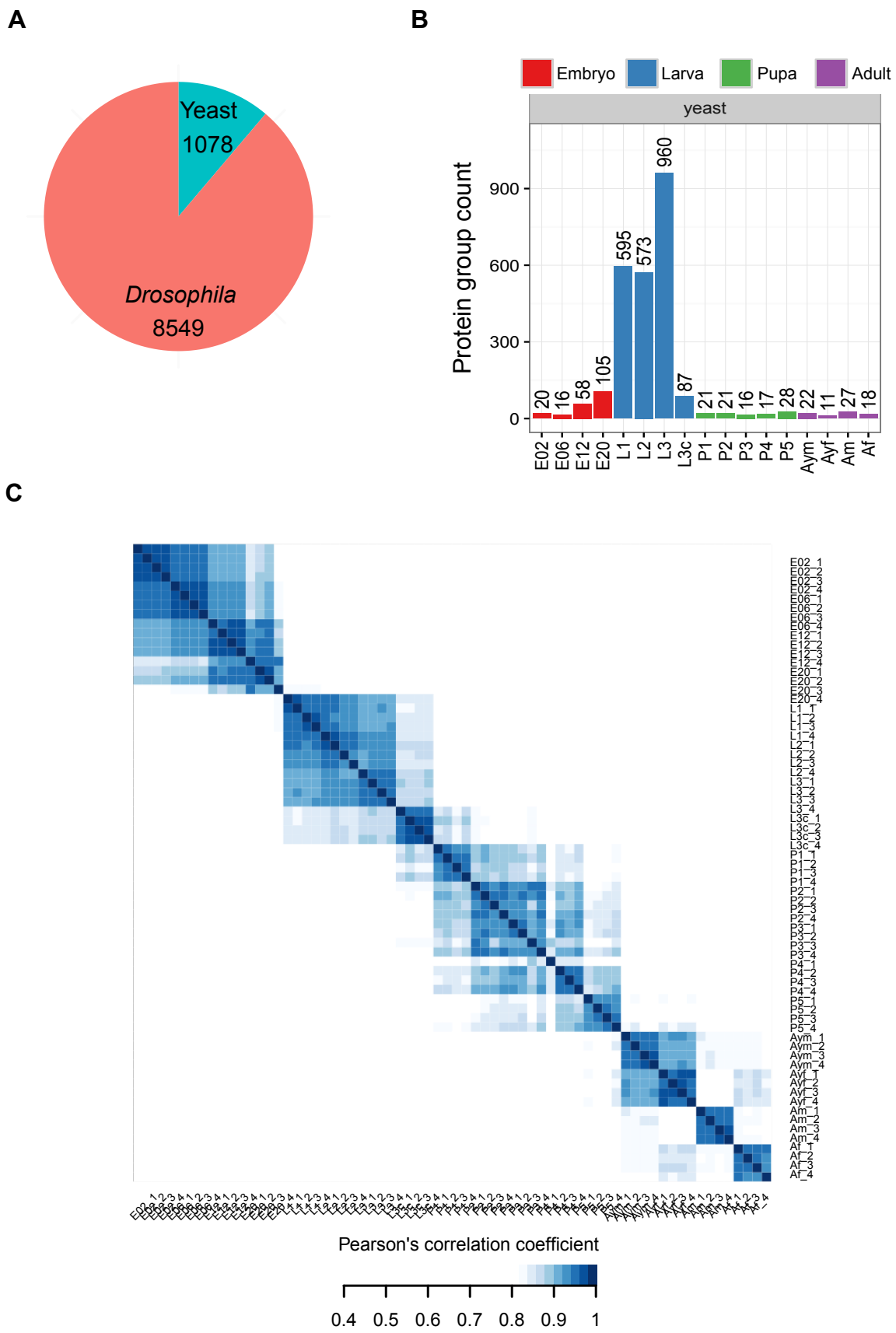
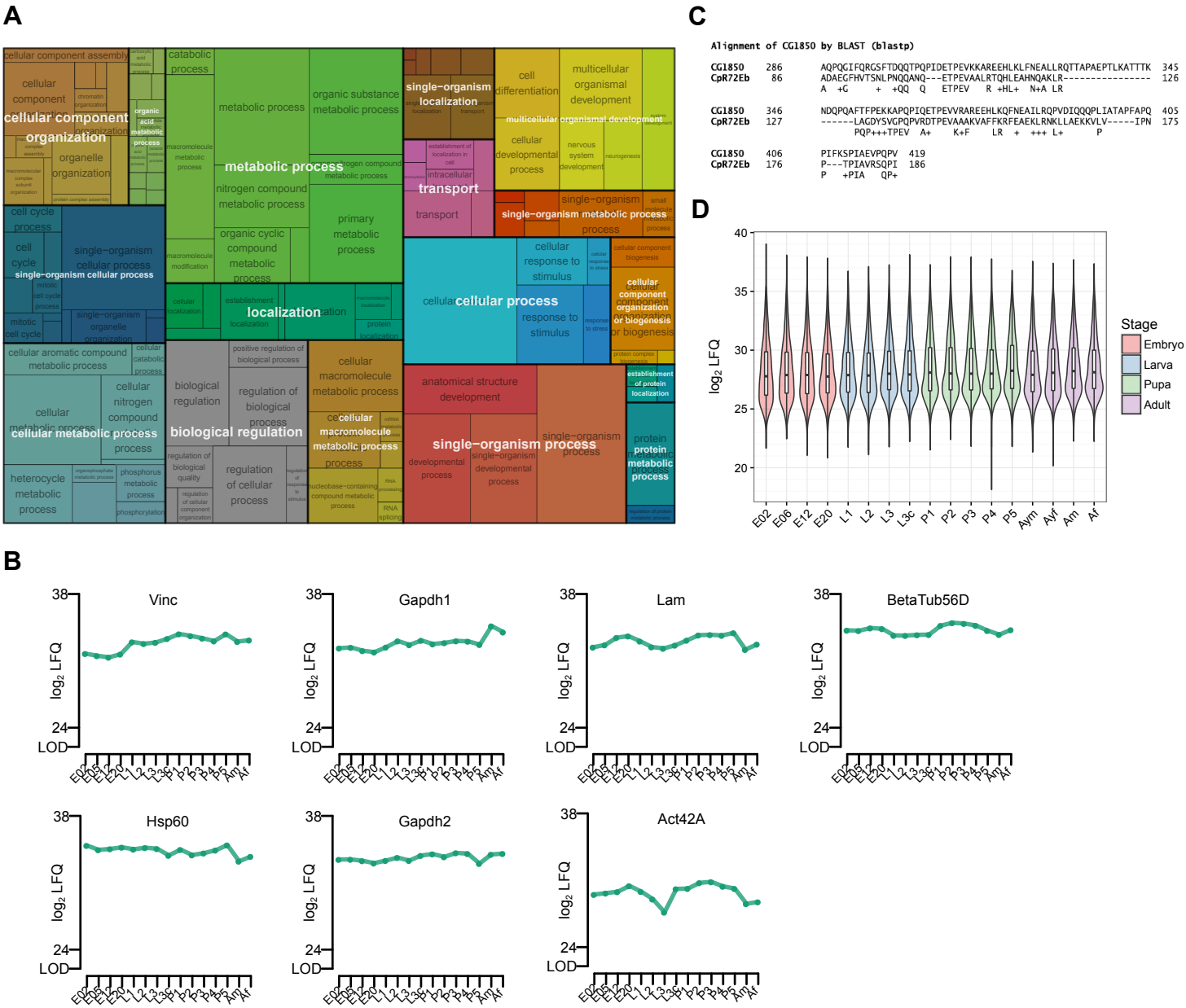


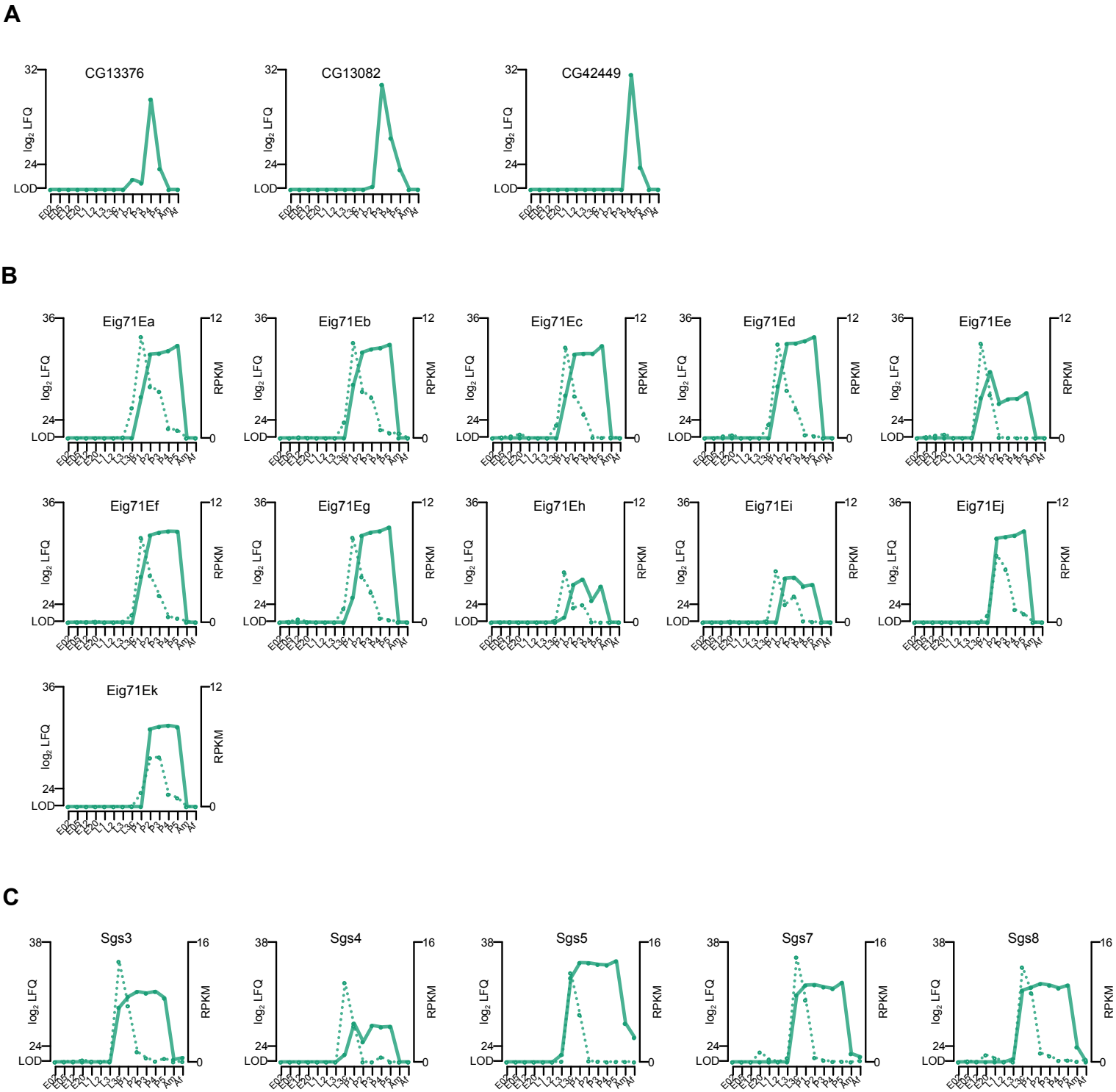
Supplemental Fig. S1:

(A) Pie chart of detected yeast and *Drosophila* protein groups. Yeast proteins represent 12 percent (1078) of the identified proteins in the screen (9627). (B) Nearly all yeast proteins are detected in feeding L1, L2 and early L3 larval stages. (C) Correlation plot for the 17 samples of the life cycle experiment with all four replicates.



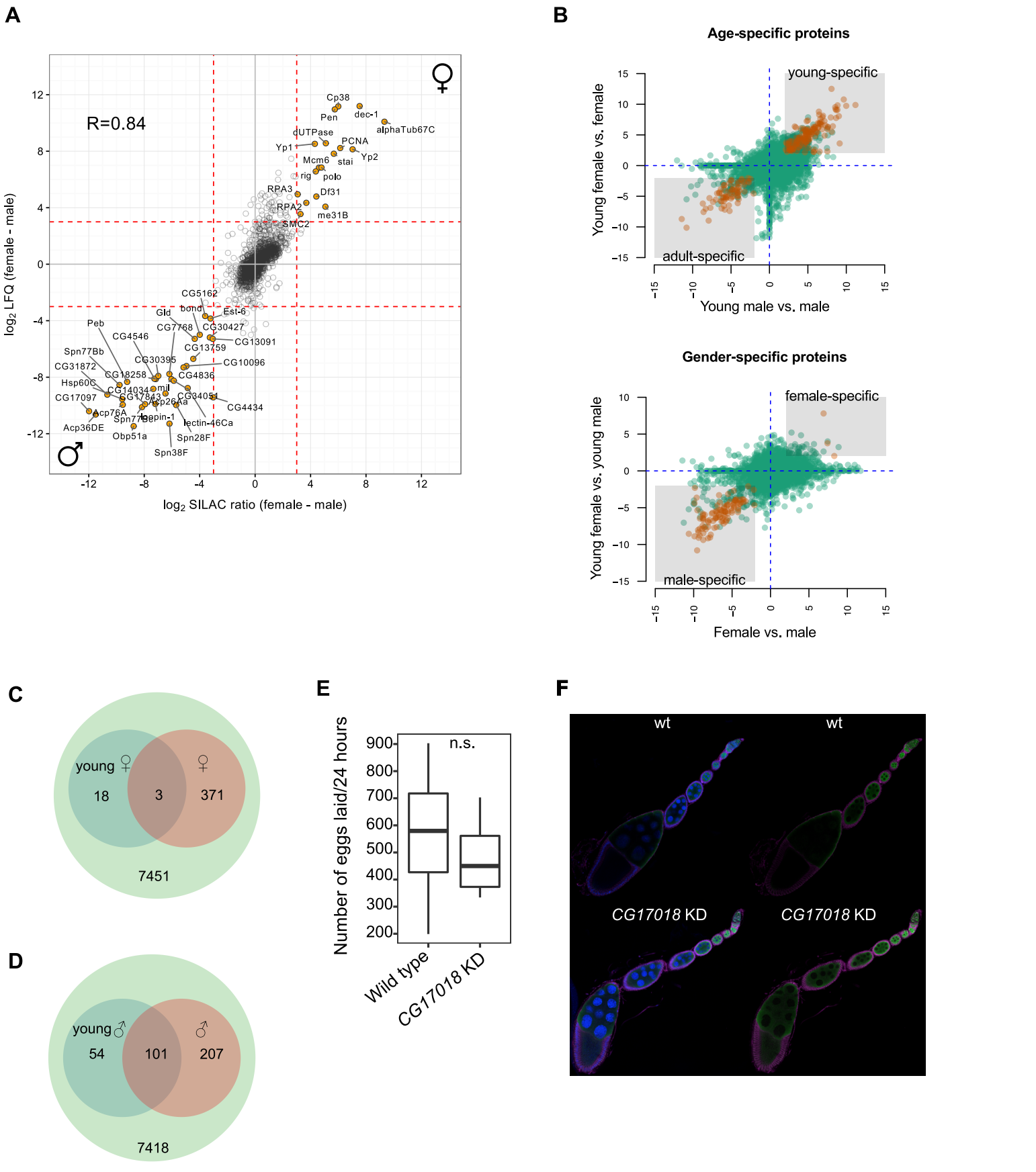
Supplemental Fig. S2:

(A) Treemap with overrepresented GO terms. The area is proportional to the size of the GO terms. (B) Alignment using BLAST between CpR72Eb and CG1850 shows a stretch of moderate sequence similarity between both proteins. (C) Overall protein expression values based on LFQ quantitation shows similar abundance levels and distribution across all conditions of the experiment. (D) Expression profiles of stably and dynamically expressed proteins annotated in Figure 2C and not displayed in Figure 2D.



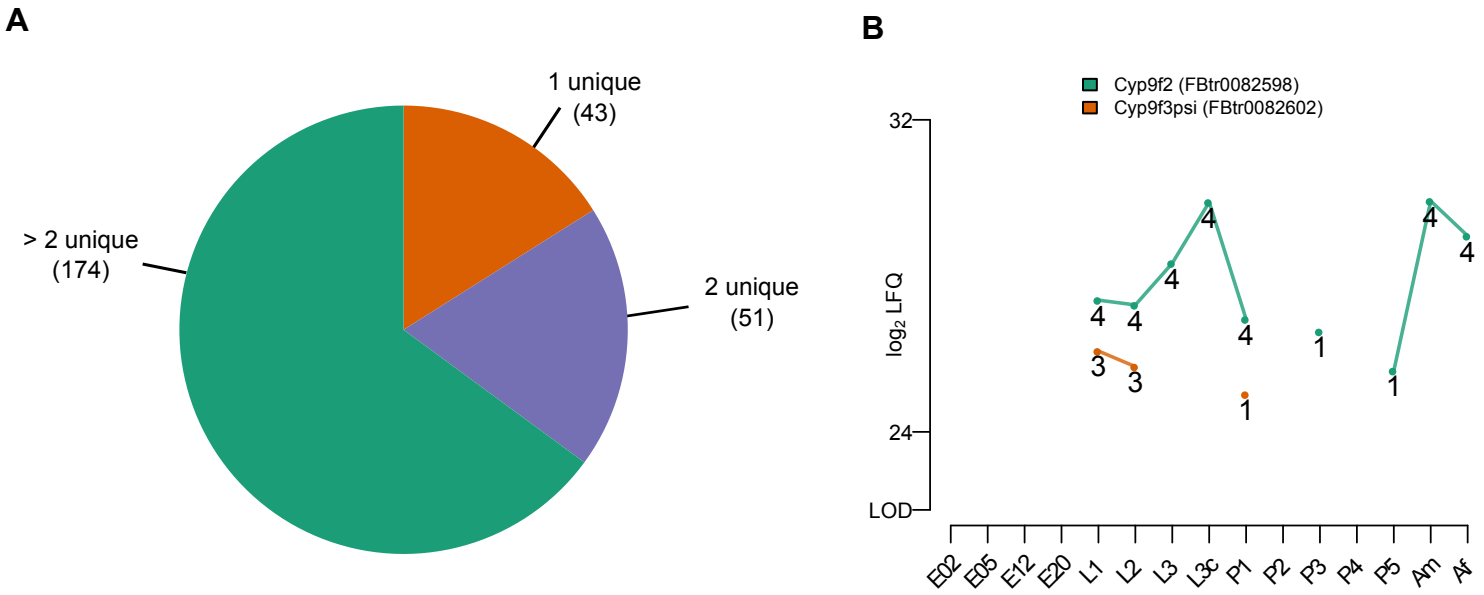
Supplemental Fig. S3:

Expression profiles of uncharacterized proteins upregulating at a single pupal stage (A). RNA (dotted line) and protein (solid line) expression profiles of all quantified Eig72E family proteins (B) and Sgs proteins (C).



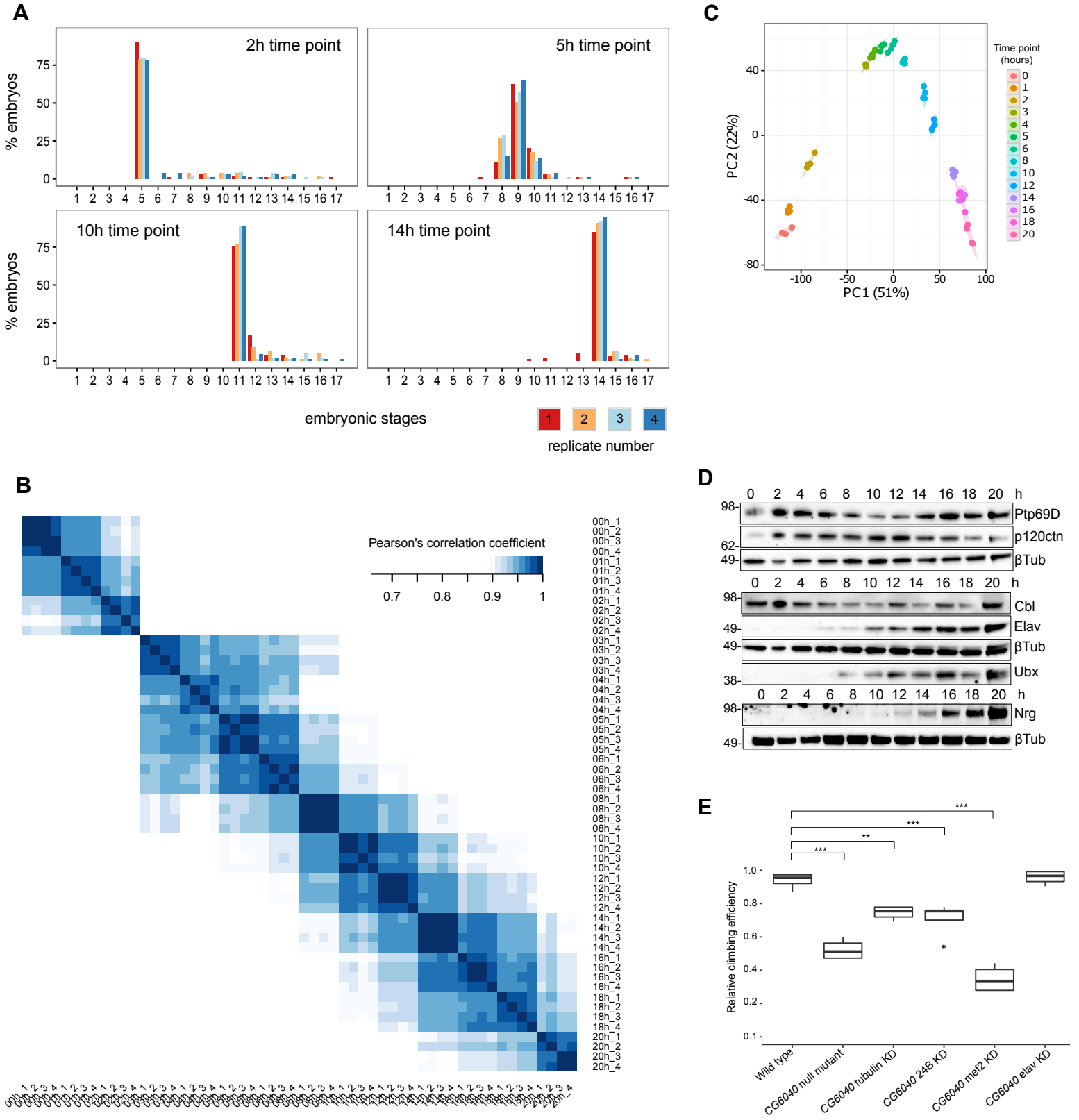
Supplemental Fig. S4:

(A) Comparison of w1118 SILAC data (Sury et al., 2010) to our Oregon R strain LFQ data. More than 8-fold enriched protein groups are highlighted (orange fill) and annotated. (B) Scatter plot showing age-specific and gender-specific proteins. (C) Venn diagram to compare female-specific proteins already present in the young fly. (D) Venn diagram to compare male-specific proteins already present in the young fly. (E) Number of laid eggs from wild type and the *CG17018* knockdown line. (F) Dissected ovaries from 5 day old female flies stained for DAPI (blue), Vasa (green) and 1B1 (purple).



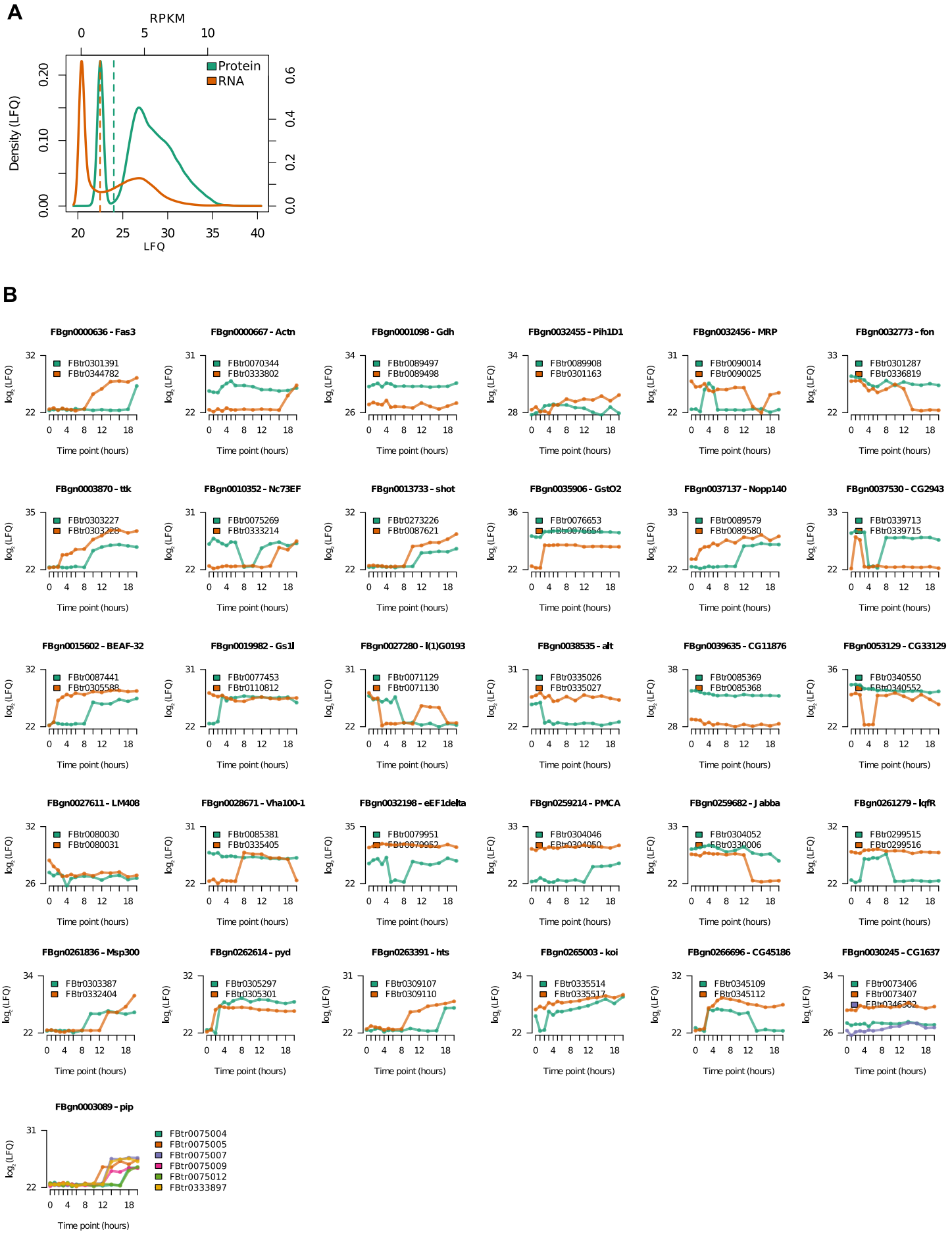
Supplemental Fig. S5:

(A) Pie chart showing the distribution of small proteins identified by the number of unique peptides. For a large majority (>80 percent) of the small proteome more than a single unique peptide was measured. (B) Developmental expression profile for the expressed previously annotated pseudogene Cyp9f3Psi in comparison to its paralog Cyp9f2.



Supplemental Fig. S6:

(A) Histogram showing the distribution of embryo stages in the four replicates of selected staged time points with a representative picture. (B) Correlation analysis of the 500 most variant proteins show high correlation between individual biological replicates. (C) PCA analysis with the first two components. Replicates are shown in identical colors and the standard error is represented by elliptic areas. (D) Westerm blot validation of chosen protein profiles in embryogenesis shown in Fig. 6. The respective tubulin loading controls are shown. (E) Ubiquitous (tubulin driver) and mesodermal (24B and mef2 driver), but not neuronal (elav) knockdown of *CG6040* results in reduced locomotion activity (Dunnett's test; ** p-value <0.005, *** p-value < 0.001). (F) *in situ* hybridization of *CG6040* and *CG1674* showing expression in muscle tissue.



Supplemental Fig. 7:

(A) Density plot of transcriptome (RPKM) and proteome (LFQ) quantitation shows bimodal distribution. The cut-offs are chosen to separate the two maxima (dashed lines). (B) Profile plots of the remaining 31 genes (not shown in Figure 7D) with isoform quantitation on protein level during the time course experiment.

Supplemental material

The developmental proteome of *Drosophila melanogaster*

Nuria Casas-Vila, Alina Bluhm, Sergi Sayols, Nadja Dinges, Mario Dejung, Tina Altenhein, Dennis Kappei, Benjamin Altenhein, Jean-Yves Roignant and Falk Butter

Supplemental Methods

Collection of embryos, larvae, pupae and adult flies

After collection, all embryo samples were dechorionated using 7.5% hypochlorite for 2 min and rinsed with water. For each time point, approximately 20 µl embryo pellets were transferred to PBS buffer for lysis and mass spectrometry measurement. To assess the homogeneity of embryonic stages of each collection, approximately 10% of the sample was fixed and staged. Unused samples were snap-frozen in liquid nitrogen and stored at -80°C. Early larval collections were performed from agar apple juice plates in 2 hour laying time windows. Crawling larvae and pupae stages were collected from flasks at respective time points and rinsed with water.

Fixation and antibody staining for embryo staging

A small fraction (approx. 10%) of each embryo collection during the time course was fixed in fixation buffer (450 µl PBS, 600 µl heptane, 70 µl 37% formaldehyde) for 20 min while agitating. To remove the vitelline membrane, heptane was exchanged for methanol and embryos were vortexed for 2 minutes. Embryos were washed several times in methanol and finally snap-frozen in liquid nitrogen. Fixed embryos were rinsed three times in PBT (0.1% Triton-X-100 in PBS) for 10 min and incubated with the 4D9-engrailed/invected-s antibody (1:7 in PBT, Developmental Studies Hybridoma Bank (DSHB)) at 4°C while slowly agitating. Embryos were washed three times in PBT for 10 min and incubated with the anti-mouse antibody conjugated with alkaline phosphatase (AP) (1:250 in PBT, Jackson ImmunoResearch) for 2 hours at RT. Three washes in PBT were followed by a 5 min wash in AP detection buffer (0.1 M NaCl, 0.05 M MgCl2, 0.1 M Tris pH

9.5, 0.1% Tween20). The AP staining solution (150 µg/mL nitro blue tetrazolium (NBT) and 75 µg/mL 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) in AP buffer) was added, embryos were transferred to a small dish and the color reaction was monitored using a binocular. To stop the AP reaction, embryos were rinsed three times in PBT and incubated for 10 min in methanol. After three PBT and three PBS washes, embryos were stored in 1 mL 70% glycerol at RT. Staging was done according to morphology and antibody staining.

Mass spectrometry sample preparation

For proteome analysis of the whole life cycle, snap-frozen samples were mechanically lysed in lysis buffer (140 mM NaCl, 10 mM Tris-HCl, pH 8.0, 0.5 mM EDTA and 1x protease inhibitor (Roche)) by bead milling using 0.5 mm diameter zircona/silica beads (Carl-Roth). Bead milling was done three times for 30 sec at 6800 rpm at 4°C in a tissue lyzer (Precellys). Homogenates were collected by centrifugation and proteins were acetone precipitated. Protein pellets were resuspended in 1x NuPAGE sample loading buffer (1x LDS) supplemented with 0.1 M DTT, boiled for 10 min at 95°C, sonicated for 10 min and proteins were separated on a 4-12% NuPAGE Bis/Tris gel for 10 min at 180 V in MOPS buffer.

For the embryonic time course analysis, embryos in PBS were homogenized with a microtube pestle, cells were pelleted at 1000 x g for 5 min at 4°C and resuspended in 1x LDS buffer complemented with 0.1 M DTT. Samples were boiled for 10 min at 80°C and proteins were separated on a 4-12% NuPAGE Bis/Tris gel for 10 min at 180 V in MOPS buffer. In-gel digestion and MS analysis was done as essentially described (Kappei et al. 2013).

Western blotting

100 embryos were homogenized in 50 µl of lysis buffer (140 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 0.5% Triton X-100) using a microtube pestle (8-10 strokes). After 30 min incubation at 4°C on a rotation wheel, the lysate was centrifuged at maximum speed for 20 min and the supernatant transferred to a new tube. The previous two steps were repeated four times until the lysate was clear. Total embryonic protein extract was separated by SDS-PAGE. Western blot

analysis with affinity purified anti-Lola antibody (1:500, kindly provided by Edward Giniger) was performed by standard methods and visualized using ultra-sensitive enhanced chemiluminescent reagent (Thermo). Anti-β-Tubulin antibody (Covance catalog #MMS-410P, BioLegend) was used at a 1:2000 dilution. Additionally, Lola-RAA/RI depleted flies were generated as previously described (Kondo and Ueda 2013) using a gRNA (GTGTTGCACGTAAAGAAGCT) in exon 21 leading to a 2-bp deletion (Chr2R:10510060-10510061).

Embryo samples of selected time points (0h, 2h, 4h, 6h, 8h, 10h, 12h, 14h, 16h, 18h, 20h) prepared for mass spectrometry were used for western blot validation of protein expression profiles during embryogenesis. Proteins were separated by SDS-PAGE and blotted onto a nitrocellulose membrane, and subsequently blocked with 5% milk powder in TBST (0.1% Tween20 in 1xTBS). The following antibodies from DSHB were used in a 1:50 dilution (5% milk powder TBST): anti-p120ctn (p1B2), anti-Nrg (BP 102), anti-Ubx (kindly provided by Christian Berger), anti-PTP69D (3F11), anti-Cbl (8C4). Anti-Elav (7E8A10, DSHB) raised in rat was used 1:200 in 5% milk powder TBST.

in situ hybridization

Primers were designed to amplify a unique region within respective coding sequences using a reverse primer containing the SP6 sequence. The PCR was performed on embryonic cDNA using Phusion DNA polymerase (NEB). Amplicon size comprised 861 bp for *lola*-RAA/RI, 1095 bp for *CG6040* and 1070 bp for *CG1674*. 250 ng of template PCR product was used to perform *in vitro* transcription using the SP6 RNA polymerase (Roche) and DIG labeled UTP (Roche). The reaction was incubated over night at 37°C and probes were carbonated to approximately 500 bp using carbonate buffer. The probes were then ethanol precipitated and resuspended in DEPC treated water to obtain a concentration of 100 ng/µl. Probes were diluted 1:50 in hybridization buffer for *in situ* hybridization. Embryos were fixed for 25 min in fixation solution (400 µl PBS, 500 µl n-heptane, 100 µl 37% formaldehyde) while shaking at RT. After washing in methanol several times the embryos were snap-frozen in liquid nitrogen and

stored at -80°C. Embryos were gradually transferred into PBT (0.1% Tween20 in PBS), followed by three washes for 15 min, and finally into HB4 hybridization buffer (50 ml formamide, 25 ml 20xSSC buffer (3 M NaCl and 0.3 mM trisodium citrate-HCl pH 7.0), 200 µl Heparin (50 mg/ml), 100 µl Tween20, 500 mg Torula Yeast RNA extract). After equilibration at RT, embryos were pre-hybridized in HB4 at 56°C for several hours. Upon denaturation of the diluted RNA probe at 80°C for 10 min, embryos were hybridized over night at 65°C. Embryos were subsequently incubated in washing buffer (formamide:2xSSC (1:1) and 0.1% Tween20) for 30 min at 65°C and transferred into PBT at RT before they were incubated with anti-DIG-AP antibody (1:1000, Roche) for 2h at RT. Upon several washes in PBT and one rinse with AP buffer probes were visualized using the NBT/BCIP solution in AP buffer (1:100). Primers used in 5'-3' orientation:

lola-RAA/RI_fwd	AACCACAACAATTGCCACACATCATC
lola-RAA/RI_rev	GAGAATGGTGTAGCTCTTGCTC
*CG6040*_fwd	CCTTTGCCGCCTTAAAACTGG
*CG6040*_rev	CGCTACCCAAGCTAATGCCG
*CG1674*_fwd	CACTAAAGCAGACCTTGTTTCG
*CG1674*_ rev	TTTCGCACTGCTGTGAAG

Locomotion assay

Freshly hatched male and female flies of the respective genotype were separated and directly placed into measuring cylinders. The locomotion was assessed using the climbing assay described previously (Bahadorani and Hilliker 2008). Flies were tapped to the bottom and flies passing 8 cm in 10 or 5 seconds, respectively, were counted. Measurements were repeated four times in two independent biological replicates for each of the shRNA expressing-lines.

Fertility assay

20 young female flies were separated upon hatching and mated with 5 males for three days. Upon fertilization, flies were transferred onto fresh agar plates every 24 hours, eggs were counted and allowed to develop for further 36 hours. The hatching rate was determined by counting the number of unhatched eggs for each

lay. Counting was done on four subsequent days. Experiments were performed in three biological replicates with two different RNAi knockdown-lines.

Cuticle preparation

Cuticle preparations were prepared as previously described (Liu and Lasko 2015). Embryos were dechorionated for 2 min with 50% bleach, washed with water, transferred into an Eppendorf tube and washed with PBT (0.1% Tween20 in PBS). The supernatant was removed and Hoyer's medium (30 g gum Arabic, 200 g chloral hydrate, 20 g glycerol ad 50 ml water) was added to cover the embryos. The tube was incubated over night at 65°C, embryos were mounted onto a glass slide and examined under dark field illumination.

Immunohistochemistry

Ovaries were dissected from 5 days-old virgin females in cold PBS and fixed in 5% formaldehyde for 20 min. Samples were washed three times for 10 min with PBT (0.3% Triton X-100 in PBS) and blocked for 20 min in PBT + 5% donkey serum at RT. Samples were incubated with primary antibodies (rabbit anti-Vasa, 1:500; mouse anti-1B1, 1:100) over night at 4°C, washed three times for 10 min with PBT and incubated with secondary antibody for two hours at RT. Samples were mounted in Vectashield and examined using a confocal Leica SP5.

Extended bioinformatics analysis

Dynamicity of protein profiles

We used the Gini ratio (Damgaard and Weiner 2000; Gini 1912) to measure the stability of protein abundance throughout time. The Gini ratio calculates a score ranging from 0 to 1, which depicts the normalized mean difference of LFQ values between every possible combination of two stages for each protein:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|}{2n^2 \mu}$$

where “n” is the number of stages, “x_i” the protein quantification (LFQ) at stage “i”, and “μ” the average protein quantification throughout time. The minimum score refers to proteins that are stably expressed regardless of their average quantitation, while proteins having high abundance in only one stage present scores close to 1.

Significantly changing proteins throughout the life cycle

We used analysis of variance (ANOVA) in order to detect significant changes in protein expression in the life cycle data set. The resulting p-values were adjusted by Benjamini and Hochberg to control for the false discovery rate (FDR). We defined a protein as differentially expressed with a cut-off of 1% FDR. To identify in which stages proteins significantly change, we used the Tukey HSD post-hoc test. The test defines the Honest Significant Difference as the minimal distance between two groups to be considered statistically significant. To have a measure of the strength of the changes, we also calculated the effect size as suggested for one-way ANOVA analysis (Cohen 1988):

$$f = \sqrt{\frac{\sum_{i=1}^{k} p_i * (\mu_i - \mu)}{\sigma^2}}$$

Stage-specific proteins of the life cycle

To classify proteins into stages, we required that they are differentially expressed at 1% FDR and either detected in only one stage (embryo, larva, pupa or adult) or showing high differences in abundance (LFQ fold change > 4) in at least two other stages.

Getting the significant changes during the embryo development

We used the *timecourse* (Tai 2007) and *qvalue* (Storey et al. 2015) packages to assess significant changes in protein expression during time. The *timecourse* package implements a multivariate empirical Bayes method to calculate

moderated T^2-statistics from longitudinal data, taking into account replicate information and correlation among gene expression along time points.

The significance of the T^2-statistics was empirically estimated using the *qvalue* package, by comparing the obtained T^2-statistics with the ones obtained from bootstrapping the original data. We performed a 1000-times bootstrap of each protein, permuting with replacement the values and calculating the statistic again. Then, we calculated the empirical p-values comparing the statistics from the original data and the pool of null statistics (bootstrapped data). To control the false discovery rate we use the *qvalue* function of the same package. Based on the distribution of our q-values, we set the significance cut-off at an FDR adjusted p-value of 0.0001.

Clustering strategies

We use Affinity Propagation (AP) (Bodenhofer et al. 2011; Frey and Dueck 2007) to cluster the differentially expressed proteins with similar expression profiles into an optimal number of clusters. AP is a well-established method to automatically calculate the best number of fitting clusters to the data. This method takes the data points as potential “exemplars” of clusters and further passing messages between points to decide which are [the best] exemplars and to which exemplar the rest belong to.

We calculated a similarity matrix between protein profiles using the negative Euclidean distance, defined as the negative squared distance between two points:

$$s = -d^r$$

Then we called AP with this similarity matrix, without setting any preference for any protein to be the exemplar of a cluster (default settings are the median of non-infinite values in the similarity matrix), and allowing up to 1000 iterations (i.e. rounds of messages passed between data points).

To reduce the number of clusters of the life cycle data set, we run AP with the preference parameter set at the 10% quantile of the distribution of similarities. For the embryo data set, the default parameters were used instead.

The goodness of the clustering was assessed using silhouette plots.

GO analysis

Using the R packages *GSEABase* (Morgan et al. 2016), *GOstats* (Falcon and Gentleman 2007) and *org.Dm.eg.db* (Carlson 2016, date stamp from the source of: 2015-Sep27) we performed gene set enrichment analyses of GO terms (biological processes only).

Due to the hierarchical nature of the GO annotation, usually many terms appear from the same set of genes. To remove redundancy of terms we scored the similarity between terms using the GOSemSim package (Yu et al. 2010). This package implements several methods to calculate the functional similarity of different terms. We used the Relevance method (Schlicker et al. 2006), based on the Information Content of two terms and their closest common ancestor.

Embryogenesis data set*:* For each cluster of the embryo data set we performed a hypergeometric test to find GO enriched categories. To assign GO terms from the clusters back to the different time points, we require for each time point that at least ¾ of the proteins comprised in each cluster were at least 2-fold enriched relative to their minimum LFQ value. To calculate the time point-specific GO terms, only terms not ubiquitously enriched in all time points were kept. Terms of each cluster were sorted by FDR and similarity scores were calculated. Eventually, only the term with the lowest p-value among similar terms (similarity score higher than 0.7 in a range between 0 and1) was kept.

Life cycle data set: For each stage of the life cycle, we performed a hypergeometric test with the stage-specific proteins.

We additionally performed GO terms analysis on the core proteome (defined as the fraction of proteins detected in all time points) and also scored the similarity between terms. The terms were summarized in a scatter plot and colored based on their similarity score: 1) we performed hierarchical clustering of the distances between terms, defined as 1-score; 2) we cut the tree into subtrees, grouping together terms with similarity higher than 0.7 (range 0-1, same threshold used

before); 3) the dimensionality of the similarity matrix was reduced to 2 dimensions using classical multidimensional scaling, which were used as xy coordinates to distribute the terms in a scatter plot colored based on the cluster assignment; 4) as the cluster representative GO term, the most enriched term (FDR) of each cluster was selected, with the size of the circle representing the number of genes a term contains.

Alternatively, we summarized the GO terms of the core proteome as a Treemap, which is a way of displaying hierarchical data using nested proportional rectangles. In our case, we colored (grouped) the terms based on the subtrees of the previously calculated hierarchical clustering of the similarity matrix. The cluster representative and the size of the rectangle were assigned as described above for the scatter plot.

Integration of RNA-seq data

RNA quantification comes from the Supplementary table S10 of Graveley et al., 2011, titled *FPKM levels for FlyBase 5.12 Transcripts from short poly(A)+ RNA-Seq*. We used this table to estimate the gene expression naively summing the FPKM values of the different transcripts quantified. In order to control the variance, the FPKM values were log-transformed.

We obtained from Ensembl 84 (Yates et al. 2016) the corresponding FlybaseName Gene ids to the Flybase Transcript ids, which were later used to merge the RNA and protein quantitation.

The plot comparing the similarity between RNA and protein expression at each time point is a false color image matrix with the pairwise Pearson correlation coefficients of all time points between the RNA and protein quantitation.

Translational delay and identification of RNA to protein translation patterns

As part of the integration of RNA and protein data, we grouped together genes that have similar RNA and protein profiles. To do this, we calculated a PCA for the protein and RNA data sets in order to accumulate the maximal variability in one dimension. The first component of the two reduced data sets was then used to cluster the genes by a k-means clustering (max. 1000 iterations, 100 starting

random centers). Based on Silhouette plots we chose 6 clusters that explained best our data.

Integration of the Fly-FISH data

We downloaded the localization data for embryos from http://fly-fish.ccbr.utoronto.ca/terms/ and complemented them with further annotation from the Ensembl release 84 (ensembl_gene_id, flybasename_gene) using Biomart. Proteins were classified into tissues based on FISH data at not only embryo stages but also taking into account RNA expression in later developmental stages. Genes were automatically clustered based on the LFQ data available, using the Affinity Propagation method described before with negative squared Euclidean distances. The average expression trend of each cluster was then calculated by lowess smoothing (Cleveland 1979, 1981). Eventually, we calculated the GO terms enriched per cluster using the R packages *GSEABase*, *GOstats* and *org.Dm.eg.db* as previously described.

Supplemental Tables

Supplemental Table 1. Filtered MaxQuant output table with calculated and imputed LFQ values.

Supplemental Table 2. GO enrichment information on the life cycle proteome.

Supplemental Table 3. Calculated dynamicity score table.

Supplemental Table 4. Clusters obtained for the complete life cycle experiment.

Supplemental Table 5. GO enrichment information on the life cycle clusters.

Supplemental Table 6. One week old male and female protein comparison.

Supplemental Table 7. Young male and female protein comparison.

Supplemental Table 8. Overview of maternally loaded RNA and proteins.

Supplemental Table 9. Filtered MaxQuant output table obtained by including ncRNA sequences.

Supplemental Table 10. Filtered MaxQuant output table obtained by including pseudogene sequences.

Supplemental Table 11. Filtered MaxQuant output table with calculated and imputed LFQ values for the embryogenesis time course.

Supplemental Table 12. Automatically generated 70 clusters for the embryogenesis proteome.

Supplemental Table 13. GO enrichment information on the embryogenesis time-course analysis clusters.

Supplemental Table 14. Tissue-specific cluster information.

Supplemental Table 15. GO enrichment information on the tissue-specific clusters.

Supplemental Table 16. Six RNA and protein clusters based on the first PCA component.

Supplemental References

Badoahrani S, Hilliker AJ. 2008. Antioxidants cannot suppress the lethal phenotype of a Drosophila melanogaster model of Huntington's disease. *Genome* **51**: 392–395.

Bodenhofer U, Kothmeier A, Hochreiter S. 2011. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27:** 2463–2464.

Carlson M. 2016. *org.Dm.eg.db: Genome wide annotation for Fly*.

Cleveland WS. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74:** 829–836.

Cleveland WS. 1981. Lowess - A program for smoothing scatterplots by robust locally weighted regression. *American Statistician* **35:** 54.

Cohen J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.) L. Erlbaum Associates, Hillsdale, New Jersey

Damgaard C, Weiner J. 2000. Describing inequality in plant size or fecundity. *Ecology* **81:** 1139–1142.

Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23:** 257–258.

Frey BJ, Dueck D. 2007. Clustering by passing messages between data points, *Science* **315:** 972–976.

Gini C. 1912. Variabilità e mutabilità, contribrito allo studio delle distribuzioni e delle relazioni statistiche. Fascicolo Ier: Introduzione - Indici di variabilità - Indici di mutabilità. Bologne.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. 2011. The developmental transcriptome of *Drosophila melanogaster. Nature* **471**: 473-479.

Kappei D, Butter F, Benda C, Scheibe M, Draskovic I, Stevense M, Novo CL, Basquin C, Araki M, Araki K, et al. 2013. HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. *EMBO J* **32**: 1681–1701.

Kondo S, Ueda R. 2013. Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics* **195**: 715–721.

Liu N, Lasko P. 2015. Analysis of RNA interference lines identifies new functions of maternally-expressed genes involved in embryonic patterning in *Drosophila melanogaster*. *G3 (Bethesda)* **5**: 1025–1034.

Morgan M, Falcon S, Gentleman R. 2016. *GSEABase: Gene set enrichment data structures and methods.*

Schlicker A, Domingues FS, Rahnenfuehrer J, Lengauer T. 2006. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* **7**: 302.

Storey D, Bass AJ, Dabney A, Robinson D. 2015. *Q-value estimation for false discovery rate control*.

Sury MD, Chen JX, Selbach M. 2010. The SILAC fly allows for accurate protein quantification in vivo. *Mol Cell Proteomics* **9**: 2173-2183.

Tai YC. 2007. *Timecourse: Statistical Analysis for Developmental Microarray Time Course Data.*

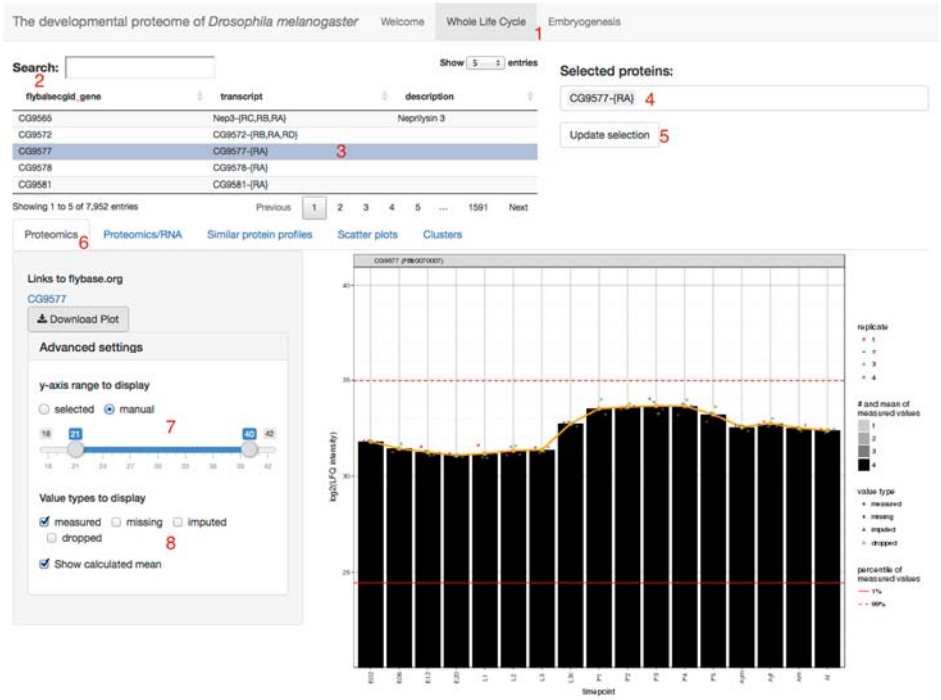
Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L et al. 2016. Ensembl 2016. *Nucleic Acids Res*. **44**: D710-D716.

Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products, *Bioinformatics* **26:** 976–978.

How-To for the web interface

This is a short manual on how to navigate the web interface accompanied with the manuscript ”The developmental proteome of *Drosophila melanogaster*”.

Contents



1 Data sets

In the top menu [1], you can select the data set to work with: “Whole Life Cycle” or “Embryogenesis”.

1.1 Whole life cycle

The data set contains the following 15 time points (measured in quadruplicates):

- E02: 0-2 hours old embryos
- E06: 4-6 hours old embryos
- E12: 10-12 hours old embryos
- E20: 18-20 hours old embryos
- L1: 40-44 hours after egg laying
- L2: 66-68 hours after egg laying
- eL3: 83-85 hours after egg laying
- L3c: crawling larvae
- P1: white pupae
- P2-P5: every 24 hours after pupation (white pupae)
- Aym/Ayf: adult young male/female flies 4 hours after hatching
- Am/Af: 1 week old adult male/female flies

1.2 Embryogenesis

Embryos were collected in quadruplicates with a 30 min laying window for the following time points: 0h, 1h, 2h, 3h, 4h, 5h, 6h, 8h, 10h, 12h, 14h, 16h, 18h, 20h.

2 Web interface structure

2.1 Search

After selecting one of the data sets, a table with all detected proteins will appear. The protein of interest can be queried in the “Search” bar [2] and selected from the

table by clicking [3]. After selecting one entry, the protein will be saved in the “update selection” window on the upper right site.

2.2 Selected proteins

This feature allows keeping information about previous searches. For a completely new search, please delete the entry in the upper right window [4] and click the update selection button [5].

2.3 Data analysis

After selecting the protein of interest, the data can be chosen from a second menu [6] with the following tabs:

- Proteomics
- Proteomics/RNA
- Similar protein profiles
- Scatter plots
- Clusters

All the graphical representations are navigable by hovering over the dots/lines. Additionally, it is possible to download any graphical representation via the “download” button.

2.3.1 Proteomics

Bar plots representing protein abundance (y-axis) throughout the collected time points of the selected data set (x-axis). Colored dots indicate the different replicates. The grey scale of the bars represents the number of replicates in which the protein was measured (grey=1, black=4). The value type legend specifies the origin of the data (for detailed information, please check supplemental methods). In each plot two thresholds are presented: 1% (continuous line) and 99% (dashed line) of the LFQ intensities found in the complete data set. The orange line

indicates the tendency of protein expression based on measured and imputed values.

2.3.2 Proteomics/RNA

Protein and RNA profiles of the selected protein are shown. The RNA data is retrieved from the modENCODE project (Graveley et al., 2011). All transcripts for each protein are color-coded and by hovering over the lines, their gene names are displayed.

2.3.3 Similar protein profiles

This feature allows searching for proteins with similar expression profiles to the protein of interest. The 100 most similar proteins are listed in the table on the left (scored by similarity) and can be searched in the search tab. By selecting proteins, expression profiles of protein and RNA will be added to the graphical output in a different color.

2.3.4 Scatter plots

Different scatter plots from the manuscript are available for interactive searches (see left menu for selection). The selected proteins from the search (if available in the scatter plot) are highlighted.

2.3.5 Clusters

Significantly changing whole life cycle protein profiles were clustered and are shown by stage (embryo, larva, pupa, adult). The cluster of interest can be selected and respective protein profiles will be visualized. The clusters that contain information about the selected proteins will be displayed (lower right).

Embryogenesis data was complemented with tissue-specific expression data. Therefore, a second tab “tissue” allows the selection of tissue and cluster number showing the respective protein expression profiles.

3 Advanced settings

This tab allows modifying different aspects of the graphical output:

- y-axis: the ordinate range can be defined [7]
- Value types [8]:
  - **measured:** calculated LFQ intensities
  - **missing:** no LFQ intensity calculated, but measured intensity
  - **imputed:** calculated by imputation (see supplemental methods)
  - **dropped:** a single measured replicate value without intensities in neighboring time points was dropped and replaced by an imputed value
- Show/hide calculated mean (orange line)