. A total of 17,218 high-quality ESTs representing 3,051 genes were obtained, corresponding to 20% of the predicted genes in the genome. The EG transcriptome is unexpectedly distinct from that of adult gonads and includes an extremely high proportion of retrotransposon-derived transcripts. We verified 101 genes preferentially expressed in the EGs by whole-mount *in situ* hybridization. Within this subset, 39 and 58 genes were expressed predominantly in germ-line and somatic cells, respectively, whereas four genes were expressed in the both cell lineages. The gonad-enriched genes encompassed a variety of predicted functions. However, genes implicated in SUMOylation and protein translation, including germ-line-specific ribosomal proteins, are preferentially expressed in the germ line, whereas the expression of various retrotransposons and RNAi-related genes are more prominent in the gonadal soma. These transcriptome data are a resource for understanding the mechanism of various cellular events during germ-line development.

expressed sequence tag | germ cell | retrotransposon | pole cell

The germ line is the only cell type that transmits genetic materials from one generation to the next during sexual reproduction. In many animal species, germ-line progenitors migrate within embryos to associate with gonadal somatic cells to form the embryonic gonads (EGs) that will later develop into a fully functional organ capable of producing gametes. In *Drosophila*, the germ-line progenitors, or pole cells, form at the posterior pole region of the early embryos (1, 2). Pole cells then migrate toward the mesodermal layer, where they associate with the specialized mesodermal cells known as somatic gonadal precursors. Eventually, the somatic cells encapsulate the pole cells to form EGs. Within the gonads, the pole cells undergo oogenesis or spermatogenesis and differentiate into germ cells during postembryonic development. Pole cells that fail to be encapsulated within the gonads eventually degenerate without producing germ cells (3).

Within the EGs, distinct cellular events associated with germ-line development occur, such as resumption of germ-line proliferation (4, 5), selection of the germ-line stem cell (6), gonad morphogenesis (7), and cellular communication between germ-line and somatic cells (8–10). Recent studies have also revealed that the male germ-line stem cell niche is already specified in the EG (ref. 11; Y.K., S.S., K. Arita, and S.K., unpublished data). Despite the importance of the EG in germ-line development, only limited information is available, regarding which genes are expressed in the EG although transcriptome data of adult testes and ovaries have accumulated (12, 13). Thus, we attempted to identify the genes expressed within the EGs by a direct and comprehensive approach.

In *Drosophila*, transcriptome analysis of individual organs and cell types has been hampered by the smallness of their size. To overcome this problem, we have developed an efficient method to isolate EGs by flow cytometry (14). We generated a cDNA library from purified gonads and obtained 17,218 valid ESTs representing 3,051 genes, all of which were examined by whole-mount *in situ* hybridization (WISH). The transcripts from 101 genes were enriched in the EG. These genes encompass a wide array of molecular and biological processes, as deduced from the Gene Ontology (GO) categories in the fly database. Here, we highlight five functional categories of genes enriched in the EG and discuss their roles.

Results and Discussion

Purification of EG by FACS and Generation of ESTs. We used FACS to isolate EG from transgenic *Drosophila* embryos harboring the germ-line marker *EGFP-vasa* (15). Embryos at 10–18 h after egg laying were homogenized without protease treatment to keep the gonad intact. From these homogenates, gonads containing both GFP-positive pole cells and GFP-negative gonadal somatic cells were separated from the remaining tissue by FACS. With this procedure, we were able to obtain a highly enriched fraction of EG, as confirmed by microscopy and quantitative PCR (14). We constructed an EG cDNA library (EG library) from poly(A)⁺ RNA from a pool of ≈25,000 FACS-sorted EGs. We sequenced 12,977 cDNA clones from the 5' end and 6,755 from the 3' end. After removing low-quality and contaminating sequences, 17,218 high-quality reads were obtained (DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank accession nos. BP540206–BP560422).

When aligned to *Drosophila melanogaster* genomic sequences, 15,384 (90.1%) ESTs mapped to euchromatic genomic regions and 434 (2.5%) to heterochromatic genomic regions. The remainder, 1,254 ESTs (7.4%), mapped to multiple loci within the genome; these included 974 highly repetitive sequences (≥10 hits in the genome). Compared with the public EST collections, this EG library includes a significantly higher proportion of repetitive sequences (Fig. 1A). Almost all of the repetitive sequences were derived from retrotransposons (Fig. 1B).

We aligned each EST with a reference transcript set in the *Drosophila* database (FlyBase, <http://flybase.net>) and assigned it to a gene. In total, we consolidated 17,072 ESTs derived from the EG library into a nonredundant set of 3,051 genes; these correspond to ≈20% of the predicted genes in *D. melanogaster*. Because our EST

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: EG, embryonic gonad; AG, adult gonad; OT, other tissue; WISH, whole-mount *in situ* hybridization; GO, Gene Ontology; NCBI, National Center for Biotechnology Information.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. BP540206–BP560422).

[§]S.S. and Y.K. contributed equally to this work.

[¶]To whom correspondence should be addressed. E-mail: skob@nibb.ac.jp.

© 2006 by The National Academy of Sciences of the USA

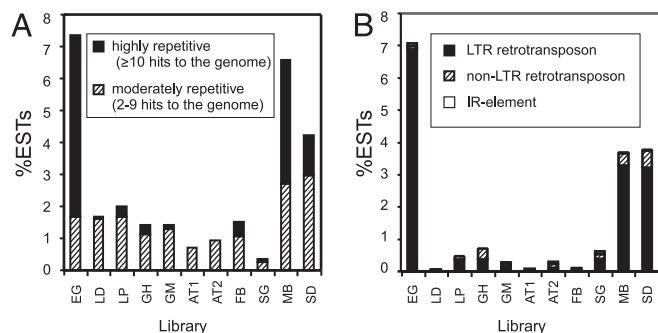


Fig. 1. The transcriptome of the EG includes highly repetitive elements, almost all of which are derived from LTR retrotransposons. (A) ESTs with multiple hits to the genome. EST sequences that match multiple loci on the genome were classified into two groups: highly repetitive (≥ 10 hits to the genome) and moderately repetitive (two to nine hits). The proportions of the repetitive ESTs are shown. LD, 0- to 22-h whole embryo; LP, larva and early pupa; GH, adult head; GM, adult ovary; AT1, testis (0- to 3-day adult); AT2, testis (1- to 5-day adult); FB, fat body of larva; SG, salivary gland of pupa; MB, mbn2 cell line; and SD, Schneider L2 culture cells. The EG library includes a significantly higher proportion of repetitive sequences than other EST libraries. (B) Proportion of transposon-derived transcripts in EST libraries. A significant accumulation of LTR retrotransposons is observed in the EG library. EST libraries generated from cell lines (MB and SD) also include many highly repetitive and retrotransposon-derived sequences. In general, the number of transposable elements is increased in cell lines (50).

analysis was nearly saturating (see *Supporting Text*, which is published as supporting information on the PNAS web site), it covers most of the gene repertoire of the EG transcriptome. All identified genes expressed in the EG are listed in Table 3, which is published as supporting information on the PNAS web site.

Using the National Center for Biotechnology Information (NCBI) UniGene-based classification, we compared the gene sets expressed by EGs, adult gonads (AGs; ovary + testis), and other tissues (OTs). Genes represented in all three collections are regarded as those with “housekeeping” functions (1,809 UniGenes; 60.8% of EG UniGene collection). Except the housekeeping genes, we observed that the proportion of genes present in both the EG and AG collections was low (Fig. 2). Only 145 genes (5.0% of the EG UniGene collection) were common to EG and AG, whereas 719 genes (24.8% of EG UniGene collection) were expressed in both EGs and OTs but not AGs. Thus, the EG transcriptome is unexpectedly different from AG. We conclude that the genetic regulation of germ-line development within the EGs is distinct from the one underlying gametogenesis in ovaries and testes.

Overview of Comprehensive Whole-Mount *In Situ* Hybridization (WISH). All 3,051 genes represented in our EG library were subjected to WISH to examine their distribution within the embryo. Overall, we obtained useful expression data for 2,388 genes. Although most of them showed ubiquitous distribution, we found that

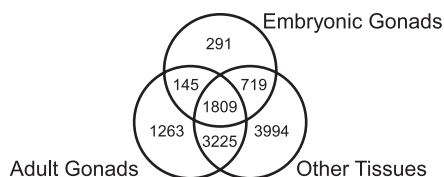


Fig. 2. Comparison of genes expressed by EGs, AGs, and OTs. The AG EST collection is a pool of adult testis libraries and adult ovary libraries. The OT EST collection is a pool of seven public EST libraries, including head, salivary gland, fat body, and culture cell EST libraries. The ESTs were collapsed into UniGene sets that were precompiled by NCBI. The three-way comparison shows that the expressed gene repertoire of the EG is quite different from that of the AG.

transcripts from 101 genes were enriched in the EG, as summarized in Table 1. We further examined their distribution within the gonads by double-staining embryos with a RNA probe for each transcript and an anti-VASA antibody to distinguish the germ-line and somatic expression of the transcripts within the gonads. We identified 39 RNAs that are expressed predominantly in pole cells, 58 that are expressed in gonadal somatic cells, and 4 that are expressed in both cell types (Table 1, Fig. 3).

We investigated the temporal expression patterns of transcripts enriched in pole cells by WISH. Embryos at various developmental stages were examined (Fig. 4, which is published as supporting information on the PNAS web site), and three major expression patterns were extracted (Types I, II and III). Transcripts with the Type I expression pattern are first observed in the pole cells during their migration through the posterior midgut epithelium and remain detectable after the coalescence of the gonads. Transcripts from *vasa*, *RpL22-like*, *RpS19b*, *CG10990*, *CG4415*, *TfIIA-S*, and *Ssb-c31b* exhibit this type of expression. Because the pole cells are transcriptionally inactive until they migrate (16), these transcripts are some of the earliest zygotic transcripts in the pole cells. Given that their transcription is initiated in the pole cells before coalescing with the gonadal somatic cells, we speculate that their expression is autonomously initiated by maternal factors partitioned into the pole cells, rather than an inductive signal from the gonadal soma. Indeed, the expression of some Type I genes also was detectable “lost” pole cells that failed to be incorporated within the gonads. Transcripts with a Type II expression pattern are observed in various tissues before gonad formation but are enriched in pole cells after they associate with the gonadal somatic cells. Transcripts for *smt3*, *Uba2*, *lwr*, *Top2*, and *grp* display this type of expression pattern. Type III expression includes transcripts that accumulate in the pole cells throughout embryogenesis. These transcripts present in the early pole cells are presumably maternal in origin, whereas zygotic transcription may occur at later stages. This type includes transcripts from *ovo*, *stai*, *Hsp26*, *Hsp27*, *Hsp83*, and *zpg*.

Functional Classification of EG-Enriched Genes. To characterize the EG transcriptome, we assigned GO terms to each EG-enriched gene according to FlyBase annotations. As shown in Table 4, which is published as supporting information on the PNAS web site, EG-enriched genes represent a broad range of biological and molecular functions. Our statistical analysis showed that some of the categories were significantly overrepresented in the list of EG-enriched genes (Table 2). Among them, five categories are highlighted and discussed in detail.

Germ-Line Development. We found that genes in the GO category “germ cell development” were overrepresented in the list of EG-enriched genes (Table 2). It is generally expected that genes responsible for germ-line development are predominantly expressed in the gonads. However, their functions are known to be required within the AGs. For example, *piwi* is expressed in the somatic cells adjacent to germ-line stem cells and is essential for stem cell self renewal (17). *zpg* is required for survival of differentiating early germ cells in AGs (18), and *armi* represses *oskar* translation in ovaries and *Ste* expression in testes (19, 20). Although their functions during embryogenesis are unclear, these genes were expressed in the EGs. A similar precocious expression has been reported for meiotic genes; a subset of the genes responsible for meiotic division is expressed in pole cells during embryogenesis, whereas meiosis is initiated later at the postembryonic stages (21). It is possible that transcription of these gametogenesis-related genes initiates in the EGs, but posttranscriptional repression restricts the function of these genes until the onset of gametogenesis. Although we cannot exclude the possibility that these genes may have additional functions, our observations are consistent with the notion that the EG acquires at least a part of the potential to carry out gametogenesis.

Table 1. EG-enriched genes identified by WISH

Gene	Tissues*	Functions (excerpt)	FlyBase ID
Expressed in germline			
<i>RpS13</i>	GO[g], CNS, LG	Ribosomal protein	FBgn0010265
<i>RpS19b</i>	GO[g], PV	Ribosomal protein	FBgn0039129
<i>CG9871</i>	GO[g], SNS	Ribosomal protein	FBgn0034837
<i>RpS5b</i>	GO[g], U	Ribosomal protein	FBgn0038277
<i>smt3</i>	GO[g], CNS	Sumoylation	FBgn0026170
<i>Uba2</i>	GO[g], CNS	Sumoylation	FBgn0029113
<i>lvr</i>	GO[g], CNS, U	Sumoylation	FBgn0010602
<i>Aos1</i>	GO[g], CNS	Sumoylation	FBgn0029512
<i>Hsp27</i>	GO[g], CNS, DV	Protein folding	FBgn0001226
<i>Hsp26</i>	GO[g], MGL	Protein folding	FBgn0001225
<i>Hsp83</i>	GO[g], OE, PNS, CNS, MG	Protein folding	FBgn0001233
<i>CG4415</i>	GO[g], PV	Unfolded protein binding	FBgn0031296
<i>Cam</i>	GO[g], CNS, OE	Calcium ion binding	FBgn0000253
<i>I(1)G0269</i>	GO[g], CNS, PNS	CTD-like phosphatase	FBgn0029067
<i>grp</i>	GO[g], CNS	Protein serine/threonine kinase; cell cycle	FBgn0011598
<i>CG2919</i>	GO[g], BR	Cytoskeleton organization and biogenesis	FBgn0037348
<i>scra</i>	GO[g], CNS	Cytoskeleton organization and biogenesis	FBgn0004243
<i>Mapmodulin</i>	GO[g], CNS	Microtubule binding	FBgn0034282
<i>stai</i>	GO[g], CNS, PNS	Microtubule binding	FBgn0051641
<i>Mcm6</i>	GO[g], CNS, MG, FB	DNA helicase	FBgn0025815
<i>Top2</i>	GO[g], CNS	DNA topoisomerase	FBgn0003732
<i>Thd1</i>	GO[g], CNS, U	Pyrimidine-specific mismatch base pair DNA N-glycosylase	FBgn0026869
<i>dUTPase</i>	GO[g], CNS, U, MG	dUTP diphosphatase	FBgn0013349
<i>TfIIA-S</i>	GO[g], CNS, U	General transcription factor	FBgn0013347
<i>Ssb-c31a</i>	GO[g], MG, BR	Transcription coactivator	FBgn0015299
<i>ovo</i>	GO[g], EP, BR	Transcription factor	FBgn0003028
<i>Fs(2)Ket</i>	GO[g], CNS, U	Importin β	FBgn0000986
<i>zpg</i>	GO[g]	Innexin channel	FBgn0024177
<i>janA</i>	GO[g], MG, U	Sex differentiation	FBgn0001280
<i>CSN3</i>	GO[g], CNS, U	Signalosome complex	FBgn0027055
<i>Uba1</i>	GO[g], CNS, U	Ubiquitin-activating enzyme	FBgn0023143
<i>CG10990</i>	GO[g]	Translation elongation factor	FBgn0030520
<i>vas</i>	GO[g]	RNA helicase	FBgn0003970
<i>CG11329</i>	GO[g]		FBgn0031848
<i>CG15930</i>	GO[g]		FBgn0029754
<i>CG18213</i>	GO[g]		FBgn0038470
<i>CG12576</i>	GO[g], CNS, MG		FBgn0031190
<i>CG14346</i>	GO[g], U		FBgn0031337
<i>Unnamed gene</i>	GO[g], U		FBgn0058460
Expressed in somatic line			
<i>412</i>	GO[s]	Retrotransposon	FBgn0000006
<i>297</i>	GO[s], CNS, U	Retrotransposon	FBgn0000005
<i>17.6</i>	GO[s], DV, PH	Retrotransposon	FBgn0000004
<i>mdg1</i>	GO[s]	Retrotransposon	FBgn0002697
<i>Quasimodo</i>	GO[s]	Retrotransposon	FBgn0062261
<i>Stalker</i>	GO[s], U	Retrotransposon	FBgn0064138
<i>Stalker2</i>	GO[s], PV	Retrotransposon	FBgn0063399
<i>Tabor</i>	GO[s], SNS	Retrotransposon	FBgn0045970
<i>Tirant</i>	GO[s]	Retrotransposon	FBgn0004082
<i>ZAM</i>	GO[s], U	Retrotransposon	FBgn0023131
<i>gtwin</i>	GO[s]	Retrotransposon	FBgn0063436
<i>armi</i>	GO[s], FB	RNA interference	FBgn0041164
<i>Dcr-2</i>	GO[s], U	RNA interference	FBgn0034246
<i>piwi</i>	GO[s]	RNA interference	FBgn0004872
<i>CG8908</i>	GO[s], U	ABC transporter	FBgn0034493
<i>CG30359</i>	GO[s], FB	Carbohydrate metabolism; transporter	FBgn0050359
<i>CG3036</i>	GO[s], PNS, GC	Membrane protein; transporter	FBgn0031645
<i>CG9935</i>	GO[s], SNS, LG, PE	Membrane protein; transporter	FBgn0039916

Table 1. (continued)

Gene	Tissues*	Functions (excerpt)	FlyBase ID
<i>CG11537</i>	GO[s], CNS, SG	Membrane protein; transporter	FBgn0035400
<i>CG1599</i>	GO[s], U	Plasma membrane protein; v-SNARE	FBgn0033452
<i>CG3074</i>	GO[s], SNS	Cathepsin B; proteolysis	FBgn0034709
<i>CG9634</i>	GO[s], U	Proteolysis	FBgn0027528
<i>m1</i>	GO[s], U	Serine-type endopeptidase inhibitor	FBgn0002578
<i>Fas1</i>	GO[s], CNS, PNS	Cell adhesion	FBgn0000634
<i>I(2)03709</i>	GO[s], MG, FB, MU	Cell cycle, DNA metabolism	FBgn0010551
<i>Wnt6</i>	GO[s], MG, MP, GC, MG	frizzled-2 signaling	FBgn0031902
<i>skf</i>	GO[s], U	Plasma membrane protein; signal transduction	FBgn0050021
<i>lbm</i>	GO[s], CNS, PNS	Tetraspanin; receptor signaling protein	FBgn0016032
<i>CG7194</i>	GO[s]	Gonad development	FBgn0035868
<i>M(2)21AB</i>	GO[s], RG, MG	Methionine adenosyltransferase	FBgn0005278
<i>mRpS24</i>	GO[s], FB, MG	Mitochondrial ribosomal protein	FBgn0039159
<i>Mocs1</i>	GO[s], MG	Mo-molybdopterin cofactor biosynthesis	FBgn0036122
<i>mud</i>	GO[s]	Mushroom body development	FBgn0002873
<i>Pros45</i>	GO[s], CNS	Proteasome complex	FBgn0020369
<i>CG10565</i>	GO[s], MG, PV, FB	Protein folding; nucleic acid binding	FBgn0037051
<i>stg</i>	GO[s]	Protein tyrosine/serine/threonine phosphatase; cell cycle	FBgn0003525
<i>CG5800</i>	GO[s], SNS, MG	RNA helicase	FBgn0030855
<i>B52</i>	GO[s], U	RNA splicing factor activity	FBgn0004587
<i>CG11447</i>	GO[s], MG, ES, HG	rRNA (uridine-2'-O)-methyltransferase	FBgn0038737
<i>zfh1</i>	GO[s], CNS, PNS	Transcription factor	FBgn0004606
<i>stc</i>	GO[s], FB	Transcription factor	FBgn0001978
<i>esg</i>	GO[s], HIB	Transcription factor	FBgn0001981
<i>ftz-f1</i>	GO[s], PV	Transcription factor	FBgn0001078
<i>neur</i>	GO[s], CNS, FB	Ubiquitin-protein ligase	FBgn0002932
<i>novel gene</i>	GO[s], CNS, PNS		FGM222E05 [†]
<i>novel gene</i>	GO[s], CNS, U		FGC026A04 [†]
<i>CG15784</i>	GO[s], CNS		FBgn0029766
<i>CG7267</i>	GO[s], FB, U		FBgn0030079
<i>CG33047</i>	GO[s], GRL		FBgn0053047
<i>CG6014</i>	GO[s], HG		FBgn0027542
<i>CG7498</i>	GO[s], LG, SNS, FB		FBgn0040833
<i>CG7224</i>	GO[s], MG, GC, PV		FBgn0031971
<i>CG5541</i>	GO[s], MG, HG, ES, PV		FBgn0030603
<i>CG14998</i>	GO[s], PNS, FB		FBgn0035500
<i>CG11050</i>	GO[s], U		FBgn0031836
<i>dpr17</i>	GO[s], U		FBgn0051361
<i>CG14072</i>	GO[s], U		FBgn0032318
Expressed in germline and somatic line			
<i>Su(var)205</i>	GO[g/s], CNS	Chromatin binding	FBgn0003607
<i>Df31</i>	GO[g/s], CNS, PNS, misc, HIB	Histone binding	FBgn0022893
<i>ran</i>	GO[g/s], CNS, BR, FB	Ras GTPase	FBgn0020255
<i>14-3-3e</i>	GO[g/s], CNS, U	Ras protein signal transduction	FBgn0020238

*BR, brain; DV, dorsal vessel; EP, epidermis; ES, esophagus; FB, fat body; GC, gastric caeca; GO, gonad; HG, hindgut; HIB, histoblast; LG, lymph gland; MG, midgut; MGL, midline glial cell; MP, Malpighian tubule; MU, muscle; OE, oenocyte; PE, pericardial cells; PH, pharynx; PNS, peripheral nervous system; PV, proventriculus; RG, ring gland; SG, salivary gland; SNS, stomatogastric nervous system; U, weak signal is ubiquitously detected; g and s in a pair of brace indicate germ-line and somatic line expression in the gonad.

[†]Instead of FlyBase ID, the clone name is shown for the novel gene.

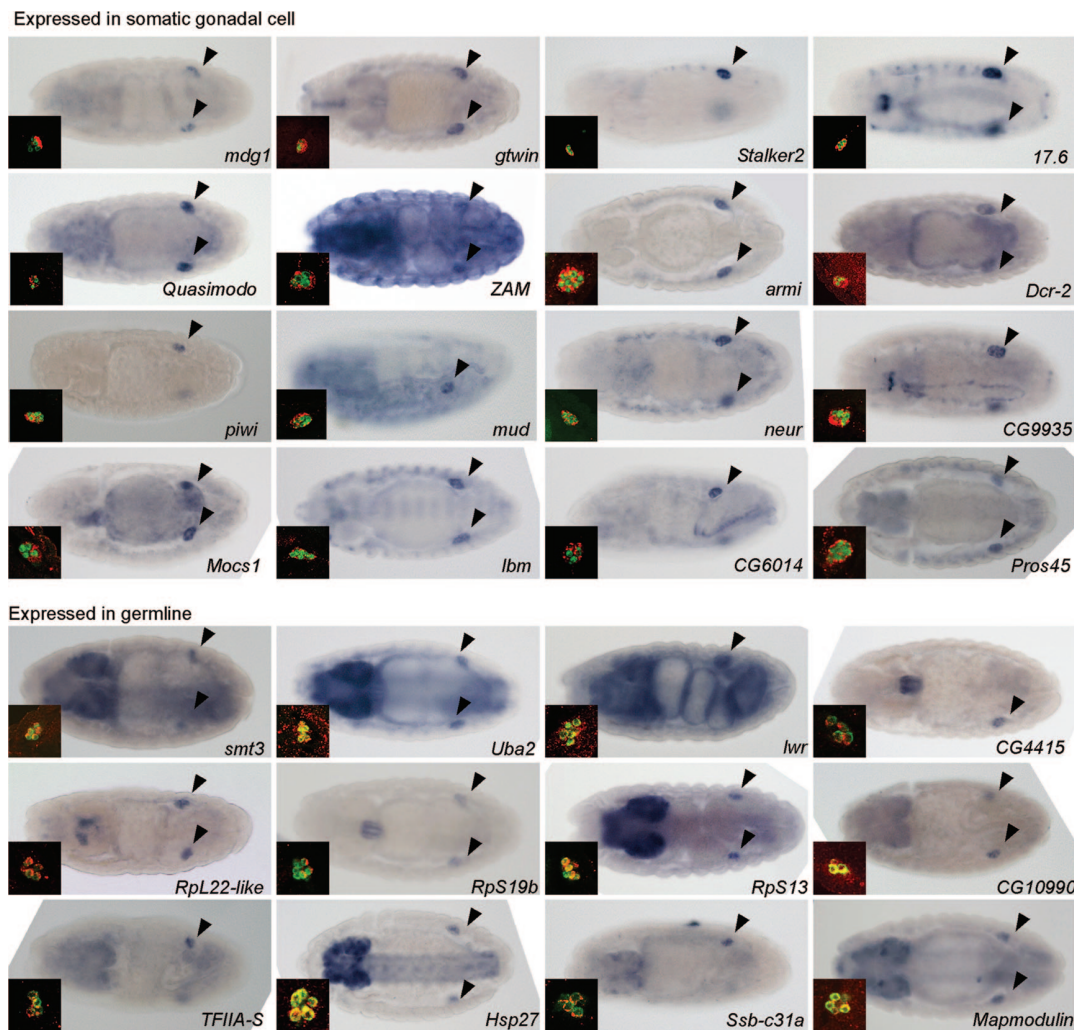


Fig. 3. Spatial expression patterns of gonad-enriched transcripts. Embryos at stage 14–16 are shown with anterior to the left. Gonads are indicated by arrowheads, and gene names are shown in the bottom right of each image. *Insets* provide the confocal microscopic image of the gonad double-stained with an antisense RNA probe for the indicated gene (red) and anti-Vasa (green), a germ-line marker.

LTR Retrotransposons. We observed that a surprisingly large number of EG ESTs (>1,000 ESTs) were derived from retrotransposons; this population corresponds to 7% of the EG EST collection. This proportion was significantly larger than in other public EST collections (Fig. 1*B*). Thus, retrotransposons are predominantly expressed in the EGs. Approximately 100 families of retrotransposons have been identified in *Drosophila* genome (22). Our EST analysis detected transcripts from various types of retrotransposons (30 families) but was dominated by those with LTRs (23 families; Table 5, which is published as supporting information on the PNAS web site). The WISH experiments reveal that at least 11 LTR retrotransposons, *17.6*, *297*, *412*, *gtwin*, *mdg1*, *quasimode*, *stalker*, *stalker2*, *tabor*, *ZAM*, and *tirant*, are expressed predominantly in the EGs (Fig. 3). These observations are in accordance with the previous reports showing that transcripts for *17.6*, *412*, *mdg1*, *297*, and *gypsy* accumulate in the EGs (23, 24). It is interesting to note that these transcripts were all detected in gonadal somatic cells rather than in pole cells (Fig. 3). Thus, we conclude that hyperexpression in the gonadal somatic cells is a common feature of various types of LTR retrotransposons.

The significance of the retrotransposon expression in gonadal somatic cells is unclear. Expression and retrotransposition in germ line would be a more effective strategy for retrotransposons to propagate them in a heritable manner from one generation to the

next. An interesting case has been reported in a specific strain called *Rev1*, in which the retrotransposon *ZAM* is expressed in the follicle cells of the adult ovaries and forms virus-like particles that transfer to neighboring oocytes (25). A similar transfer has been reported for the virus-like particles originating from the *gypsy* retrotransposon in the ovaries of *flamenco* mutant females (26). This translocation of virus-like particles may couple with yolk transfer from follicle cells to the oocytes by exo- and endocytosis and/or through gap junctions (25, 27). Retrotransposons may exploit the intimate link between the follicle cells and the oocytes to obtain additional access to gametes. This somatic expression may circumvent a host defense against retrotransposons in the germ line (28, 29). However, it is worthwhile to note that the expression of *ZAM* and *gypsy* retrotransposons are detectable only in certain genetic backgrounds, such as *Rev1* and the *flamenco* mutant, respectively (25, 27). Thus, their expression is normally repressed in the follicle cells. In contrast, in our experiments, transcripts from various retrotransposons are preferentially expressed in gonadal somatic cells during normal embryogenesis. One possibility is that their early transcription is regulated differently, and the transcripts are inactivated by a posttranscriptional regulatory mechanism (see below).

RNAi. Among the *Drosophila* genes, 11 are annotated to be associated with “RNAi” (FlyBase), a mechanism by which dsRNA

Table 2. Functional classification of gonad-enriched genes based on GO terms

GO name (GO ID)	No. genes enriched in EG	No. genes in the genome	P value	Genes
Biological process				
Germ cell development (0007281)	5	157	0.012	<i>zpg, vas, armi, 14-3-3e, piwi</i>
DNA replication (0006260)	5	128	5.0E-03	<i>armi, 1(2)03709, Mcm6, Top2, Thd1</i>
Protein folding (0006457)	4	134	0.029	<i>CG10565, Hsp26, Hsp27, Hsp83</i>
Protein import into nucleus (0006606)	3	49	0.0087	<i>Fs(2)Ket, lwr, smt3</i>
Response to heat (0009408)	3	51	0.0097	<i>Hsp26, Hsp27, Hsp83</i>
RNA interference (0016246)	3	11	9.9E-05	<i>armi, Dcr-2, piwi</i>
Molecular function				
Microtubule binding (0008017)	3	76	0.028	<i>Mapmodulin, scra, stai</i>
SUMO-activating enzyme activity (0019948)	2	2	7.4E-05	<i>Aos1, Uba2</i>
Cellular component				
Cytosolic ribosome (0005830)	4	99	0.011	<i>CG9871, RpS13, RpS19b, RpS5b</i>

Overrepresented categories are chosen.

induces gene silencing. Transcripts from nine RNAi-related genes are constituents of the EG EST collection. Among them, three genes, *piwi*, *armi*, and *Dcr-2*, were expressed predominantly in the EGs by WISH (Fig. 3). These transcripts were all detectable in the gonadal somatic cells rather than pole cells in the EGs (Fig. 3). The functions of *piwi* and *armi* have been investigated in the AG. *piwi* is an Argonaute-family gene necessary for germ-line stem cell renewal in ovaries (17), and *armi* is required for polarization of the oocyte (20) and for silencing of *Stellate* gene in male germ cells (19). However, the functions of these genes in somatic gonadal cells during embryogenesis are not yet clear. We propose that a RNAi-mediated gene silencing mechanism is active in the somatic cells of the EGs.

RNAi-mediated mechanisms contribute to host defenses against transposons and viruses (30–32). A subset of mutations that disable the RNAi mechanism mobilizes families of transposable elements. For example, the LTR retrotransposons *gypsy* and *ZAM* are regulated by a mechanism that depends on *piwi* (30–32). In *Drosophila*, transposons and repeated sequences, including *P-element*, *Stellate*, *I-element*, and *gypsy*, are repressed by a trans-silencing mechanism termed “cosuppression” that targets any transposons containing homologous sequences to the “trigger” transcripts by small interfering RNA (siRNA; refs. 33–35). Based on the aforementioned observations that transcripts from various LTR retrotransposons and the RNAi-related genes are both enriched in the gonadal somatic cells of the embryos, we hypothesize that LTR-retrotransposon transcripts would be the “trigger;” they are processed by RNAi pathway to produce siRNA, which in turn silences the retrotransposons in the following developmental stages.

This hypothesis is supported by our observations that *Dcr-2* but not *Dcr-1*, the two *Drosophila* dicer homologs, is predominantly expressed in the EGs, because *Dcr-2* is responsible for the production of small interfering RNA (siRNA) from dsRNA, whereas *Dcr-1* is for microRNA-triggered gene silencing (36–38). A recent analysis of the small RNAs expressed during *Drosophila* embryogenesis has identified a large number of repeat-associated siRNAs, which are complementary to repetitive elements, including retro-

transposons (39). Although the distribution of these small RNAs remains unclear, it is likely that *Dcr-2* and the other RNAi-related genes process transcripts from the LTR retrotransposons in the gonadal soma. Further studies examining the role of these RNAi-related genes in the EGs are required to investigate this hypothesis.

SUMOylation. We found that almost all of the components required for SUMOylation are expressed predominantly in pole cells. SUMO is a member of the ubiquitin-like protein family that regulates cellular function by binding covalently to a variety of target proteins (40). SUMO (*smt3*) is one of the most highly represented transcripts in the EG EST collection (80 ESTs; Table 3). Our WISH analysis revealed that *smt3* RNA is enriched in pole cells as well as in the CNS (Fig. 3). Similarly, E1 and E2 components, which are encoded by *Uba2*, *Aos1*, and *lwr* genes, are all concentrated in the pole cells (Fig. 3). In addition, these transcripts exhibit quite similar temporal expression (Type III; see above). The common spatiotemporal expression pattern suggests that SUMOylation occurs in pole cells within the gonad of developing embryos.

A large fraction of the SUMO substrates identified by global proteomics and studies *in silico* contribute to transcription (41, 42). Thus, SUMOylation may regulate germ-line gene expression by posttranslational modification of transcription factors. Indeed, our computational analysis reveals that the EG-EST collection contains a number of potential substrates for SUMOylation, including proteins involved in transcription (data not shown). To understand the role of SUMOylation in germ-line development, we are attempting to identify SUMO substrates with genetic and biochemical approaches.

Germ-Line-Specific Ribosomal Proteins. Four genes (*RpL22-like*, *RpS19b*, *RpS5b*, and *RpS13*) encoding cytosolic ribosomal proteins are preferentially in pole cells within the EGs (Fig. 3). We found that three (*RpL22-like*, *RpS19b*, and *RpS5b*) of the four genes have paralogs (*RpL22*, *RpS19a*, and *RpS5a*, respectively) in the genome, and these paralogs are expressed ubiquitously throughout late embryos (14). Thus, *RpL22*, *RpS5a*, and *RpS19a* are used universally, and *RpL22-like*, *RpS5b*, and *RpS19b* have a specialized role in the germ line.

In addition to the ribosomal proteins, transcripts encoding translational regulators also are expressed preferentially in pole cells. For example, CG10990, which encodes a translational repressor distantly related to eIF4G and PDCD4 (43), was detected in pole cells in late embryos (Fig. 3). A *Drosophila* homolog of mammalian RpS13, which interacts with PDCD4 in HeLa cells (44), is expressed in the pole cells (Fig. 3). In addition, *vasa*, which has sequence similarity to eIF4A, is zygotically activated in pole cells during their migration to the gonads, and the transcriptions of *RpL22-like*, *RpS19*, and *CG10990* are all activated in pole cells at nearly same time as that of *vasa* (data not shown). It is interesting to note that these translation-related genes (*RpS5b*, *RpS19b*, *RpL22-like*, *CG10990*, and *vasa*) are also up-regulated in the germ-line stem cells of adult ovaries (45). These germ-line-specific components may be essential for the translational regulatory mechanisms required for germ-line development.

Alternatively, it is possible that the germ-line-specific ribosomal proteins carry out extraribosomal functions. It is plausible that the duplicated genes for ribosomal proteins acquire novel functions unrelated to their paralogs. This view is supported by our data that the germ-line-specific paralogs of the *RpL22* and *RpS19* families are more divergent than the universal ones; for example, the *D. melanogaster* RpL22 protein sequence is 57% identical to human RPL22, whereas germ-line-specific RpL22-like displays only 44% identity (14). Novel functions of ribosomal proteins have been reported. In human cells infected with Epstein–Barr virus, an appreciable portion of the RpL22 is not associated with ribosomes but is located in the nucleoplasm, where RpL22 binds to a small viral RNA (46). In addition, RpL22 has been identified as a protein

associated with telomerase RNA (47). Thus, we speculate that the *Drosophila* paralogs of ribosomal proteins have acquired novel functions that contribute to germ-line development.

The mechanism for germ-line-specific expression of the paralogs of ribosomal proteins is not yet clear. The similarity of their spatiotemporal expression is consistent with these genes being regulated in a coordinated fashion by germ-line-specific transcriptional machinery. An interesting case has been reported in *Ascaris lumbricoides* (48). Its genome encodes both germ-line- and soma-specific ribosomal proteins homologous to *RpS19*. A paralog, *RpS19G*, is expressed predominantly in the germ line but is eliminated from the genome of all somatic cells by chromatin diminution during early development. Instead, the other paralog, *RpS19S*, is expressed in the soma. Thus, we speculate that the differential expression of the ribosomal protein paralogs (and probably their function) is intimately related to the regulatory mechanism underlying germ-line development.

Perspectives. Here we describe the gene expression data obtained from our EST analysis of purified EGs. Our transcriptome data provide unique genetic information to help in the understanding of gonad development. Furthermore, the spatiotemporal expression data of the gonad-enriched genes are useful for studying the regulatory mechanism of germ-line- and gonadal soma-specific gene expression and function. The general transcription factor *TjflA-S* and transcription coactivator *Ssb-c31a* (Fig. 3) may be involved in germ-line-specific gene regulation at the transcriptional level. Recent studies indicate that transmembrane proteins are involved in gonad morphogenesis and the establishment of the germ-line–stem-cell niche within the EG (refs. 11 and 49; Y.K., S.S., K. Arita, and S.K., unpublished data). Our list of gonad-enriched genes includes many genes encoding membrane proteins that are predicted to function in cell–cell interactions, in signaling, and as transporters (Table 1). Functional analysis of these genes will help us to understand the mechanism of gonad formation, the germ–soma interaction, and the establishment of the germ-line–stem-cell niche within this specialized organ.

- Santos AC, Lehmann R (2004) *Curr Biol* 14:R578–R589.
- Williamson A, Lehmann R (1996) *Annu Rev Cell Dev Biol* 12:365–391.
- Hay B, Jan LY, Jan YN (1988) *Cell* 55:577–587.
- Asaoka-Taguchi M, Yamada M, Nakamura A, Hanyu K, Kobayashi S (1999) *Nat Cell Biol* 1:431–437.
- Sonnenblick BP (1994) in *Biology of Drosophila*, ed Demerec M (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 62–167.
- Asaoka M, Lin H (2004) *Development (Cambridge, UK)* 131:5079–5089.
- DeFalco TJ, Verney G, Jenkins AB, McCaffery JM, Russell S, Van Doren M (2003) *Dev Cell* 5:205–216.
- Mukai M, Kashikawa M, Kobayashi S (1999) *Development (Cambridge, UK)* 126:1023–1029.
- Jenkins AB, McCaffery JM, Van Doren M (2003) *Development (Cambridge, UK)* 130:4417–4426.
- Wawersik M, Milutinovich A, Casper AL, Matunis E, Williams B, Van Doren M (2005) *Nature* 436:563–567.
- Le Bras S, Van Doren M (2006) *Dev Biol* 294:92–103.
- Andrews J, Bouffard GG, Cheadle C, Lu JN, Becker KG, Oliver B (2000) *Genome Res* 10:2030–2043.
- Parisi M, Nuttall R, Edwards P, Minor J, Naiman D, Lu J, Doctolero M, Vainer M, Chan C, Malley J, et al. (2004) *Genome Biol* 5:R40.
- Shigenobu S, Arita K, Kitadate Y, Noda C, Kobayashi S (2006) *Dev Growth Differ* 48:49–57.
- Sano H, Nakamura A, Kobayashi S (2002) *Mech Dev* 112:129–139.
- Van Doren M, Williamson AL, Lehmann R (1998) *Curr Biol* 8:243–246.
- Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H (1998) *Genes Dev* 12:3715–3727.
- Tazuke SI, Schulz C, Gilboa L, Fogarty M, Mahowald AP, Guichet A, Ephrussi A, Wood CG, Lehmann R, Fuller MT (2002) *Development (Cambridge, UK)* 129:2529–2539.
- Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, Cook HA, Koppetsch BS, Theurkauf WE, Zamore PD (2004) *Cell* 116: 831–841.
- Cook HA, Koppetsch BS, Wu J, Theurkauf WE (2004) *Cell* 116:817–829.
- Mukai M, Kitadate Y, Arita K, Shigenobu S, Kobayashi S (2006) *Gene Expr Patterns* 6:256–266.
- Ashburner M, Golic K, Howley R (2005) *Drosophila, a Laboratory Handbook* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
- Brookman JJ, Toosy AT, Shashidhara LS, White RA (1992) *Development (Cambridge, UK)* 116:1185–1192.
- Ding D, Lipshitz HD (1994) *Genet Res* 64:167–181.

Materials and Methods

Fly Stocks. EGs were collected from *EGFP-vasa* embryos (15) by FACS. *y w* flies were used for WISH analysis. Detailed procedures for WISH are described in *Supporting Text*.

Construction of a cDNA Library from the FACS-Sorted EGs. The EG was isolated from *EGFP-vasa* transgenic embryos at 10–18 h after egg laying, as described (14). Microscopically, >99% of the total particles obtained by FACS were gonads (the number of particles we counted was >400). The remaining particles (<1%) were small noncellular clumps. A cDNA library was generated from $\approx 7 \mu\text{g}$ of total RNA, which was purified from 25,000 gonads by using the SMART system (Clontech, Mountain View, CA), as described (14). Two plasmids were used for cDNA construction, the pDNR-LIB vector (Clontech) and the pGEM-T Easy vector (Promega, Madison, WI). Information about primers used for EST sequencing is available in Table 6, which is published as supporting information on the PNAS web site. The clone name and corresponding EST accession no. used for each synthesis of the RNA probe are listed in Table 7, which is published as supporting information on the PNAS web site.

EST Sequencing and Informatics. Each EST was sequenced, processed, and annotated as described (14). Detailed information is provided in *Supporting Text*. The bioinformatics analyses on (i) the analysis of repetitive ESTs, (ii) functional annotation based on GO and the statistical analysis, and (iii) the EST comparison among public EST collections and our EG library are also available in *Supporting Text*.

We thank Dr. T. Akiyama (Azabu University, Fuchinobe, Japan) and Beckman Coulter (Fullerton, CA) for cell sorting, Dr. A. Nakamura (RIKEN, Kobe, Japan) for an anti-VASA antibody, and Mr. K. Hashiyama for WISH experiments. This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science, and Technology and the National Institute of Agrobiological Sciences and by the Core Research for Evolutional Science and Technology project of the Japan Science and Technology Agency.

- Leblanc P, Dessel S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B, Vaury C (2000) *J Virol* 74:10658–10669.
- Song SU, Kurkulos M, Boeke JD, Corces VG (1997) *Development (Cambridge, UK)* 124:2789–2798.
- Waksmanski SL, Woodruff RI (2002) *J Insect Physiol* 48:667–675.
- Aravin AA, Klenov MS, Vagin VV, Bantignies F, Cavalli G, Gvozdev V A (2004) *Mol Cell Biol* 24:6742–6750.
- Sijen T, Plasterk RH A (2003) *Nature* 426:310–314.
- Buchon N, Vaury C (2006) *Heredity* 96:195–202.
- Waterhouse PM, Wang M-B, Lough T (2001) *Nature* 411:834–842.
- Kavi HH, Fernandez HR, Xie W, Birchler JA (2005) *FEBS Lett* 579:5940–5949.
- Ronsseray S, Josse T, Boivin A, Anxolabehere D (2003) *Genetica* 117:327–335.
- Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA (2001) *Curr Biol* 11:1017–1027.
- Sarot E, Payen-Groschène G, Bucheton A, Pelissou A (2004) *Genetics* 166:1313–1321.
- Okamura K, Ishizuka A, Siomi H, Siomi MC (2004) *Genes Dev* 18:1655–1666.
- Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, Wang X (2003) *Science* 301:1921–1925.
- Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW (2004) *Cell* 117:69–81.
- Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T (2003) *Dev Cell* 5:337–350.
- Johnson ES (2004) *Annu Rev Biochem* 73:355–382.
- Zhou F, Xue Y, Lu H, Chen G, Yao X (2005) *FEBS Lett* 579:3369–3375.
- Wohlschlegel JA, Johnson ES, Reed SI, Yates JR, III (2004) *J Biol Chem* 279:45662–45668.
- Yang H-S, Jansen AP, Komar AA, Zheng X, Merrick WC, Costes S, Lockett SJ, Sonenberg N, Colburn NH (2003) *Mol Cell Biol* 23:26–37.
- Kang M-J, Ahn H-S, Lee J-Y, Matsushashi S, Park W-Y (2002) *Biochem Biophys Res Commun* 293:617–621.
- Kai T, Williams D, Spradling AC (2005) *Dev Biol* 283:486–502.
- Toczyski DP, Matera AG, Ward DC, Steitz JA (1994) *Proc Natl Acad Sci USA* 91:3463–3467.
- Le S, Greider CW, Sternglanz R (2000) *Mol Biol Cell* 11:999–1010.
- Etter A, Bernard V, Kenzelmann M, Tobler H, Müller F (1994) *Science* 265:954–956.
- Mathews WR, Ong D, Milutinovich AB, Van Doren M (2006) *Development (Cambridge, UK)* 133:1143–1153.
- Potter SS, Brorein WJ, Jr, Dunsmuir P, Rubin GM (1979) *Cell* 17:415–427.