

Materials and Methods

MICROARRAY CONSTRUCTION AND DATA ACQUISITION

cDNA microarrays were constructed with ~9,700 *Drosophila* cDNA elements representing 5,081 different genes. We sequenced the 5' ends of 80% of the cDNAs to verify their identities. The results considered below are for the 4,028 sequence-verified, unique genes. Poor quality spots on the microarrays were filtered from the raw data before analysis. A spot was considered of poor quality if >25% of pixels in both channels were within two standard deviations of background. Sample and array variability was determined by calculating the correlation coefficient between independent hybridizations of the two independently prepared duplicate mRNA samples for each of the experimental points. The median correlation coefficient was 0.89 between the duplicate samples. Ratio data from the microarray experiments were normalized using the Genepix software (<http://www.axon.com/>).

Each hybridization is a comparison of one time point to the reference sample. The reference sample provides a constant denominator for all fluorescence ratio calculations, allowing a comparison of the relative levels of transcripts of each gene in different experimental samples. Normalization is calculated so that the average ratio of signals from the experimental and reference sample equals one. Based on previous studies of RNA complexity across development (1), each time point is likely to have a similar level of complexity, and thus our normalization method will likely avoid distorting the data. All hierarchical clusters were made using Cluster and visualized using Treeview (9), unless otherwise indicated.

SAMPLE COLLECTIONS

Embryo samples were collected as one-hour egg lays which were allowed to develop for the desired interval. At least two pre-lays were performed before each collection. The embryos were dechorionated and snap frozen in liquid nitrogen. Aliquots from each collection were taken prior to freezing and fixed for morphological verification of developmental stage. Only tightly staged embryo collections were used for RNA isolation. Larvae were raised from 6-hour embryonic egg lays, synchronized at hatching, and aged appropriately and stage-verified by morphology.

Pupal samples were synchronized by collection during the white prepupal stage, and aged appropriately. Adult flies were synchronized by collection within the first 24 hours of eclosion (0-24 hour adults). The flies were cultured as a mixed population to the appropriate age, transferring to fresh food every 1-2 days. The adults were harvested by shaking into tubes and immediately snap freezing in liquid nitrogen. The animals were then sorted by sex, on a plastic sheet over dry ice. All developmental samples were prepared using the Canton S wild type strain and were raised and aged at 25°C. For all samples the total RNA was isolated using Trizol reagent (Gibco-BRL) and the PolyA+ was extracted using the PolyATtract mRNA isolation system (Promega). For each experimental sample, polyA+ RNA was prepared from two independent collections of whole organisms, reverse transcribed into fluorescent cDNA probes, and hybridized independently to cDNA microarrays; the average of the two values is presented

DROSOPHILA REFERENCE RNA

Animals were maintained at 25⁰C and approximately 50% humidity. A detailed protocol of how the RNA was isolated is described below. The animals were collected as follows:

EMBRYOS

Three embryo collections were taken:

1. 0-8 hour After Egg Lay (AEL)
2. 8-16 hour AEL
3. 16-24 hour AEL

A small aliquot of each collection was formaldehyde fixed and the rest of the collection was snap frozen in liquid nitrogen. The fixed embryos were used to verify that the range of developmental stages expected was present within each eight hour collection. RNA was isolated from 60g of frozen embryos, as described below. A typical yield of total RNA is ~8-10 mg/1g of embryos, which is approximately 80-100 µg of PolyA+ RNA.

LARVAE

Three larval collections were taken:

1. 24-30 hr AEL (early first instar) 6 hr egg lay from 3 houses. Larvae collected from egg lay plates by rinsing down with water into mesh-bottomed beaker. Larvae were then snap frozen in liquid nitrogen.
2. 57-63 hr AEL (mid second instar) 6 hr egg lay from 3 houses. Embryos collected from egg lay plates by rinsing down with water into mesh-bottomed beaker. Transferred to large tupperware container (1 ft x 2.5 ft) filled with 1 inch layer of macerated fly food. Second instar larvae collected by filtration and snap frozen.
3. 96-102 hr AEL (mid-late third) 20 cages were seeded with 1 ml of 0.06g/ml 0-6hr embryos. Third instar larvae collected by filtration and snap frozen.

The frozen animals were weighed, and transferred to Trizol reagent. 15 g of larvae from each of the three time periods (24-30 hr, 57-63 hr and 96-102 hr) was used for total RNA isolation. Each 15g collection of larvae yielded ~30-45 mg total RNA, which was subsequently incubated with Oligo(dT) beads. A yield of ~1 mg PolyA+ RNA was obtained for each collection. At this point the resultant samples from the first round of PolyA+ isolation were pooled (3 mg) and re-applied to the Oligo(dT) beads. The final larvae polyA+ yield was 2.5 mg.

PUPAE

40 cages were seeded with 1 ml of 0.06g/ml 0-6hr embryos.

Animals were scraped from the walls of 5 cages at 120 hrs, 144hrs., 168hrs., 192hrs., and 216 hrs. and snap frozen in liquid nitrogen and then stored at -80⁰C.

The frozen animals were weighed, and pooled in Trizol reagent. The following masses were used to make the PolyA+ mRNA: 120 hrs 14g, 144hrs 14g, 168hrs 14g, 192hrs 14g, 216 hrs 7g

63g of animals yielded 50 mls at 10mg/ml total RNA. 90mgs of total RNA was used for each batch of polyA+ purification and 2 purifications were performed. A total of 3.8mg polyA+ RNA was isolated.

ADULTS

40 cages were seeded with 1 ml of 0.06g/ml 0-6hr embryos

The cages were cleared on day 9 and animals were collected from 5 cages/ collection starting day 10. The animals were shaken into 50ml conical tubes, snap frozen and stored at -80°C.

Animals were collected on the following days and 15g of animals from each time point was used to make total RNA: day 10, day 11, day 12, day 13, day 14.

A total of 75g of animals yielded 380 mg of total RNA

A total of 1.9 mg polyA+ RNA was isolated.

FINAL MIXTURE OF REFERENCE SAMPLE

The relative complexity of RNA from each stage of the reference sample was assessed by comparing the gene expression profile of each collection to a mini-reference sample that was produced by pooling all stages at an equal ratio. Based on this analysis PolyA+ RNA from each stage was pooled in the following ratio: 6g Embryo RNA (2g from each 8hr sample): 2.5g Larva RNA: 3.5g Pupa RNA: 1g Adult RNA. This mixture was used as a reference sample for all experiments in this study. See below for large scale RNA isolation for reference sample.

ANALYSIS OF VARIANCE TESTS

One-way analysis of variance (ANOVA) tests were performed using the wild type time course data to identify genes that change with time as the independent factor.

ANOVA tests were performed using the one-hour time points from embryogenesis (not the overlapping time points) through the 19-20 hour embryogenesis time point, all larval and pupal time points, and in one test all adult male time points, and in another test all adult female time points. The union of these two analyses was used to identify the set of genes that changed significantly. Numbers of genes that fluctuate significantly during the life cycle, and corresponding estimates of the percentage of genes genome-wide that change, were calculated using p-value thresholds of $p < 0.001$ (3,483 or ~86%, expect 4 false positives), $p < 0.01$ (3,683 or ~91%, expect 40 false positives) and $p < 0.05$ (3,810 or ~95%, expect ~200 false positives). Thus, a minimum of 86% of genes we assayed change.

One-way ANOVA tests were performed to identify genes that changed with time as the independent factor in each developmental stage. It should be noted that the varying number of time points sampled during each stage changes the power of each test. For embryogenesis, all the one-hour time points (not the overlapping time points) through the 19-20 hour embryogenesis time points were included in the analysis (Table S7). All larval and pupal time points were included in the analyses of those stages (see Tables S8-9). Two-way ANOVA tests were performed on all adult time points (with sex and time as the independent factors) to identify genes that change significantly with time (see Table S10). Two-way ANOVA tests were performed on only the 5-30 day time points

from males and females (with sex and time as the independent factors) to identify genes that change after adult development is complete (Table S11).

PEAK FINDING ALGORITHM:

The dynamic range was determined for each gene, defined as the third highest ratio minus the third lowest ratio. The two highest and two lowest ratios were discarded to avoid counting artifactual extremes. A gene was considered activated if the ratio for two successive time points was in the top half of the dynamic range and at least two successive time points prior had ratios below the bottom quarter of the dynamic range. Genes were considered activated at the first time point if that ratio was in the top half of the dynamic range. All onsets, or periods of activation, for each gene were similarly calculated.

HIERARCHICAL CLUSTER BY STAGE

For this analysis adult data from only females was used. Males were excluded so that the analysis would not be affected by the large group of male germ line genes that are expressed throughout the pupal and adult periods.

PROTEIN FUNCTIONAL GROUP ANALYSIS

Protein functional groups were derived from the Gene Ontogeny project. Those genes not yet defined by GO were assigned by the BDGP sequence grouping project. The proportion of genes within a family called “active” or “off” was defined as those in the top or lower 25% of that gene’s dynamic range of enrichment across development.

IDENTIFICATION OF MATERNALLY DEPOSITED GENES

We used a two-step method to determine which genes encode maternally deposited transcripts. First we selected the set of 3,219 genes whose expression changes at least four fold across development (ANOVA $p<0.001$). Second we used a training set of maternal genes to develop a simple criterion that was applied to these 3,219 genes; genes with expression levels $>2/3$ of their dynamic range during the first two time points (0-1.5 hrs) were considered maternal. This criterion was chosen empirically and correctly identified all 31 maternal genes while excluding all 14 zygotic genes from a test set known to be expressed either maternally or zygotically. 1,212 of the 3,219 genes have expression kinetics similar to known maternal genes. This is approximately 38% of the genes tested (1,212/3,219) and approximately 30% of the total genes on the array (1,212/4,028). Thus we estimate that at least 30% of the genes in the genome are maternally deposited.

Maternal genes passed strict maternal expression criteria if (i) their average expression in adults was >2 -fold higher than the average across larval and pupal life (L24-P80), and (ii) their average expression in larval and pupal life always remained less than 50% of the dynamic range of expression across all development. SOM analysis for Fig. 3D and E performed by Genecluster (Maternal gene lists are Tables S12-S17).

IDENTIFICATION OF ZYGOTICALLY ACTIVATED GENES

To identify early zygotic genes based on molecular criteria, we selected genes that are not represented in maternal RNA but whose transcript levels increased at least 2.5-fold above their initial level (at 0-1.5 hrs AEL) during the first 6.5 hours of development. This criterion correctly identified 13 of 14 characterized strict zygotic genes in our test set and excluded 31 of 31 maternal genes. The only zygotic test gene not to pass was *huckebein* (CG9768), which shows the expected early induction, but the magnitude was not sufficient to pass the cutoff (Zygotic gene lists are Tables S18-S22)

Genes satisfied transient early zygotic expression criteria if (i) they had a >4-fold enrichment in average expression during 0-6.5 hr of development versus the remainder of development and (ii) remained below 50% of their dynamic range of development during larval, pupal, and adult life. SOM analysis for Fig. S2C performed by Genecluster (10).

SEARCH PARAMETERS FOR dMef2 BINDING SITES

The 5kb upstream of the predicted gene and first 5kb of intronic DNA were searched for pairs of dMef2 binding site motifs. A pair was defined as two sites within 1,500 bp of each other. Each pair had to contain at least one instance of the canonical sequence YTAWWWWTAG, the other site could be the slightly more degenerate YTAWWWWTAR. One out of twenty of the 4,028 genes represented on the array had at least one pair of Mef2 binding sites. In contrast, 15 out of 23 genes in the muscle cluster (Fig. 4A) contain at least one pair of dMef2 binding sites; this indicates about a 13-fold enrichment for sites in genes within the cluster compared to the other 4,028 genes.

IDENTIFICATION OF SEX-SPECIFIC GERMLINE AND SOMATIC GENES

In addition to the 78 developmental samples described, we also examined transcription in males and females separately at one larval period (105 hour) and one pupal period (48 hours). Inspection of sex-sorted larval and pupal data points demonstrates that the expression of genes in the two germline clusters (Fig 5A) is sex-specific at these early stages, as well.

Additional germline genes were defined as those genes that had a significantly higher expression ($p < 0.01$) in wild type adults of one sex compared to *tudor* progeny adults of the same sex (one-tailed Student's t-Test, See Tables S26 and S27). Because the *tudor* progeny data set is smaller than the wild type data set (2 and 8 time periods, respectively) the statistical power is inadequate to distinguish between elevated somatic and germline gene expression for all 1,685 genes (two-way ANOVA, $P < 0.001$) that differed between adult male and females. A larger *tudor* mutant dataset should allow classification of the remaining genes. Genes expressed in the germline in both sexes (Fig 5A and Table S28 for gene list) were defined as those genes that had a significantly higher expression ($p < 0.01$) in both adult wild type males and females as compared to adult male and female *tudor* progeny, respectively (one-tailed Student's t-test).

Somatic genes were defined as those genes that had a significantly different expression level between adult females and males (two-way ANOVA, $P < 0.001$), a significantly different expression level between female and male *tudor* progeny (one-tailed student's t-test $p < 0.01$) and a ≥ 2 -fold difference in expression level between female and male *tudor* progeny (Tables S29 and S30).

IDENTIFICATION OF EYE GENES

The *eya*² allele used is a viable mutation that results in loss of the compound eye, without affecting the ocelli or other tissues. Many genes encoding phototransduction proteins are expressed both in the compound eyes and the simple eyes, the ocelli, so *eya* mutants may show residual expression of phototransduction genes.

Our two criteria for eye genes were: (1) a >58% average depletion in 0-24 hour *eya* adults relative to similarly staged wild type flies, and (2) that this decrease constituted at least one third of the log₂ range between the wild type 0-24 hour level and the minimum level across the entire wild type time course. If data for both sexes was not available, one sex was used. These criteria identified 8 of 11 known eye genes and excluded all 3 non-eye genes in our training set. The training set was constructed from the 22 of 52 genes annotated in Flybase to be involved in phototransduction that were present on the array. Of these, 6 were excluded, as they were not expected to be eye-specific in their expression, such as the TGFb homolog *dpp* and the insulin pathway gene *chico*.

Fly-Human-BLAST:

Fly genes encoding proteins with a BLAST e-score of 10–50 or lower were classified as having at least one highly conserved human homolog. Genes encoding proteins with a BLAST score less than 10–50 but equal to or greater than 10–5 were classified as having at least one less conserved human homolog. Genes encoding proteins with a BLAST score of greater than 10–5 were classified as fly-specific. Using these criteria, each of these categories contained approximately one-third of the genes on the microarray.

LARGE SCALE RNA ISOLATION FOR THE REFERENCE SAMPLE

1) Total RNA preparation:

Prepare total RNA using TRIzol

Pre-ordering and preparation:

- TRIzol (1ml per 100 mg of tissue (Gibco/BRL 200 ml cat. # 15596-018))
- Centrifuge tubes (VWR cat. # 21020-072 pack of 36)
- 40ml glass on glass homogenizer
- Polytron

Protocol: Follow manufactures instructions with minor modifications

- Thaw the samples on ice in TRIzol.. The embryonic time points were snap frozen in 50ml disposable VWR tubes, the pellets can be tapped into one of the 250ml tubes. Add appropriate about of TRIzol (1ml/100mg) and thaw at RT. (The thawing is

faster if the embryos are not tightly packed before snap freezing). It is a lot quicker to thaw and homogenize in a small volume of TRIzol, about a third the recommended amount, and then increase the volume of TRIzol afterwards.

- Polytron the embryos at full speed for approximately 30 seconds to a minute. (This increased the speed of homogenization which was made more difficult due to the Chorion and high density).
- Homogenize using the tight rod until you cannot see anymore intact embryos, larvae or adults, Add homogenized sample to a fresh 250ml tube.
- Increase the volume of the homogenized embryos to the recommended level with TRIzol and incubate at RT for 5 min.
- Centrifuge at 11,000g (~8,900rpm in a Beckman JA-14) 10 min.
(This removes a lot of the Chorion and Vitelline membranes and makes the interphase of the phenol/chloroform spin more compact)
- Add 0.2ml of chloroform per 1ml of TRIzol, shake vigorously for 15 seconds and incubate at RT for 2-3 min.
- Centrifuge @ 11,000g for 15 min at 4°C
- Remove upper aqueous phase to a new 250ml tube (**be very careful not to touch the interface, it is better to leave some of the aqueous phase near the interface**).
- Add 0.7 volumes of isopropyl alcohol. Incubate at RT for 10 min and then centrifuge @ 11,000g for 10 min at 4°C
- Wash the pellet with 1ml of 75% ethanol/DEPC SDW per 1 ml of the original TRIzol volume. *(75% EtOH: remove 117.5 ml of EtOH from a new pint bottle if 200 proof ethanol and add 117.5 ml of DEPC SDW)*
- Centrifuge at 11,000g for 5 min at 4°C.
- Air dry briefly for ~1 min and resuspend in DEPC SDW, store at -80°C. **Do not over-dry total RNA.** (For resuspending, leave in a 55°C waterbath for ~30 min, mixing by pipeting). The RNA will go into solution at a concentration range of 10mg/ml and should be cloudy.

2) PolyA+ isolation:

Pre-ordering and preparation:

- Oligo (dT)25 cellulose beads (NEB cat# 1408 250mg)

We used it in the ratio of ~15mg total RNA to one order (250mg) of beads and we regenerated once. Therefore to isolate PolyA+ from ~165mg of total we used 6 orders of beads regenerated once.

- Stock Solutions (RNase free, made in baked glassware with DEPC SDW)

DEPC SDW	~ 4 liters total
1M Tris pH 7.5	500ml
0.5M EDTA	500ml
5M NaCl	1000ml
10N NaOH	200ml
3M NaAcetate pH5.2	100ml

- Nalgene Oakridge centrifuge tubes, max. vol capacity 42 ml.
VWR cat# 21009-386, pk of 10, HHMI price \$31.36, Stanford price \$26.88 (in stock)

Protocol:

Minor modifications to the manufactures instructions

<u>Loading Buffer (LB)</u>	<u>2x Loading Buffer</u>	<u>Low Salt Buffer</u>	<u>Elution Buffer</u>
500mM NaCl	1M NaCl	100mM NaCl	10mM Tris
pH7.5			
20mM Tris-HCl pH7.5	40mM Tris pH7.5	20mM Tris pH7.5	1mM EDTA
1mM EDTA	2mM EDTA	1mM EDTA	

- Pool three orders of beads into a 50ml VWR tube (for ~40 mg of total RNA)

(We isolated PolyA+ from 165 mg of total RNA for each time point in two rounds of purification. 82.5 mg of total RNA was added to two 50 ml tubes with three orders of beads in each. After washing and elution of the PolyA+, these beads were regenerated and incubated again with a fresh 82.5 mg of total RNA.)

- Once the appropriate amount of beads have been pooled into a 50ml tube, centrifuge in Beckman CS-GR centrifuge @3,500 rpm, 3-4 min. (*put brake on low)
- Pipette off supernatant and incubate with 30ml LB for 2 min and centrifuge as above.
- Add equal volumes of 2xLB to total RNA and incubate at 65°C for 5 min, cool on ice for 5min
- Add RNA to the pellet of the washed beads and incubate on shaker (gently) for 15min at RT. Centrifuge @3,500 rpm for 4 min
- Take off supernatant into the original RNA tube and heat at 65°C for 5 min, cool on ice for 5min. Re-incubate with the beads on a shaker for 15 min, centrifuge.
- Wash beads 6 times with 30ml of LB
- Wash once with 30ml of low salt buffer
- Elute off PolyA+ by adding 11ml elution buffer that is pre-warmed to 70°C. Incubate at 70°C for 3min. Centrifuge beads
- Pipette the supernant into a new Nalgene Oakridge centrifuge tubes.
- Re-elute beads with a second 11ml of elution buffer at 70°C. Incubate at 70°C for 3min. Centrifuge beads and pipette supernatant into a new Oakridge tube.
(As we were using two 50 ml tubes for each time point, we had 4x 11ml of eluted RNA)
- The Eluted RNA is precipitated in the Oak Ridge centrifuge tubes in the following ratio;
 - 11ml PolyA+ elution
 - 1.1ml of 3M NaAcetate (pH5.2)
 - 26.6ml 100% Ethanol @-20°C (2.2 vol)

- Centrifuge @ 12,000 rpm in a Beckman JA-20 rotor (~18,000g), 20 min at 4°C (*With this amount of RNA the pellet should be visible*)
- Wash with 20ml 75% Ethanol/DEPC, centrifuge @ 12,000 rpm for 5 min. Remove as much ethanol as possible, briefly air dry. (*75% EtOH: remove 117.5 ml of EtOH from a new pint bottle if 200 proof ethanol and add 117.5 ml of DEPC SDW*). **Do not pour off the 75% ethanol as the pellet is loose.**
- Resuspend each tube in 400µl DEPC followed by 1x 100µl DEPC. (Final volume ~2ml from 4 tubes)

Re-generation of the beads:

- Rinse each 50ml tube once with 30ml 0.1N NaOH and then incubate with a fresh 40ml of 0.1N NaOH on a shaker at RT for 1hr. (*This hydrolyses any remaining bound RNA*)
- The beads are washed 3 times with >30ml of DEPC (until the pH of the DEPC is ~7.0)
- Wash once with 30ml of LB

The beads are now ready to purify a second batch of total RNA. (*Use the same batch of used Oligo dT for the same time point*).

- The beads are incubated with the total RNA and washed as described above.
- The RNA from this second batch of PolyA+ isolation is precipitated as before.
- The beads are regenerated by incubating with 0.1N NaOH

In order to increase the PolyA+:rRNA ratio the precipitated PolyA+ is passed through the same re-generated beads a second time, following the protocol above.

*** It is important to precipitate the PolyA+ elution's before it is added onto the beads to be purified a second time. The volume ratio between the beads and the PolyA+ is crucial.**

Re-purification of PolyA+:

- The two 50ml tubes of re-generated beads are pooled into one 50ml tube
- The precipitated PolyA+ (~4ml of DEPC SDW) was incubated in a 15ml tube with an equal volume of 2x LB at 65°C for 5min , ice 5min and then incubated with the beads for 15min on a shaker.
- After centrifugation the supernatant was incubated at 65°C for 5min , ice 5min and then incubated with the beads for 15min on a shaker.
- The beads are washed as described above with 6x 40ml LB, 1x 40ml low salt buffer.
- The re-purified PolyA+ is eluted off the beads in 2x 10ml of LB at 70°C.
- The RNA is precipitated in new Oak Ridge centrifuge tubes in the following ratio;
 10ml PolyA+ elution
 1ml of 3M NaAcetate (pH5.2)
 24.2ml 100% Ethanol @-20°C (2.2 vol)
- Centrifuge @ 12,000 rpm in a Beckman JA-20 rotor (~18,000g), 20 min at 4°C (With this amount of RNA the pellet should be visible)

- Wash with 20ml 75% Ethanol/DEPC, centrifuge @ 12,000 rpm for 5 min. Remove as much ethanol as possible, briefly air dry.
- Resuspend in 2x 200 μ l DEPC followed by 2x 100 μ l DEPC. Final volume ~600 μ l. Remove a small aliquot for A₂₆₀ reading, store the RNA at -80°C.
- Regenerate beads as described above and store in LB at 4°C.

Each of the co-first authors contributed substantially to every aspect of the project, from sample collection to analysis. The following individuals were primarily involved and may be contacted for further details on the following aspects: B.H.N., construction of microarrays and data collection; E.J., global analysis; F.I., cell biology, maternal zygotic and human disease analysis; E.F., muscle analysis and M.N.A., sexual dimorphism of gene expression.

Web Supplemental Figure 1. Patterns of gene use and reuse. **(A)** The left column shows three simple transcript profiles. CG5958 is a gene of unknown function whose induction occurs in early embryogenesis and is maintained. The transiently induced genes, such as CG1733, have short peaks of intense expression and are not expressed at other points in development. CG0159 (BEAF-32) is an example of a transiently repressed gene. Its expression is sharply reduced following fertilization, and increases again almost immediately. The center column shows three examples of bimodal, or reinduced, genes. Each of these genes is a representative example of a large group of genes whose transcripts rise and fall in a similar fashion. Their induction is staggered; the peak of the early gene (*Amalgam*) is in early embryogenesis and at the larval/pupal transition, while the intermediate gene (*igloo*) and the late reinduced gene (CG17814) are each induced slightly later. The third column shows several genes with unusual patterns of expression. CG6611 has three sharp peaks of expression, at the E/L boundary, L/P boundary, and the prepupal-pupal boundary but, surprisingly, not at the P/A boundary. The other two genes also show unusual patterns of expression. **(B)** A subset of genes from the early embryonic-early pupal group (blue bracket in (Fig 1C)) and the late embryonic-late pupal group (purple bracket in (Fig 1C)) are shown ordered by hierarchical clustering. Genes in the latter group that encode components of terminally differentiated muscle are highlighted. Line plots of examples from both classes are shown in the top and bottom graphs in the central column of Web Supplemental Fig. 1A.

Web Supplemental Figure 2. Maternal and Zygotic genes **(A)** Selected clusters generated from a SOM analysis of the 322 rapidly degraded maternal genes in (Figure 2B). A simple pattern of gradual decline encompasses 49 genes, among which are several components of DNA replication (blue) and DNA repair (red). A rapidly degraded and reinduced group of 26 genes is shown, among which are *BEAF-32* (black) and *thiolase* (green). Of special interest is a SOM cluster enriched for genes present in both male and female germlines (see Fig. 5A). This group of 24 genes contains *exuperantia* (red), the germline sex determination gene *stand still* (green), and a ubiquitin hydrolase (black). **(B)** A subset of 36 genes from (Fig 2B) showing >10-fold reduction in expression over the first 0-6.5 hr of development. Individual gene traces are shown, with

specific genes highlighted in the panel. The gene, CG18543 (black), with the most dramatic reduction in expression measured is of unknown function, and the most rapidly degraded is *Nacalp*ha (blue). Genes passing the cutoff due to a single aberrant data point were discarded by eye. Other genes highlighted include *bruno* (green), *exuperantia* (red), *BEAF-32* (pink). (C) Three patterns of gene expression changes among the early zygotic genes. The three clusters were identified using an SOM analysis. The induced and maintained cluster includes a predicted phosphatase CG3239 (black) among its 33 genes. The reinduced cluster contains many neuronal genes (green) and cell adhesion genes (red). The male germline cluster is composed of 30 unknown genes, including CG7886 (black) and CG11703 (pink). (D) A subset of 53 genes from (Fig 2E) which are induced >10-fold between 0 and 6.5 hr of development. Sample genes are highlighted, including CG15634 (black), which encodes a novel protein and has the earliest onset of expression of all genes analyzed. Other genes highlighted: *tailless* (blue), *odd paired* (pink), *wingless* (red), *Antennapedia* (green).

Web Supplemental Figure 3. Coordinate regulation of components of macromolecular complexes and physiological processes across development. (A) Specific clusters of genes involved in particular biological functions, including gene sets encoding components of protein complexes like the ribosome, proteasome, and COP.

Web Supplemental Figure 4. Developmental expression profiles of known muscle genes. A comprehensive search of the literature identified 12 genes that are known to be involved in *Drosophila* muscle contraction and are present on our arrays. The developmental expression profile for all 12 genes is shown. Five of these genes, indicated in red, were contained within a cluster of 23 genes (known as the muscle cluster, Fig 4A in the paper) due to the high similarity in their expression profile throughout both larvae and adult muscle development. When the correlation coefficient of this cluster is reduced from 0.861 to 0.834, the size of the cluster is increased from 23 to 38 genes, and includes *bent* (*Drosophila* projectin). The other 6 of the 12 known genes did not group with the muscle cluster as their expression profile differed during either the embryonic-larval time points or the metamorphosis-adult time points. For example the

gene *flap wing* is strongly enriched during larval life and is not highly expressed in adults, while *flightin* is not expressed during embryo or larval life and is strongly enriched during the end of metamorphosis and in adults. This suggests that larval and adult muscles share some common proteins that are expressed during both stages of muscle development and, in addition, also express some genes that are specific to either the larval or adult muscles, reflecting the unique functions of these two muscle types.

Web Supplemental Figure 5. RNA Blot analysis of 4 germline genes.

(A) RNA blot analysis of 2 male and 2 female enriched germline genes. The blots contain polyA mRNA from 0-5 day male and female adults, derived from two wild type strains (Canton S [CS] and Oregon R [OR]) and adult *tudor* progeny (TUD). RNA probes were derived from Berkeley Drosophila Genome Project (BDGP) ESTs: GH26310 (CG12423, Kelch-related), GH07062 (CG5555, BRCA-1 associated), LD22180 (CG17252, BCL-7 like), LD37603 (CG1962, paraneoplastic antigen-related [PCD]). The ethidium bromide stained gels are shown beneath the blots.

Web Supplemental Figure 6. Distribution of human homologs and fly-specific genes in gene expression clusters. BLASTP was performed with the predicted translations of each gene represented on the microarray and all predicted human proteins. The ratio of human homologs to fly-specific genes for each cluster was calculated by $\log_{10}(H/F*C)$, where H = the number of genes with highly conserved human homologs (BLAST scores $<10^{-100}$), F = the number of fly-specific genes (BLAST scores $>10^{-5}$), and C is a correction factor (0.87) to account for the relative proportion of fly-specific sequences to highly conserved sequences present on the array. The numbers of highly conserved genes and fly-specific genes present in each cluster are available at <http://flygenome.yale.edu/Lifecycle>. To focus on the comparison between highly conserved genes and fly-specific genes, genes with intermediate BLAST scores (10^{-100} to 10^{-5}) were excluded. A positive value represents an excess of human homologs while a negative value represents an excess of fly-specific genes. Chi square analysis was performed to determine if the number of observed highly conserved genes, or fly-specific genes, deviated from the expected number: *** p<0.001, ** p<0.01. Only those clusters with enough genes to perform a Chi Square analysis are shown (20 out of 49). The overall distribution of highly

conserved verses fly-specific genes in clusters appears nearly random, although there are notable exceptional clusters. The identities of the genes in clusters showing significance are enriched with the following classes of genes: ribosomal genes (H=36, F=3), zygotically expressed pattern formation genes (H=24, F=5), male germline genes (H=77, F=100), and genes encoding peptide hormones, peptidases, peritrophins or cuticle proteins (H=3, F=12).

Web Supplemental Figure 7. Overview of gene expression and analyses. All 4028 sequence-confirmed genes are shown with results from all developmental time points and mutant animals. Genes have been hierarchically clustered; groups of genes with similar expression patterns have been demarcated with green, yellow and red bars (correlation coefficients of 0.8, 0.9 and 0.95). Many of the analyses presented are summarized here in the columns to the right of the gene expression data. Genes that pass the criteria for a given column's condition are marked as a colored line. Particular clusters of interest discussed in the text are shaded light blue underneath all columns.

Web Supplemental Tables 1-22 and 24-30

Gene lists for analyses presented.

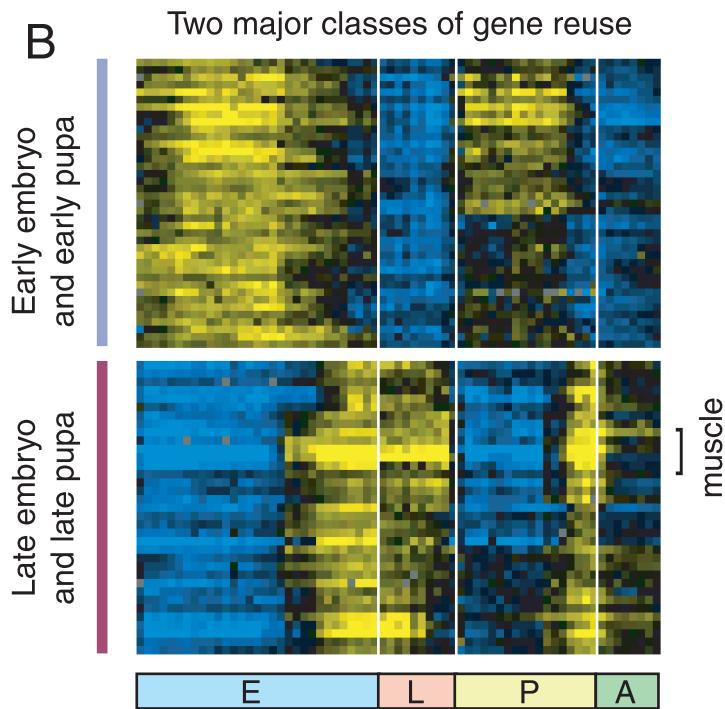
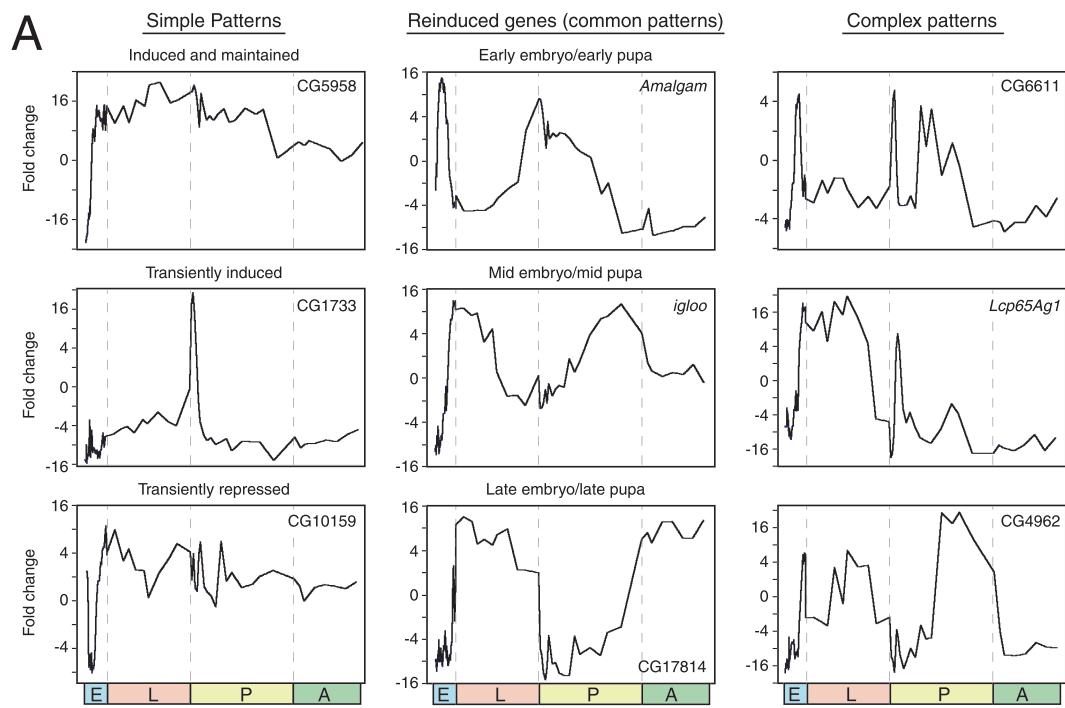
In the first column the computed gene identifier is shown. In the next column information about each spotted DNA is shown as : a "+" to indicate the DNA was re-sequenced, the gene name abbreviation, the protein functional group, BLAST hits (demarcated by asterisks), predicted protein domains (enclosed by brackets), computed gene identifier, EST identifier, cytological position, which instance of genes with multiple spotted DNAs on array, and our unique identifier.

Web Supplemental Table 23. dMef2 binding site analysis and in situ data summary for genes within the muscle cluster. The muscle cluster contains 23 genes that have tightly co-ordinated expression throughout development. The genomic region of all 23 genes were analyzed for pairs of dMef2 binding sites. In situ hybridization analysis was performed on 9 genes of unknown function within the cluster; 7 contained dMef2 binding sites and 2 did not. 6 out of 7 of the genes with dMef2 binding sites are expressed in muscle. The 7th gene, CG8256, that contained dMef2 binding sites, as well as the two

genes that did not contain obvious binding sites (CG9098 and CG7565) are expressed in the nervous system. The EST that was used for the *in situ* is indicated in brackets after the gene's CG name.

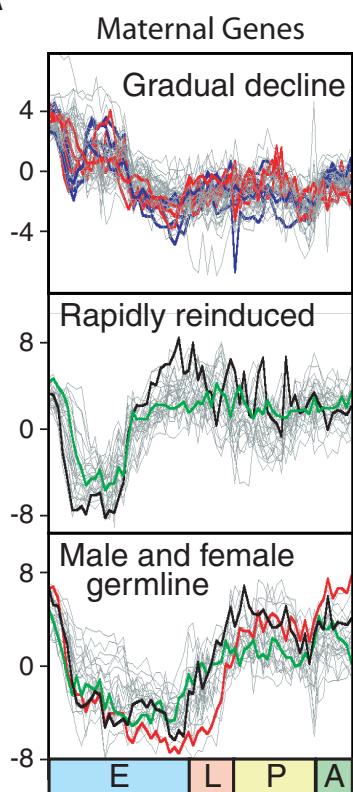
(Note: CG9432 is expressed in somatic muscle, personal communication with David Micklem, CG6020 is expressed in pharageal muscle, the rest of the 'expressed in muscle' genes are expressed in the somatic and/or visceral muscles).

Web Supplemental Figure 1

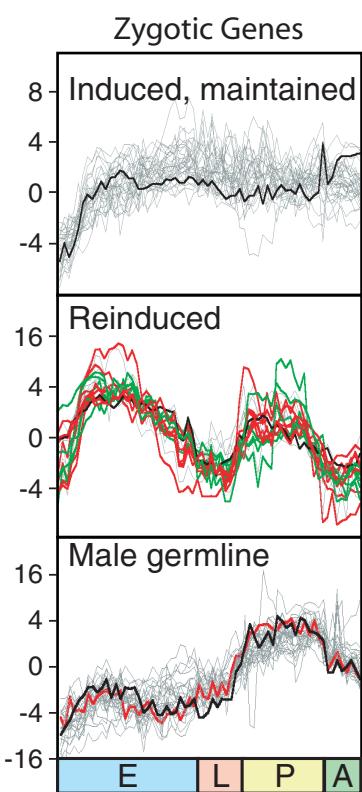


Web Supplemental Figure 2

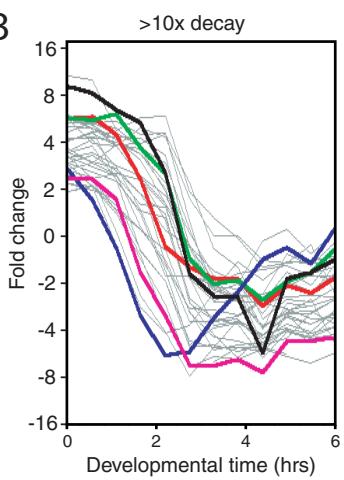
A



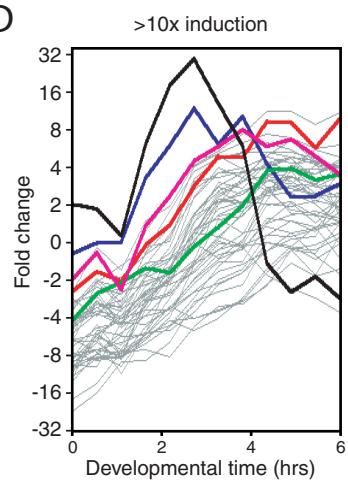
C



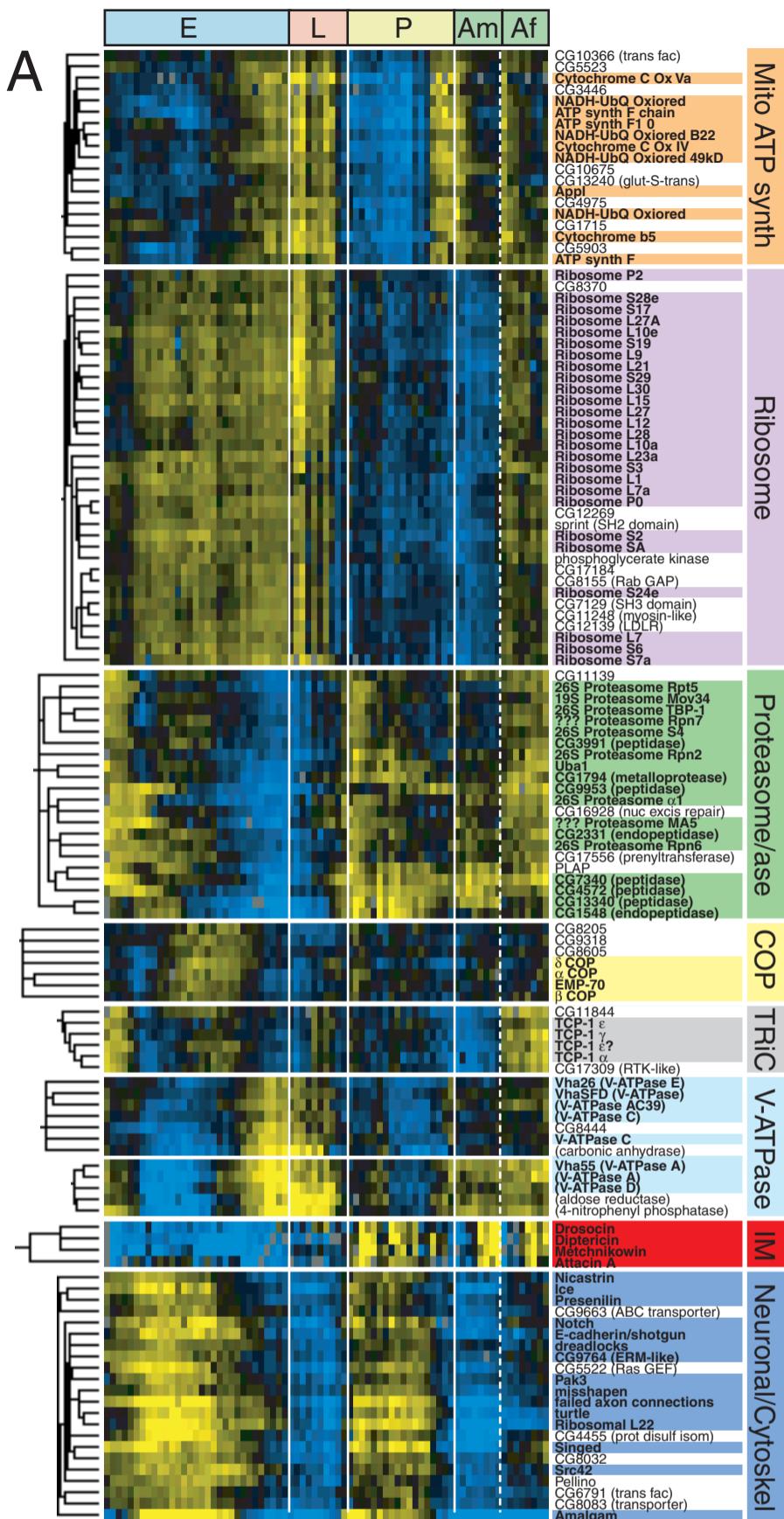
B



D

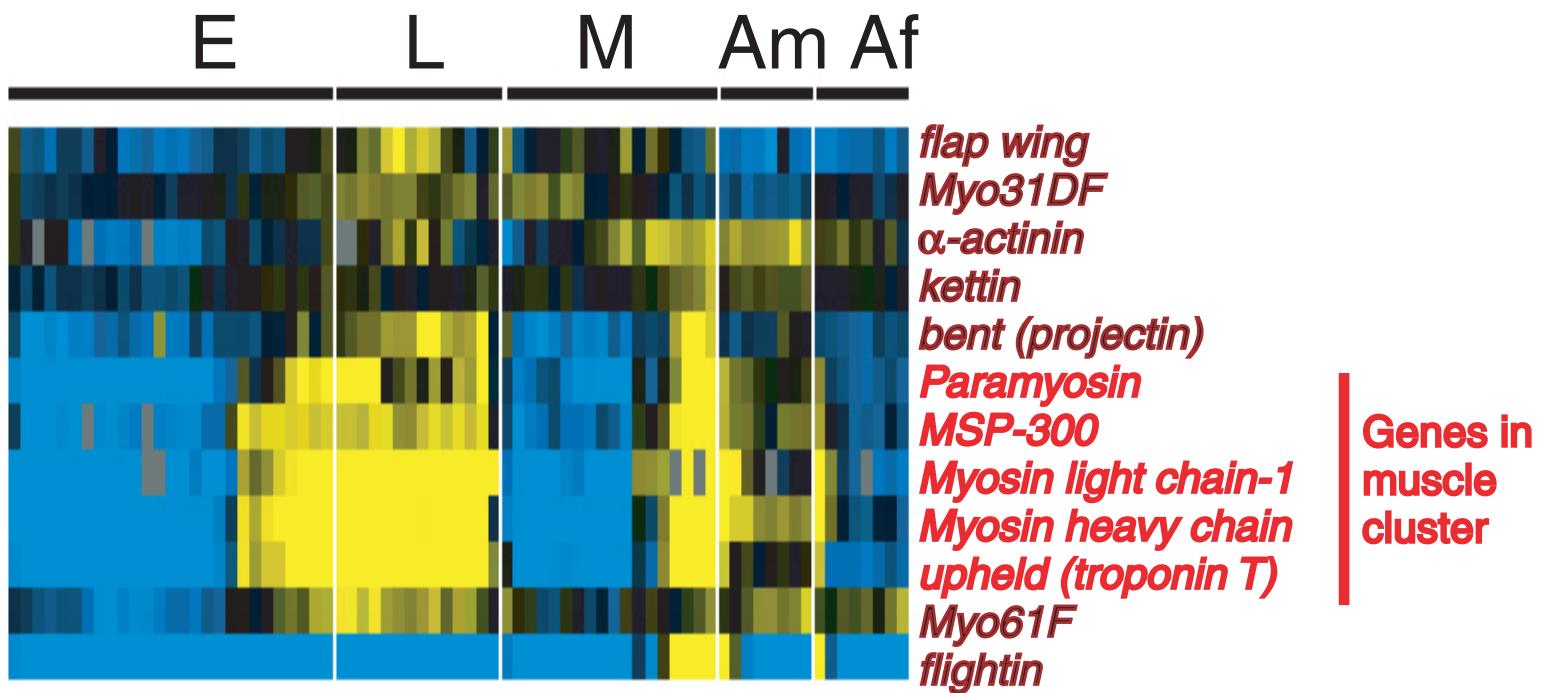


Web Supplemental Figure 3



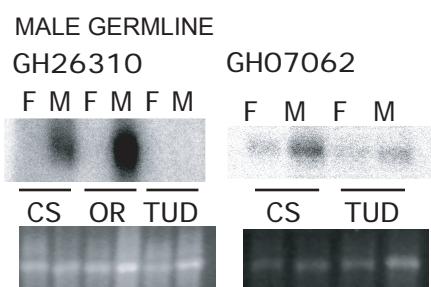
Web Supplemental Figure 4

Expression profiles of known muscle genes

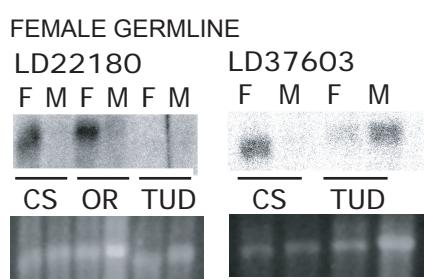


Web Supplemental Figure 5

A

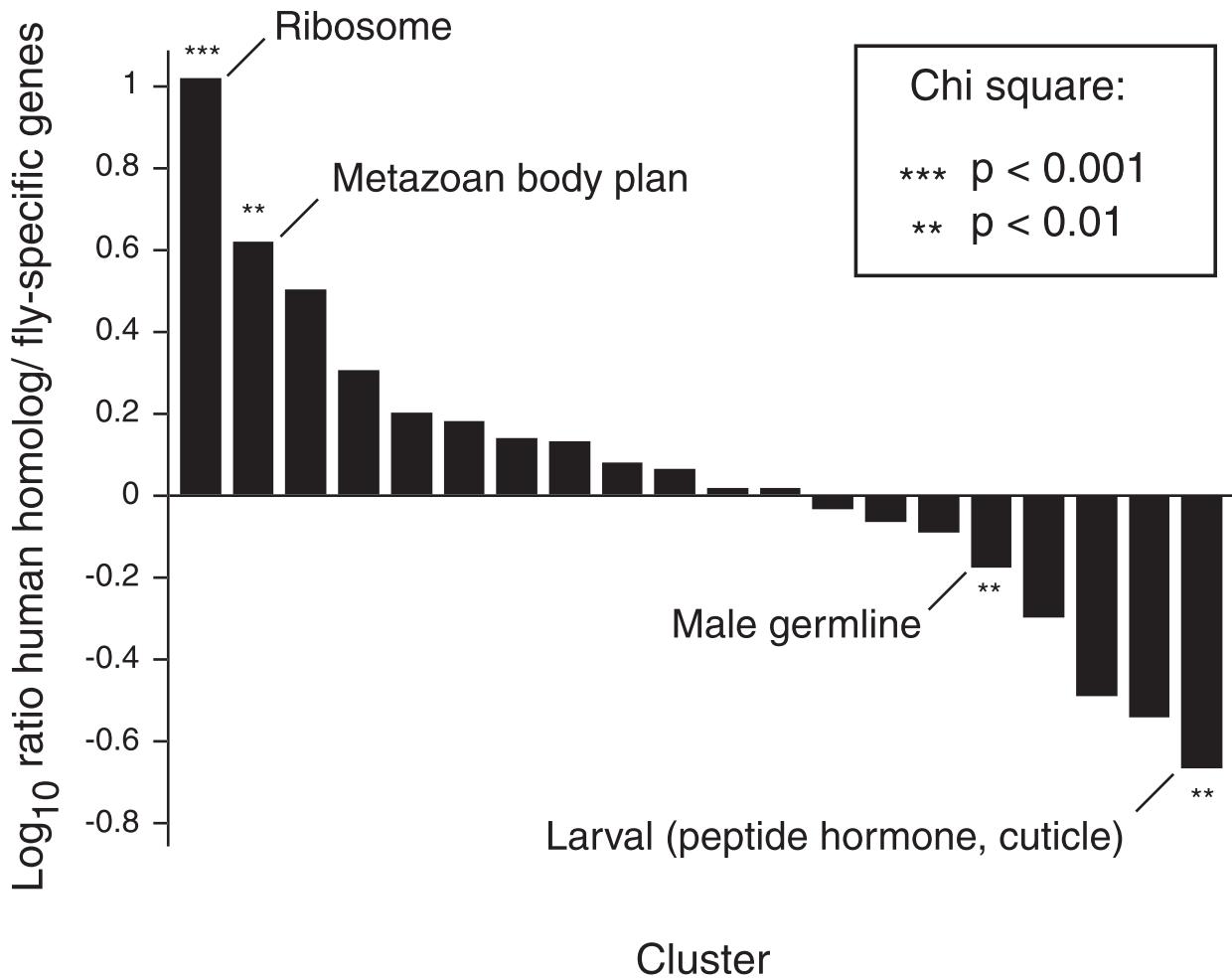


B



Web Supplemental Figure 6

Human-fly gene conservation within clusters of high correlation



Web Supplemental Figure 7

