



**HAL**  
open science

## **Genes involved in *Drosophila melanogaster* ovarian function are highly conserved throughout evolution**

Sébastien Elis, Alice Desmarchais, Emilie Cardona, Sophie Fouchécourt, Rozenn Dalbies-Tran, Thuy Thao Vi Nguyen, Violette Thermes, Virginie Maillard, Pascal Papillier, Svetlana Uzbekova, et al.

### ► To cite this version:

Sébastien Elis, Alice Desmarchais, Emilie Cardona, Sophie Fouchécourt, Rozenn Dalbies-Tran, et al.. Genes involved in *Drosophila melanogaster* ovarian function are highly conserved throughout evolution. *Genome Biology and Evolution*, 2018, 10 (10), pp.2629-2642. 10.1093/gbe/evy158 . hal-01991460

**HAL Id: hal-01991460**

**<https://hal.science/hal-01991460v1>**

Submitted on 23 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Title:

# Genes involved in *Drosophila melanogaster* ovarian function are highly conserved throughout evolution

Sebastien Elis<sup>1#</sup>, Alice Desmarchais<sup>1#</sup>, Emilie Cardona<sup>2</sup>, Sophie Fouchecourt<sup>1</sup>, Rozenn Dalbies-Tran<sup>1</sup>, Thaovi Nguyen<sup>2</sup>, Violette Thermes<sup>2</sup>, Virginie Maillard<sup>1</sup>, Pascal Papillier<sup>1</sup>, Svetlana Uzbekova<sup>1</sup>, Julien Bobe<sup>2</sup>, Jean-Louis Couderc<sup>3</sup> and Philippe Monget<sup>1\*</sup>

<sup>1</sup> UMR PRC, CNRS, IFCE, INRA, Université de Tours, 37380, Nouzilly, France

<sup>2</sup> LPGP, INRA, 35000 Rennes, France

<sup>3</sup> GReD Laboratory, Université Clermont Auvergne - CNRS UMR 6293- INSERM U1103, Clermont-Ferrand, France

#Equivalent contribution

\*Author for Correspondence: Philippe Monget, INRA Centre Val de Loire, Physiologie de la Reproduction et des Comportements, 37380 Nouzilly, France

Phone : 0033 2 47427797

Fax : 0033 2 47427743

Email address: [philippe.monget@inra.fr](mailto:philippe.monget@inra.fr)

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## **ABSTRACT**

This work presents a systematic approach to study the conservation of genes between fruit flies and mammals. We have listed 971 *Drosophila* genes involved in female reproduction at the ovarian level and looked for orthologues in the *Ciona*, zebrafish, coelacanth, lizard, chicken and mouse. Depending on the species, the percentage of these genes with at least one orthologue varies between 69% and 78%. In comparison, only 42% of the *Drosophila* genes have an orthologue in the mouse genome ( $p < 0.0001$ ), suggesting a dramatically higher evolutionary conservation of ovarian genes. The 177 *Drosophila* genes that have no orthologue in mice and other vertebrates correspond to genes involved in mechanisms of oogenesis that are specific to the fruit fly or the insects. Among 759 genes with at least one orthologue in the zebrafish, 73 have an expression enriched in the ovary in this species (RNA-seq data). Among 760 genes that have at least one orthologue in the mouse; 76 and 11 orthologues are reported to be preferentially and exclusively expressed in the mouse ovary, respectively (based on the UniGene EST database). Several of them are already known to play a key role in murine oogenesis and/or to be enriched in the mouse/zebrafish oocyte, while others have remained unreported. We have investigated, by RNA-seq and RT-qPCR, the exclusive ovarian expression of 10 genes in fish and mammals. Overall, we have found several novel candidates potentially involved in mammalian oogenesis by an evolutionary approach and using the fruit fly as an animal model.

**Key words:** evolution; conservation; reproduction; ovary

## 1. INTRODUCTION

Several examples of conservation of genes of the biochemical and molecular mechanisms involved in various biological functions have been previously documented, in particular between flies and mammals (McGary, et al. 2010). For example, *hox* genes are involved in morphogenesis and cell differentiation in the fly and in the morphogenesis of the vertebrate embryo (Mark, et al. 1997); the *ovo* gene is implicated in epidermal differentiation in the fruit fly and in hair formation in the mouse (Dai et al. 1998).

In *Drosophila*, contrary to mammals, oogenesis spans the entire life of a female, due to the presence of stem cells in the most anterior part of the ovary in a region called the germarium. Although different at first sight, *Drosophila* and vertebrate oocyte formation and maturation might share more common features than previously thought (Matova and Cooley 2001). Firstly, meiosis is an important common step in the precursors of oocytes as well as packing the oocyte with RNA, proteins and organelles that are crucial for the first steps of oogenesis. Moreover, within the animal kingdom, neo-oogenesis exists in several species, not only in insects but also in molluscs and teleost fishes (Billard 1987; Craig-Bennett 1931; Eggert 1931; Grier et al. 2007; Hann 1927; Wallace and Selman 1990). In *Drosophila*, many experiments have shown that the self-renewal of germline stem cells (GSCs) takes place in a niche, a physical structure made of signalling cells, that functions to maintain stem cell self-renewal and to prevent stem cell differentiation (Spradling et al. 2001; Watt and Hogan 2000). Renewal of the ovarian reserve during adult life is widely found in teleost fishes, and, in some salmonid species, there is an increase in the number of oogonia following spawning. In the rainbow trout (*Oncorhynchus mykiss*) postovulatory ovary, cell nests can be found in the germinal epithelium that contain oogonia, early diplotene oocytes and prefollicle cells (Grier et al. 2007). In medaka (*Oryzias latipes*), recently characterised ovarian cords within the germinal ovarian epithelium have been hypothesised to be reminiscent of the germarium from the *Drosophila* ovary (Nakamura et al.

2010). In mammalian ovaries, this type of niche has never been observed neither based on morphological studies nor molecular markers.

Recently it has been shown, by analysing how primary oocytes are formed from germ cells in the mouse embryo, that germ cell cysts exist in mouse ovaries similar to those found in *Drosophila*. Cells of these cysts are connected by ring canals and result from the synchronous divisions but incomplete cytokinesis of primordial germ cells. As in *Drosophila*, the future murine oocyte seems to be “nursed” by surrounding sister cells transferring some of their cytoplasm to this cell including Golgi, mitochondria, centrosomes and likely RNAs and proteins. Eventually, this unique cell will become a primary oocyte, while the other cells of the cyst die, like nurse cells in *Drosophila* (Pepling and Spradling 2001). In both *Drosophila* and mice, oocyte transcripts encode the so-called maternal genetic determinants that will support the earliest stages of embryogenesis (Laver et al. 2015). Mechanisms that control mRNA localisation and regulate translation play central roles in specification of the anteroposterior and dorsoventral axes of the oocyte and future embryo and have been studied in detail in *Drosophila* (Lasko 2012).

Many genes have been shown to be involved in oogenesis in *Drosophila*, and some of them have an orthologue in the mouse and zebrafish that is also involved in oogenesis. It is the case with *mos* and *Aurora*, involved in meiosis, and the mRNA binding-proteins *orb/cpeb1* (Extavour 2009; Schindler et al. 2012; Zheng and Dean 2007). We have previously identified genes encoding mRNA binding proteins that are highly expressed in the mouse oocyte, based on their phylogenetic relationship with fruit flies and nematode orthologues known to be involved in oogenesis in both species (Drouilhet et al. 2008). However, systematically identifying the orthologues of all fly genes in mammals and other vertebrates remains undone. In the present work, we have recovered a list of 971 fruit fly genes whose mutations affect ovary, egg, follicle cells, oocyte and/or germarium according to the FlyBase database

(*flybase.org*) or reported in the literature to affect *Drosophila* oogenesis. We have then systematically looked for orthologues of these genes in the *Ciona*, zebrafish, coelacanth, lizard, chicken and mouse to i) investigate their phylogenetic evolution and the evolution of ovarian function from the fly to vertebrates, ii) identify genes that are specific of the fly and/or the insects as well as genes that are conserved between these species and iii) identify putative orthologues of genes whose expression are enriched or exclusively expressed in the mouse and/or zebrafish ovary, using *in silico* and RNA-seq data, respectively. To further validate the evolutionary-conserved expression pattern of a subset of genes, we further investigated their expression using another model teleost species, the medaka, which is evolutionarily distant from the zebrafish.

## 2. MATERIALS AND METHODS

### 2.1 *In silico* analyses

A list of genes of *Drosophila melanogaster* involved in female reproduction at the ovarian level was generated by compiling FlyBase lists of *Drosophila* mutants affecting ovarian phenotypes (<http://flybase.org/>). Five different affected tissues or cells were chosen to increment the list: ovary, egg, follicle cells, oocyte and germarium. An additional list of candidate genes described as being involved in oogenesis in the literature was also added (Bastock and St Johnston 2008; Kumano 2012; McKearin and Christerson 1994; Spradling 1993; Stein and Stevens 2014). The *Drosophila* gene list thus included 356 genes affecting the egg, 223 genes affecting the oocyte, 317 genes affecting follicle cells, 184 genes affecting the germarium, 688 genes affecting the ovary and 337 genes involved in oogenesis; the latter have been previously reported in the literature. Once redundancy is taken into account, the fruit fly gene list was composed of 971 unique genes (Supplementary Table 1).

Orthologues of *D. melanogaster* genes (genome version BDGP6) were extracted from the Ensembl database using the BioMart tool (<http://www.ensembl.org/biomart/martview/>) for chordates: *Ciona intestinalis* (genome version KH), *Latimeria chalumnae* (coelacanth, genome version LatCha1), *Danio rerio* (zebrafish, genome version GRCz10), *Anolis carolinensis* (anole lizard, genome version AnoCar2.0), *Gallus gallus* (chicken, genome version Gallus\_gallus-5.0) and *Mus musculus* (mouse, genome version GRM38.p5) (Supplementary Table 1). The orthology percentage of the 971 *Drosophila* genes list was also retrieved in the Ensembl metazoa database (<http://metazoa.ensembl.org/biomart/martview/>) for sponge (*Amphimedon queenslandica*, genome version Aqu1), cnidarians (*Nematostella vectensis*, genome version ASM20922v1), platyhelminthes (*Schistosoma mansoni*, genome version ASM23792v2), Ecdysozoa : arthropods (*Drosophila erecta*, genome version dere\_caf1; *Anopheles gambiae*, genome version AgamP4; *Apis mellifera*, genome version Amel\_4.5), roundworms

(*Caenorhabditis elegans*, genome version WBcell235), Lophotrocozoa : molluscs (*Octopus bimaculoides*, genome version PRJNA270931), annelids (*Capitella teleta*, genome version Capitella teleta v1.0) and echinoderms (*Strongylocentrotus purpuratus*, genome version Spur\_3.1).

For all mouse orthologues of *Drosophila* genes, EST (expressed sequence tag) expression data from all tissues was collected from the UniGene database (<http://www.ncbi.nlm.nih.gov/unigene>). The ratio of ovarian expression was calculated as the sum of ovary, oocyte, fertilised and unfertilised ovum in transcripts per million divided by the sum of all adult female tissues. The ratio of testicular expression was calculated as the sum of testis, epididymis and prostate in transcripts per million divided by the sum of all adult male tissues (Supplementary Table 2).

For zebrafish orthologues, RNA-seq data were extracted from the PhyloFish database (Pasquier et al. 2016). The ratio of ovarian expression was calculated as the ovary expression divided by the sum of all adult female tissues (Supplementary Table 3). The ratio of testicular expression was calculated as the testis expression divided by the sum of all adult male tissues (Supplementary Table 3). For medaka, orthologues were identified using the list of manually curated orthology relationships established between zebrafish and medaka, using the spotted gar as an outgroup of teleosts (Braasch et al. 2016). Characterisation of the gene lists was performed using gene ontology and DAVID 6.8 (<https://david.ncifcrf.gov/home.jsp>) (Huang et al. 2009, 2008).

## **2.2 Biological material**

All experimental protocols were carried out in accordance with the European Directive 2010/63/EU on the protection of animals used for scientific purposes and the recommendations of the French Ministry of National Education, Higher Education and Research. The brain, heart, liver, spleen, skeletal muscle (quadriceps), kidney, bladder, uterus and ovaries were collected



from three adult female mice of the Swiss strain (Janvier, Le Genest St Isle, France). Testes were also collected from three adult male mice. Tissues were frozen in liquid nitrogen and stored at -80 °C until their use for RNA isolation.

For medaka and zebrafish, all sampled fishes originated from the INRA LPGP experimental facility. Fishes were reared and handled in strict accordance with French and European policies and guidelines of the INRA LPGP Institutional Animal Care and Use Committee (# 25M10), which approved this study. In medaka, all tissues were collected from 8 adult females ( $0.431 \pm 0.05$  g) and 8 adult males ( $0.425 \pm 0.05$  g). Zebrafish fish tissue originated from adult fish, as previously described (Braasch et al. 2016; Pasquier et al. 2017). All samples were immediately frozen in liquid nitrogen and subsequently stored at -80 °C until RNA extraction.

### **2.3 Gene expression analysis**

Total RNA was extracted from murine tissues using TRI reagent, as recommended by the manufacturer's instructions. Then, potential genomic DNA contamination was eliminated using RQ1 RNase-free DNase (Promega, Charbonnières-les-Bains, France), and the RNA quantity was assessed using a NanoDrop ND-1000 spectrophotometer (Nyxor Biotech, Paris, France). Reverse transcription (RT) was carried out on 1 µg of total RNA with the iScript Select CDNA synthesis kit (BioRad, Marnes La Coquette, France), according to the manufacturer's protocol. Real-time PCR was performed in a total volume of 20 µL containing each primer at a final concentration of 125 nM (see sequences of specific primer pairs in Table 1), 5 µL of RT reaction solution (diluted 1:25) and 10µL of IQ™ SYBR® Green Supermix (BioRad, Marnes La Coquette, France), according to the manufacturer's instructions. Real-time PCR reactions were run on a MyiQ™ cyclor (BioRad, Marnes La Coquette, France), with the following thermal conditions: denaturation at 95 °C for 3 min, then 40 cycles of a three-step protocol (denaturation at 95 °C for 30 s, specific annealing at 60 °C for 30 s and elongation at 72 °C for 30 s) followed by acquisition of a melting curve. Gene expression was normalised using the

housekeeping gene *Rpl19* with stable expression under our conditions. The normalised values

of relative expression (R) were calculated according to the following equation:  $R = \frac{(E^{-Cq \text{ Gene}})}{(E^{-Cq \text{ Rpl19}})}$ ,

where Cq is the cycle threshold and E is PCR efficiency for each primer pair. The specificity of the amplified fragment was controlled by checking the amplicon size by electrophoresis. Normalised gene expression is expressed as mean  $\pm$  SEM of three female/male mice.

In medaka, total RNA from several tissues was extracted using the Trizol method (MRC, USA), according to the manufacturer's instructions. The quantity and quality of each RNA were assessed by measuring their absorbance at 260 nm and 280 nm using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Courtaboeuf, France) associated with ND-1000 V3 7.0 software. A DNase treatment to remove residual DNA was carried out using the Ambion DNase free kit, according to the manufacturer's instructions. First-strand cDNA was synthesised with 540 ng of total RNA in each reaction system using the Thermo Scientific Maxima first strand cDNA synthesis system, according to the manufacturer's protocol. At this step, luciferase control RNA (Promega, Charbonnières-les-Bains, France), 50 pg per 500 ng of total RNA, was added to each sample to allow for data normalisation. All cDNAs were diluted 1:10 with nuclease-free water and stored at -20 °C until used as templates in real-time quantitative PCR (RT-qPCR). Primers were designed so that they were overlapping an intron when it was possible (Primer3 software) using known sequences from the PhyloFish database. The primer sequences are shown in Table 2. To determine the RT-qPCR efficiencies of each primer pair used, standard curves were generated using six serial dilutions (cDNA dilution from 1:5 to 1:160) of a pool from eight ovary samples. The ability of primers was validated when the amplification efficiency varied between 90% and 110%. RT-qPCR was carried out using a StepOnePlus™ Real-Time PCR System (Applied Biosystems, Villebon Sur Yvette, France) using Promega's GoTaq® qPCR Master Mix following the manufacturer's recommendations. Each PCR run included duplicates of each sample and also negative controls (reverse

transcriptase-free and water RNA-free samples). The reactions were mixed in a volume of 10  $\mu\text{L}$  containing 5  $\mu\text{L}$  SYBR® Premix, 4  $\mu\text{L}$  cDNA (diluted 1:10) and 0.5  $\mu\text{L}$  each of the 6  $\mu\text{M}$  forward and reverse primers. After initial denaturation at 95 °C for 15 min, 40 cycles of amplification were carried out starting at 95 °C for 15 s, followed by 1 min at 60° C, with a final extension at 60 °C for 1 min and 95 °C for 15 min. Data were subsequently normalised using exogenous luciferase transcript abundance in samples.

## **2.4 Statistical Analysis**

The proportion of genes exhibiting orthologues, enriched expression at the ovarian level or enriched gene ontology were compared using Chi square analyses.

Gene expression was compared between tissues using non-parametric analysis of variance (permutational ANOVA) (R package *lmp*) (Wheeler 2010) with the Tukey post-hoc test (R package *np*) (Konietschke et al. 2015)), R version 3.3.1 (R\_Core\_Team 2015). A difference with  $p \leq 0.05$  was considered significant.

### 3. RESULTS

#### 3.1 Chordate orthologues

The tables presenting the 971 fruit fly genes and their orthologues in chordate species, as well as EST/RNA-seq expression of corresponding murine/zebrafish orthologue genes are provided in Supplementary Tables 1, 2 and 3. Among these 971 fruit fly genes, 81.8% (794 genes) possess a chordate orthologue in at least one of the following species: *C. intestinalis* (69.3%), *L. chalumnae* (coelacanth, 77.4%), *D. rerio* (zebrafish, 78.2%), *A. carolinensis* (anole lizard, 77.0%), *G. gallus* (chicken, 76.4%) and/or *M. musculus* (mouse, 78.3%). Considering only the mouse, 78.3 % of the fruit fly genes (760 genes) have mouse orthologues, with a total of 1556 mouse orthologues. Only 41.7 % of the complete fruit fly gene repertoire (17559) have mouse orthologues (7318 genes) according to the Ensembl BioMart database. Therefore, our fruit fly gene list showed a significant increase in the proportion of genes possessing a mouse orthologue ( $p < 0.0001$ ) or a vertebrate orthologue (Figure 1) in comparison to the fruit fly gene repertoire. Given the high percentage of the studied genes that are conserved in the fruit fly and mouse, we investigated the conservation of these genes in other phyla. Among the 971 fruit fly genes, 66.0% possess an orthologue in sponge (*A. queenslandica*, 641 genes) and 72.0% possess an orthologue in cnidarians (*N. vectensis*, 699 genes). Only 34.6% and 38.6% of the complete fruit fly genome (17559) have orthologues in *A. queenslandica* (6067 genes) and *N. vectensis* (6784 genes), respectively. Therefore, our fruit fly gene list showed a significant increase in the proportion of genes possessing an orthologue in sponge and cnidarians ( $p < 0.0001$ ) compared to the fruit fly gene repertoire.

To determine which GO (gene ontology) was especially enriched throughout evolution, a GO analysis was performed on the 1556 mouse orthologue genes and the 1849 zebrafish genes corresponding to the 971 fruit fly genes (Figures 2 and 3; Supplementary Tables 4 and 5), and they were compared to the murine and zebrafish genome, respectively. Enriched GO terms

corresponding to at least 20 of the 1556 or 1849 mouse or zebrafish genes, respectively, were investigated. Only GO terms with a Bonferroni corrected  $p$  value  $< 0.01$  were considered significantly enriched compared to the complete genome. In the mouse, concerning the biological process GO category, ontologies related to signal transduction (peptidyl-tyrosine phosphorylation, peptidyl-serine phosphorylation, regulation of protein phosphorylation, canonical Wnt signalling pathway, transmembrane receptor protein tyrosine kinase signalling pathway and cellular response to insulin stimulus) ( $p < 0.0001$ ), regulation of translation ( $p < 0.0001$ ) and to cell cycle (regulation of cell cycle and actin cytoskeleton organisation) ( $p < 0.0001$ ) were enriched more than 4-fold in the 1556 mouse orthologue genes compared to the mouse genome (Figure 2, Supplementary Table 4). Concerning molecular function GO category, ontologies related to kinase, endopeptidase and phosphatase activities ( $p < 0.0001$ ) as well as related to mRNA binding ( $p < 0.0001$ ) were also enriched more than 4-fold in the 1556 mouse gene list compared to the mouse genome (Figure 3, Supplementary Table 4).

In the zebrafish, similar ontologies were found significantly increased compared to the genome. Concerning biological process GO category, ontologies related to signal transduction (peptidyl-tyrosine autophosphorylation, integrin-mediated signalling pathway, canonical Wnt signalling pathway, transmembrane receptor protein tyrosine kinase signalling pathway and peptidyl-serine phosphorylation) ( $p < 0.0001$ ), cell migration ( $p < 0.0001$ ) and actin cytoskeleton organisation ( $p < 0.01$ ) were enriched more than 3-fold in the 1849 zebrafish orthologue genes compared to the zebrafish genome (Figure 2, Supplementary Table 5). Concerning molecular function GO category, ontologies related to kinase and endopeptidase activities ( $p < 0.0001$ ), motor activity ( $p < 0.0001$ ) and microtubule binding ( $p < 0.0001$ ) as well as related to mRNA binding ( $p < 0.0001$ ) were also enriched more than 3-fold in zebrafish genes compared to the zebrafish genome (Figure 3, Supplementary Table 5). A lot more GOs are enriched in the mouse gene list compared to the zebrafish gene list. Indeed, approximately

all GOs that are enriched in the zebrafish gene list are also found enriched in the mouse gene list. Therefore, additional GOs are found enriched in the mouse gene list but not in the zebrafish gene list (i.e., regulation of translation, cellular response to insulin stimulus, positive regulation of neuron apoptotic process, negative regulation of ERK1 and ERK2 cascade, transforming growth factor beta receptor signalling pathway and histone deacetylase binding).

### **3.2 *Drosophila* genes with no orthologues in chordate**

Among the 971 fruit fly genes, 18.2% of the fruit fly genes have no chordate orthologues (177/971) (Supplementary Table 1). We investigated the evolution of these genes in order to detail the absence of these genes (Figure 4). Concerning the Ecdysozoa phylum, 94.4 % of these genes have orthologues in the closest species of *Drosophila* (*Drosophila erecta*), which is a significantly higher conservation percentage compared to the complete *Drosophila melanogaster* genome (72.8%). In other insects phyla (*A. gambiae* and *A. mellifera*, respectively), 57.9 and 52.8 % of the 177 *Drosophila* genes without vertebrate orthologues, are present, which is not different from the average conservation of the complete *Drosophila melanogaster* genome (51.8% and 47.4%, respectively). All the other invertebrate phyla analysed showed significantly lower conservation percentages of the 177 genes list compared to the average conservation of the complete *Drosophila melanogaster* genome: 12.4 % have roundworm orthologues (*C. elegans*) vs 31.4%; concerning the Lophotrochozoa phylum, 23.6 % of the genes have a mollusc orthologue (*O. bimaculoides*) vs 41.6% and concerning Echinoderms phylum, only 19.7 % of the genes have a *S. purpuratus* orthologue vs 40.0%. This result could suggest that most of these genes are either lost after insect phyla or appeared throughout insect evolution.

### **3.3 *In silico* ovarian enriched expression in mice**

The UniGene database enabled us to class the mouse orthologues regarding their ovarian expression. Among the 1556 mouse orthologues, 76 genes had an ovarian expression ratio above 0.5 (meaning that more than 50% of the total number of transcripts from all tissues of these genes are in the ovarian compartment) (Supplementary Tables 2 and 6). Among these genes, 11 mouse genes are described as strictly expressed in the ovaries (*Kpna7*, *Aurkc*, *Nanos2*, *Pabpn1l*, *T*, *Cts7*, *Eif4e1b*, *Tdrd1*, *Rbmxl2*, *Pin1rt1* and *Spocd1*), including 3 genes that are also strictly expressed in the testes of males (*Kpna7*, *Tdrd1* and *Rbmxl2*).

Moreover, 52 genes had a testis expression above 0.5 (Supplementary Table 7). Among these genes, 14 mouse genes are described as strictly expressed in the testis but not in the mouse ovary (*Fscn3*, *Gm4064*, *Tdrd6*, *Hmgb4*, *Hus1b*, *4921511C20Rik*, *Gm382*, *Enthd1*, *Meiob*, *Capza3*, *Gm732*, *Mycs*, *DIPas1* and *Rbm31y*).

GO analysis was used to determine which GO was enriched when selecting genes on a preferential ovarian expression ratio above 0.5 (76 mouse genes) compared to the complete list of 1556 mouse orthologue genes (Supplementary Table 8). Enriched GO terms corresponding to at least 4 of the 76 mouse genes, respectively, were investigated. Only GO terms with a raw p value < 0.05 were considered significantly enriched compare to the 1556 mouse gene list. Concerning biological process GO category, the ontologies related to the meiotic cell cycle (p = 0.0026) and regulation of translation (p = 0.045) were significantly enriched (7.8-fold and 3.0-fold, respectively) compared to mouse genes (Figure 5). Concerning molecular function GO category, the mRNA binding ontology was significantly (p = 0.029) enriched (2.9-fold) in mouse genes, exhibiting a preferential ovarian expression compared to all 1556 mouse genes.

### **3.4 Ovarian enriched expression in zebrafish (RNA-seq)**

The RNA-seq data obtained on zebrafish tissues allowed the classification of zebrafish orthologues based on their ovarian expression (Supplementary Table 3). Among the 1849

zebrafish orthologues, 73 genes had an ovarian expression ratio above 0.4 (meaning that more than 40% of the total number of transcripts from all tissues of these genes are in the ovarian compartment) (Supplementary Table 3). As zebrafish data consisted of RNA-seq data and not EST data, a 0.4 ratio was chosen instead of 0.5. Indeed, contrary to UniGene data, expression in several tissues is observed for all the zebrafish genes studied, likely due to the efficiency of RNA-seq methodology. Interestingly, 20 genes were common between the mouse and zebrafish (Supplementary Table 6). Indeed, 20 mouse genes and their 20 zebrafish orthologues had an ovarian preferential expression: *Kpna7*, *Pabpn1l*, *Aurkb*, *Eif4e1b*, *Rnf17*, *Foxr1*, *Tdrd9*, *Pld6*, *Cpeb1b*, *Ccnb1*, *Kif23*, *Igf2bp1*, *Kif14*, *Ncapg*, *Lsm14b*, *Mybl2b*, *Mcm10*, *Mcph1*, *Dlgap5* and *Cdc7*.

GO analysis was used on genes showing a preferential ovarian expression ratio above 0.4 (73 zebrafish genes) compared to the complete list of 1849 zebrafish orthologue genes (Supplementary Table 9). Enriched GO terms corresponding to at least 4 of the 73 zebrafish genes, respectively, were investigated. Only GO terms with a raw p value < 0.05 were considered significantly enriched compared to the 1849 zebrafish gene list. Concerning biological process, the ontologies related to DNA methylation involved in gamete generation (p = 0.0005; 21-fold enriched), regulation of translation (p = 0.0028; 7.7-fold enriched), DNA replication (p = 0.011; 8.1-fold enriched) and microtubule-based movement (p = 0.011; 8.1-fold enriched) were significantly enriched compared to zebrafish genes (Figure 5). Concerning molecular function, the ontologies related to RNA binding (p = 0.004; 3.0-fold enriched), ATPase activity (p = 0.0038; 11.3-fold enriched) and microtubule motor activity (p = 0.018; 6.8-fold enriched) were significantly enriched compared to the 1849 zebrafish genes.

Ontologies enriched in zebrafish ovarian genes and related to ATPase activity, microtubule-based movement or microtubule motor activity were not enriched in mouse ovarian genes. On the contrary, ontologies enriched in mouse ovarian genes and related to the



negative regulation of translation or gene silencing by RNA were not enriched in zebrafish ovarian genes.

### **3.5 *In vivo* confirmation of ovarian enriched expression**

Among the mouse orthologues exhibiting an *in silico* ovarian enriched expression, the expression of 11 genes was checked by qPCR: *Eife1b*, *Foxr1*, *Kpna7*, *Lin28a*, *Nanos2*, *Pin1rt1*, *Rnf17*, *Slc25a31*, *Spocd1*, *Tdrd1* and *Tdrd9* (Figure 6). Eight of these genes showed significantly enriched expression in the ovary compared to nongonadal tissues (*Eife1b*, *Kpna7*, *Lin28a*, *Nanos2*, *Pin1rt1*, *Rnf17*, *Spocd1* and *Tdrd1*). Moreover, the expression in the ovary is higher compared to all other tissues for six of these genes: *Eife1b*, *Kpna7*, *Lin28a*, *Nanos2*, *Pin1rt1* and *Spocd1*. Interestingly, a significantly enriched expression in the testis was found for six genes compared to nongonadal tissues: *Kpna7*, *Rnf17*, *Slc25a31*, *Spocd1*, *Tdrd1* and *Tdrd9*.

Among the genes exhibiting an ovarian enriched expression in zebrafish RNA-seq data, seven were selected for further analysis in medaka using qPCR: *kpna7*, *pabpn11*, *bokb*, *mos*, *lsm14b*, *cdt1* and *orc1* (Figure 7). All genes had predominant expression in the ovary with the exception of *bokb* that was equally expressed in the eyes. Significant testicular expression was also observed for *pabpn11*, *cdt1* and *orc1*.

## 4. DISCUSSION

In the present study, we investigated the vertebrate orthologues of *Drosophila* genes involved in ovarian function. For the first time, we showed, from a list of 971 *Drosophila* genes, that around 80% of them were conserved throughout evolution. We also reported that 73 and 76 orthologues of these genes, in the zebrafish and mouse, respectively, were likely to exhibit preferential ovarian expression, 20 genes being common between both species.

### 4.1 Genes shared by the fly and the vertebrates

*Drosophila* genes studied in this work showed high conservation throughout evolution. Indeed, an increased percentage of vertebrate orthologues of these *Drosophila* genes was found compared to the complete genome (80% versus 42%), in each species investigated. Moreover, orthology percentage was also increased in all other phyla investigated, including sponge, cnidarians, lophotrochozoa, ecdysozoa and echinoderm, suggesting the critical importance of these genes. The other point to consider is that the orthology percentage we measured between vertebrate species and *Drosophila* is an underestimation. Indeed, due to the high stringency of the phylogenetic trees in Ensembl, we evidenced that already described mammal orthologues were not considered in the BioMart tool. For example, surprisingly, *Vasa* and *Mos* *Drosophila* genes were not predicted by BioMart to have an orthologue in vertebrates, although they do possess one (Elis et al. 2008; Gebauer and Richter 1997; Horie and Kotani 2016; Song et al. 2016). Interestingly, the conservation of “ovarian genes” is even higher for our ovarian genes than for “disease-focused genes” that averaged 65% to 70% from *Drosophila* to human (Hu et al. 2011; Wangler et al. 2017).

Among the GO enriched in the conserved genes, mRNA binding was reported as well as GO related to posttranslational regulation (phosphorylation) and signalling pathways. The processes of mRNA biogenesis, translation, alternative splicing, nuclear exportation, cytosolic

localisation and finally degradation are modulated by RNA-binding proteins, which are of primary importance in the oocyte (Akagi et al. 2000). Several mechanisms contributing to this translational control, including conserved RNA-binding proteins that control translation by 3' UTR binding and the assembly of mRNA-protein complexes called mRNPs, have been investigated not only in invertebrates but also in vertebrates (Boateng et al. 2017; Drouilhet et al. 2008; Pushpa et al. 2017; Rosario et al. 2017; Tsukamoto et al. 2017).

#### **4.2 Conserved genes preferentially expressed in the ovary**

In this study, among the fly genes that have orthologues, we have found that the expression of 73 and 76 of them are enriched in, or strictly specific to the ovary, in the zebrafish and mouse, respectively, including 20 genes that are preferentially expressed in the ovary in both vertebrate species. In the 73 and 76 zebrafish and mouse gene lists, respectively, the GO analysis suggested that mRNA binding function is increased compared to the full list of orthologues. Interestingly, among the 20 common genes between the zebrafish and mouse, 4 genes encode proteins involved in RNA binding and/or processing (*pabpn11*, *Lsm14*, *Eif4e1b* and *Cpeb1*). In particular, *pabpn11* binds, stabilises and localises *osk* mRNA in *Drosophila* oocytes and might be involved in similar functions in vertebrates (Vazquez-Pianzola et al. 2011). The *Lsm14* protein, known to associate with RNA to form the core domain of ribonucleoprotein particles, is involved in mRNA translation and distribution in *Xenopus* oocytes (Ladomery and Sommerville 2015). *Eif4e1b* was previously shown to be expressed in the mouse, *Xenopus* and zebrafish oocytes (Evsikov and Marin de Evsikova 2009) and to be involved in mRNA cap binding (Kubacka et al. 2015). *Cpeb1* was also previously shown to regulate mRNA translation in the mouse (Sousa Martins et al. 2016) and *Xenopus* oocytes through the formation of a ribonucleic protein complex (Minshall et al. 2007). Four additional ovarian preferentially expressed genes, common between the zebrafish and mouse, can be related to RNA binding

and processing (*igf2bp1* and *Rnf17*, *Tdrd9* and *plD6*). Indeed, the *igf2bp1* gene encodes a protein that binds mRNA, but to our knowledge, its role in the oocyte is still not known. *Rnf17*, *Tdrd9* and *plD6* are all involved in piRNA processing. Indeed, the invalidation of *Rnf17* (RING finger protein 17) leads to male sterility due to a blockage of spermiogenesis (Pan et al. 2005), likely through the regulation of piRNA content and of PIWI proteins (Wasik et al. 2015). However, the same study showed that females are fertile. So, the role of this conserved gene in oogenesis remains to be determined. Moreover, *Tdrd9* (tudor domain containing 9) encodes an ATP binding RNA helicase, and *plD6* (or MITOPLD) is involved in the silencing of retrotransposons in male germ cells in the mouse (Shoji et al. 2009; Watanabe et al. 2011). A role of *plD6* in piRNA processing was recently shown in the mouse oocyte, even if it seems to be dispensable in females, in contrast to males (Kabayama et al. 2017). In testes, molecular studies of germinal granules strongly implicate them in the transport, storage, localisation, stability and regulation of translation of mRNA (Houston and King 2000; Voronina et al. 2011). Other functions are also represented in this common gene list between the zebrafish and mouse. Indeed, three genes encode serine/threonine kinases (*AuroraC* and *Cdc7*) and a cyclin (*CCNBI*). *AuroraC* and *CCNBI* are known to play a role in the processes of mitosis and meiosis in oocytes (Schindler et al. 2012; Uzbekova et al. 2008), and *Cdc7* is conserved from yeast to human and is essential for cell proliferation and embryonic development (Silva et al. 2006).

Six genes are involved in processes of cytokinesis, general chromosome relocation, DNA replication, recombination and repair and might play similar roles in the oocyte (*Kif23*, *Kif14*, *Ncapg* (or non-SMC condensin I complex, subunit G), *Mcm10* (minichromosome maintenance 10), *Mcph1* and *Dlgap5* (large (*Drosophila*) homolog-associated protein 5) (Bonnet et al. 2015; Geiman et al. 2004; Liang et al. 2010; Liao et al. 2013; Lim et al. 2011; Liu et al. 2013; Reubens et al. 2015; Samwer et al. 2013). The gene encoding Kpna7, a karyopherin involved in the transport of molecules between cytoplasm and nucleus, has been

shown to be necessary for the early steps of embryonic development in the pig (Wang et al. 2012). Lastly, two genes encode transcription factors: *FOXR1*, a forkhead DNA binding protein whose expression was described in human oocytes (Virant-Klun et al. 2013), but its role in oogenesis remains unknown; and *Mybl2*, involved in cell cycle progression and known to suppress autophagy in the pig ovary via the upregulation of the *VDAC2* gene (Yuan et al. 2015). Overall, these genes are highly conserved during evolution. Their specific ovarian expression seems to be well conserved as well and might play a universal role in oogenesis in both invertebrates and vertebrates. Of note, several of the other genes whose expression is highly enriched in the mouse and/or zebrafish ovaries, but under the threshold we fixed, have not been studied in the oocyte so far. Thus, future studies should also focus on their characterisation in oocytes: peptidyl-prolyl cis/trans isomerase *Pin1rt1*, *Spocd1*, T Brachyury transcription factor, *Slc25a31*, *Naa11* or N(Alpha)-Acetyltransferase 11.

#### **4.3 Limitations: false negative genes, no orthologue found**

As mentioned before, one limitation of this study is that the stringency of the method used (Ensembl database and BioMart tool) prevented us, in some cases, from finding an orthologue even when we knew the vertebrate orthologue does exist. Therefore, it is possible that, concerning the supposed genes present in flies but not vertebrates, due to the evolutionary distance between these species, these genes have a true vertebrate orthologue, but that the high stringency of the phylogenetic trees recovered from the Ensembl database does not allow to identify them without any ambiguity (i.e., *Vasa*, *mos* and *brca2* are *Drosophila* genes known to have vertebrate orthologues). Moreover, for the same reason of evolutionary distance, the conserved synteny could not be used between *Drosophila* and vertebrates to complete or strengthen our phylogenetic trees.

#### 4.4 Genes present in the fly and not in vertebrates

Nevertheless, in the list of 971 genes that are involved in ovarian function in the fly, 177 are supposed to have no orthologue in vertebrates. Some of these genes are known to be specific to oogenesis in flies, such as border cell formation and migration (six genes implicated in EGFR signalling, *aos*, *grk*... and four implicated in JAK/STAT signalling, *apt*, *upd*...), germ line stem cell functions (*bam*, *ote*...), oocyte fate determination (*dap*, *ranshi* and *stwl*), ring canal formation with nurse cells (*sosie*, *akap200*...), dorsoventral polarity (*spz*, *trk*...), anteroposterior axis specification (*bcd*, *osk*, *cup*...), chorion and vitelline membrane formation (*cp16*, *cp36*, *psd*...) or dorsal appendage formation (*Gfrl*, *brk*...). One of the most interesting genes is *oskar*. Indeed in *Drosophila*, it seems that there is a hierarchical model for the organisation of germ plasm, *oskar* being necessary and sufficient to initiate a cascade of interactions that assemble germ plasm at its site of localisation (Mahowald 2001). Oskar initiates granule formation likely through direct interactions with Vasa, which in turn recruits Tudor (Anne 2010; Breitwieser et al. 1996), Valois and PIWI factors (Kirino et al. 2009; Nishida et al. 2009). Germ granule mRNAs (nanos, germ cell-less and polar granule components) are transferred from the nurse cells to the oocyte (reviewed in (Becalska and Gavis 2009)). Surprisingly, some other genes are involved in more universal functions, such as apoptosis (*hid*, *rpr*, *Strica*...) or meiosis (*mei-218*, *ord*, *cort*...), suggesting that these genes are specific to these processes in *Drosophila* or insects.

#### 4.5 Genes involved in oogenesis in the fly and expressed in the testis in the mouse

Surprisingly, some genes involved in oogenesis in the fly have a testis specific expression in the mouse, suggesting a role in spermatogenesis, such as *tra2/Rbmx12* and *tud/tdrd6*. It is also the case in the *ovo* gene, which has been shown experimentally to be involved in spermatogenesis but not oogenesis in the mouse (Dai et al. 1998).

## **5. CONCLUSION**

In conclusion, through this evolutionary *in silico* analysis using a *D. melanogaster* model, we have identified several vertebrate orthologues of genes involved in ovarian function and whose expression is highly enriched or specifically expressed in the ovary. Several of these genes are not known to play a role in oogenesis and represent new functional candidates to be tested by invalidation.

## **FUNDING**

This work was supported by The Institut National de la Recherche Agronomique (INRA).

## **ACKNOWLEDGEMENTS**

We acknowledge Megane Bregeon and Fatima Patel for participating in this work.

## 6. REFERENCES

- Akagi T, et al. 2000. Molecular characterization of a mouse heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its tissue-specific expression. *Gene*. 245:267–273.
- Anne J. 2010. Targeting and anchoring Tudor in the pole plasm of the *Drosophila* oocyte. *PLoS One*. 5:e14362. doi: 10.1371/journal.pone.0014362
- Bastock R, St Johnston D 2008. *Drosophila* oogenesis. *Current Biology* 18: R1082–R1087. doi: <https://doi.org/10.1016/j.cub.2008.09.011>
- Becalska AN, Gavis ER. 2009. Lighting up mRNA localization in *Drosophila* oogenesis. *Development*. 136:2493–2503. doi: 10.1242/dev.032391
- Billard R. 1987. The reproductive cycle of male and female brown trout (*Salmo trutta fario*): a quantitative study. *Reprod Nutr Dévelop*. 27:29–44.
- Boateng R, Nguyen KCQ, Hall DH, Golden A, Allen AK. 2017. Novel functions for the RNA-binding protein ETR-1 in *Caenorhabditis elegans* reproduction and engulfment of germline apoptotic cell corpses. *Dev Biol*. 429:306–320. doi: 10.1016/j.ydbio.2017.06.015
- Bonnet A, Servin B, Mulsant P, Mandon-Pepin B. 2015. Spatio-temporal gene expression profiling during in vivo early ovarian folliculogenesis: integrated transcriptomic study and molecular signature of early follicular growth. *PLoS One*. 10:e0141482. doi: 10.1371/journal.pone.0141482
- Braasch I, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 48:427–437. doi: 10.1038/ng.3526
- Breitwieser W, Markussen FH, Horstmann H, Ephrussi A. 1996. Oskar protein interaction with Vasa represents an essential step in polar granule assembly. *Genes Dev*. 10:2179–2188.
- Craig-Bennett MA. 1931. The reproductive cycle of the three-spined stickleback, *Gasterosteus aculeatus*, Linn. *Philos Trans Royal Soc B*. 219:197–279. doi: 10.1098/rstb.1931.0005
- Dai X, et al. 1998. The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes Dev*. 12:3452–3463.
- Drouilhet L, et al. 2008. Use of combined in silico expression data and phylogenetic analysis to identify new oocyte genes encoding RNA binding proteins in the mouse. *Mol Reprod Dev*. 75:1691–1700. doi: 10.1002/mrd.20914
- Eggert B. 1931. Die geschlechtsorgane der gobiiformes und blenniiformes. *Zwiss Zool*. 139:249–558.
- Elis S, et al. 2008. Search for the genes involved in oocyte maturation and early embryo development in the hen. *BMC Genomics* 9:110.
- Evsikov AV, Marin de Evsikova C. 2009. Evolutionary origin and phylogenetic analysis of the novel oocyte-specific eukaryotic translation initiation factor 4E in Tetrapoda. *Dev Genes Evol*. 219:111–118. doi: 10.1007/s00427-008-0268-2
- Extavour C. 2009. Oogenesis: making the mos of meiosis. *Curr Biol*. 19:R489–R491. doi: 10.1016/j.cub.2009.05.015
- Gebauer F, Richter JD. 1997. Synthesis and function of Mos: the control switch of vertebrate oocyte meiosis. *Bioessays*. 19:23–28. doi: 10.1002/bies.950190106
- Geiman TM, et al. 2004. Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res*. 32:2716–2729. doi: 10.1093/nar/gkh589
- Grier HJ, Uribe MC, Parenti LR. 2007. Germinal epithelium, folliculogenesis, and postovulatory follicles in ovaries of rainbow trout, *Oncorhynchus mykiss* (Walbaum, 1792) (Teleostei, protacanthopterygii, salmoniformes). *J Morphol*. 268:293–310. doi: 10.1002/jmor.10518



- Hann HW. 1927. The history of the germ cells of *cottus bairdii* girard. *J Morphol* 43:427–497. doi: 10.1002/jmor.1050430207
- Horie M, Kotani T. 2016. Formation of mos RNA granules in the zebrafish oocyte that differ from cyclin B1 RNA granules in distribution, density and regulation. *Eur J Cell Biol.* 95:563–573. doi: 10.1016/j.ejcb.2016.10.001
- Houston DW, King ML. 2000. Germ plasm and molecular determinants of germ cell fate. *Curr Top Dev Biol.* 50:155–181.
- Hu Y, et al. 2011. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinform.* 12:357. doi: 10.1186/1471-2105-12-357
- Huang da W, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13. doi: 10.1093/nar/gkn923
- Huang da W, Sherman BT, Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4:44–57. doi: 10.1038/nprot.2008.211
- Kabayama Y, et al. 2017. Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes. *Nucleic Acids Res.* 45:5387–5398. doi: 10.1093/nar/gkx027
- Kirino Y, et al. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol.* 11:652–658. doi: 10.1038/ncb1872
- Konietschke F, Placzek M, Schaarschmidt F, Hothorn LA. 2015. nparcomp: an R software package for nonparametric multiple comparisons and simultaneous confidence intervals. *J Stat Softw.* 64:1–17.
- Kubacka D, et al. 2015. Distinct features of cap binding by eIF4E1b proteins. *J Mol Biol.* 427:387–405. doi: 10.1016/j.jmb.2014.11.009
- Kumano G 2012. Polarizing animal cells via mRNA localization in oogenesis and early development. *Development, Growth & Differentiation* 54: 1-18. doi: doi:10.1111/j.1440-169X.2011.01301.x
- Ladomery M, Sommerville J. 2015. The Scd6/Lsm14 protein xRAPB has properties different from RAP55 in selecting mRNA for early translation or intracellular distribution in *Xenopus* oocytes. *Biochim Biophys Acta.* 1849:1363–1373. doi: 10.1016/j.bbagr.2015.10.002
- Lasko P. 2012. mRNA localization and translational control in *Drosophila* oogenesis. *Cold Spring Harb Perspect Biol.* 4:a012294. doi: 10.1101/cshperspect.a012294
- Laver JD, Marsolais AJ, Smibert CA, Lipshitz HD. 2015. Regulation and function of maternal gene products during the maternal-to-zygotic transition in *Drosophila*. *Curr Top Dev Biol.* 113:43–84. doi: 10.1016/bs.ctdb.2015.06.007
- Liang Y, et al. 2010. BRIT1/MCPH1 is essential for mitotic and meiotic recombination DNA repair and maintaining genomic stability in mice. *PLoS Genet.* 6:e1000826. doi: 10.1371/journal.pgen.1000826
- Liao W, et al. 2013. Silencing of DLGAP5 by siRNA significantly inhibits the proliferation and invasion of hepatocellular carcinoma cells. *PLoS One.* 8:e80789. doi: 10.1371/journal.pone.0080789
- Lim HJ, Jeon Y, Jeon CH, Kim JH, Lee H. 2011. Targeted disruption of Mcm10 causes defective embryonic cell proliferation and early embryo lethality. *Biochim Biophys Acta.* 1813:1777–1783. doi: 10.1016/j.bbamcr.2011.05.012
- Liu J, et al. 2013. MKlp2 inhibitor paprotrain affects polar body extrusion during mouse oocyte maturation. *Reprod Biol Endocrinol.* 11:117. doi: 10.1186/1477-7827-11-117
- Mahowald AP. 2001. Assembly of the *Drosophila* germ plasm. *Int Rev Cytol.* 203:187–213.
- Mark M, Rijli FM, Chambon P. 1997. Homeobox genes in embryogenesis and pathogenesis. *Pediatr Res* 42:421–429. doi: 10.1203/00006450-199710000-00001

- Matova N, Cooley L. 2001. Comparative aspects of animal oogenesis. *Dev Biol.* 231:291–320. doi: 10.1006/dbio.2000.0120
- McGary KL, et al. 2010. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A.* 107:6544–6549. doi: 10.1073/pnas.0910200107
- McKearin D, Christerson L. 1994. Molecular genetics of the early stages of germ cell differentiation during *Drosophila* oogenesis. *Ciba Foundation Symposium* 182: 210-219; discussion 219-222.
- Minshall N, Reiter MH, Weil D, Standart N. 2007. CPEB interacts with an ovary-specific eIF4E and 4E-T in early *Xenopus* oocytes. *J Biol Chem.* 282:37389–37401. doi: 10.1074/jbc.M704629200
- Nakamura S, Kobayashi K, Nishimura T, Higashijima S, Tanaka M. 2010. Identification of germline stem cells in the ovary of the teleost medaka. *Science.* 328:1561–1563. doi: 10.1126/science.1185473
- Nishida KM, et al. 2009. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *EMBO J.* 28:3820–3831. doi: 10.1038/emboj.2009.365
- Pan J, et al. 2005. RNF17, a component of the mammalian germ cell nuage, is essential for spermiogenesis. *Development.* 132:4029–4039. doi: 10.1242/dev.02003
- Pasquier J, et al. 2017. Evolution of gene expression after whole-genome duplication: New insights from the spotted gar genome. *J Exp Zool B Mol Dev Evol.* 328:709–721. doi: 10.1002/jez.b.22770
- Pasquier J, et al. 2016. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genom.* 17:368. doi: 10.1186/s12864-016-2709-z
- Pepling ME, Spradling AC. 2001. Mouse ovarian germ cell cysts undergo programmed breakdown to form primordial follicles. *Dev Biol.* 234:339–351. doi: 10.1006/dbio.2001.0269
- Pushpa K, Kumar GA, Subramaniam K. 2017. Translational control of germ cell decisions. *Results Probl Cell Differ.* 59:175–200. doi: 10.1007/978-3-319-44820-6\_6
- R\_Core\_Team. 2015. R: a language and environment for statistical computing. Vienna, Austria: R-project.
- Reubens MC, et al. 2015. Mcm10 is required for oogenesis and early embryogenesis in *Drosophila*. *Mech Dev.* 138:291–299. doi: 10.1016/j.mod.2015.09.002
- Rosario R, Childs AJ, Anderson RA. 2017. RNA-binding proteins in human oogenesis: Balancing differentiation and self-renewal in the female fetal germline. *Stem Cell Res.* 21:193–201. doi: 10.1016/j.scr.2017.04.008
- Samwer M, et al. 2013. The nuclear F-actin interactome of *Xenopus* oocytes reveals an actin-bundling kinesin that is essential for meiotic cytokinesis. *EMBO J.* 32:1886–1902. doi: 10.1038/emboj.2013.108
- Schindler K, Davydenko O, Fram B, Lampson MA, Schultz RM. 2012. Maternally recruited Aurora C kinase is more stable than Aurora B to support mouse oocyte maturation and early development. *Proc Natl Acad Sci U S A.* 109:E2215–E2222. doi: 10.1073/pnas.1120517109
- Shoji M, et al. 2009. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell.* 17:775–787. doi: 10.1016/j.devcel.2009.10.012
- Silva T, Bradley RH, Gao Y, Coue M. 2006. *Xenopus* CDC7/DRF1 complex is required for the initiation of DNA replication. *J Biol Chem.* 281:11569–11576. doi: 10.1074/jbc.M510278200
- Song K, et al. 2016. Expression pattern of mouse Vasa homologue (MVH) in the ovaries of C57BL/6 female mice. *Med Sci Monit.* 22:2656–2663.

- Sousa Martins JP, et al. 2016. DAZL and CPEB1 regulate mRNA translation synergistically during oocyte maturation. *J Cell Sci.* 129:1271–1282. doi: 10.1242/jcs.179218
- Spradling AC. 1993. Developmental genetics of oogenesis. The development of *Drosophila melanogaster*. 1: 1-70. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Spradling A, Drummond-Barbosa D, Kai T. 2001. Stem cells find their niche. *Nature.* 414:98–104. doi: 10.1038/35102160
- Stein DS, Stevens LM 2014. Maternal control of the *Drosophila* dorsal–ventral body axis. *Wiley interdisciplinary reviews. Developmental biology* 3: 301-330. doi: 10.1002/wdev.138
- Tsukamoto T, et al. 2017. LIN-41 and OMA ribonucleoprotein complexes mediate a translational repression-to-activation switch controlling oocyte meiotic maturation and the oocyte-to-embryo transition in *Caenorhabditis elegans*. *Genetics.* 206:2007–2039. doi: 10.1534/genetics.117.203174
- Uzbekova S, et al. 2008. Spatio-temporal expression patterns of aurora kinases a, B, and C and cytoplasmic polyadenylation-element-binding protein in bovine oocytes during meiotic maturation. *Biol Reprod.* 78:218–233. doi: 10.1095/biolreprod.107.061036
- Vazquez-Pianzola P, Urlaub H, Suter B. 2011. Pabp binds to the *osk* 3'UTR and specifically contributes to *osk* mRNA stability and oocyte accumulation. *Dev Biol.* 357:404–418. doi: 10.1016/j.ydbio.2011.07.009
- Virant-Klun I, Knez K, Tomazevic T, Skutella T. 2013. Gene expression profiling of human oocytes developed and matured in vivo or in vitro. *Biomed Res Int.* 2013:879489. doi: 10.1155/2013/879489
- Voronina E, Seydoux G, Sassone-Corsi P, Nagamori I. 2011. RNA granules in germ cells. *Cold Spring Harb Perspect Biol.* 3:a002774. doi: 10.1101/cshperspect.a002774
- Wallace RA, Selman K. 1990. Ultrastructural aspects of oogenesis and oocyte growth in fish and amphibians. *J Electron Microscop Tech.* 16:175–201. doi: 10.1002/jemt.1060160302
- Wang X, et al. 2012. KPNA7, an oocyte- and embryo-specific karyopherin alpha subtype, is required for porcine embryo development. *Reprod Fertil Dev.* 24:382–391. doi: 10.1071/RD11119
- Wangler MF, Hu Y, Shulman JM. 2017. *Drosophila* and genome-wide association studies: a review and resource for the functional dissection of human complex traits. *Dis Model Mech.* 10:77–88. doi: 10.1242/dmm.027680
- Wasik KA, et al. 2015. RNF17 blocks promiscuous activity of PIWI proteins in mouse testes. *Genes Dev.* 29:1403–1415. doi: 10.1101/gad.265215.115
- Watanabe T, et al. 2011. MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev Cell.* 20:364–375. doi: 10.1016/j.devcel.2011.01.005
- Watt FM, Hogan BL. 2000. Out of Eden: stem cells and their niches. *Science.* 287:1427–1430.
- Wheeler B. 2010. *lmPerm*: permutation tests for linear models.
- Yuan J, et al. 2015. MYBL2 guides autophagy suppressor VDAC2 in the developing ovary to inhibit autophagy through a complex of VDAC2-BECN1-BCL2L1 in mammals. *Autophagy.* 11:1081–1098. doi: 10.1080/15548627.2015.1040970
- Zheng P, Dean J. 2007. Oocyte-specific genes affect folliculogenesis, fertilization, and early development. *Semin Reprod Med.* 25:243–251. doi: 10.1055/s-2007-980218

## 7. TABLES

Table 1 : Mouse oligonucleotide sequences

Abbrev.	Accession number	Gene	Forward primer	Reverse primer	bp	E %
<i>Eif4e1b</i>	NM_001033269	eukaryotic translation initiation factor 4E family member 1B	GGTTGGGAATTGTTTTGAGG	CCTGGTACCCAATGATGGTC	185	99.9
<i>Foxr1</i>	NM_001033469	forkhead box R1	CTTCCATCTGATTGCCTTGG	TAACGACCCAGGAGAATTGG	180	108
<i>Kpna7</i>	NM_001013774	Karyopherin alpha 7	TGGAGATGGTGACAGTTTGC	CACCAGAGACCATGTGGATG	173	103.9
<i>Lin28a</i>	NM_145833	lin-28 homolog A	TGGACACCAGAAAAGGGACT	GATGGGGAAAGGACAGACAA	211	102.7
<i>Nanos2</i>	NM_194064	Nanos homolog 2	CGACCAGGCTCATACTCA	GAGCAGAGCAGAGGGGACTA	179	102.2
<i>Pin1rt1</i>	NM_001033768	Protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting 1, retrogene 1	CAAGGGGGTGACAACAGTCT	GCGTTGGTGATGTGATTGAA	193	95.8
<i>Rnf17</i>	NM_001033043	Ring finger protein 17	CCCTTAAGTGACTCGGGAAA	TCTGCAAGCCCCATTTCTAC	200	104.7
<i>Rpl19</i>	NM_009078	Ribosomal protein L19	CCTCCAGGCCAAGAAGGAAG	GGGCAACAGACAAAGGCTTG	151	87.8
<i>Slc25a31</i>	NM_178386	Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 31	GCCTTCCTAGAGACACCTGT	CCCAAGCAACAGGAAAATGT	245	73.6
<i>Spocd1</i>	XM_017320505	SPOC domain containing 1	CAGGTGAAAATGGTCCTGCT	TCTTTGGGACCTTGTTTCGAC	181	118.6
<i>Tdrd1</i>	NM_031387	Tudor domain containing 1	AAACAATCCAACCAACCAAAGT	GAAAGAGGAGGAAGCAGCAA	205	102.1
<i>Tdrd9</i>	NM_029056	Tudor domain containing 9	CTAGTGTGTGATGGGCCAAA	TCCACTCGTAGGGTTTCTCG	157	98.3

Table 2 : Medaka oligonucleotide sequences

Abbrev.	Accession number	Forward primer	Reverse primer
<i>kpna7</i>	XM_011494462.2	ACCAAAAGCCTTCTGTGCAG	GGGAAGGAGACCACATGTGA
<i>bokb</i>	XM_011486847.2	GGTGGTCCAAAGCCGAAATT	CACTCCAGCTCGTCTCCTAG
<i>lsm14b-201</i>	XM_011475365.2	AGTACCAGGAATTCAGGGCC	AACTCTGGATTTGGCGAGGA
<i>mos</i>	XM_020702078.1	GCACCTCCGTCATTCATTGG	CTCTTCACCTTCTTCAGCGC
<i>pabpn11-201</i>	XM_011490104.2	ATCAAAGCCAGGGTTCAGGA	TAGAAAGGCCCCAGGTTGAGG
<i>cdt1-201</i>	XM_004076871.3	ATTGTCTCAGACCACAGCCA	TGCTTCACGAGGGAGATCAG
<i>orc1-201</i>	XM_020702576.1	GGAAATCTGCGAGCACTCAG	CATTCAAGGCCTCCATCAG

## 8. FIGURE LEGENDS

### **Figure 1: *Drosophila* genes possessing a chordate orthologous gene**

Number and percentage of orthologues of the 971 *Drosophila melanogaster* genes and of the complete *Drosophila* gene repertoire were obtained for each species considered using the Ensembl metazoa gene and Ensembl gene databases and the BioMart tool, as described in Material and Methods section. Comparison of orthologue percentage of the 971 selected genes and the *Drosophila* gene repertoire were performed using Chi square tests. \* indicates a significant difference ( $p < 0.0001$ ).

### **Figure 2: Biological process GO enrichment**

Fold enrichment of biological process GO in either zebrafish or mouse orthologue genes compared to the 971 *Drosophila* genes. GO analyses were performed using DAVID 6.8 (<https://david.ncifcrf.gov/home.jsp>). Only significantly enriched biological processes were represented ( $p < 0.05$ , Bonferroni corrected).

### **Figure 3: Molecular function GO enrichment**

Fold enrichment of molecular function GO in either zebrafish or mouse orthologue genes compared to the 971 *Drosophila* genes. GO analyses were performed using DAVID 6.8 (<https://david.ncifcrf.gov/home.jsp>). Only significantly enriched molecular functions were represented ( $p < 0.05$ , Bonferroni corrected).

### **Figure 4: *Drosophila* genes possessing no chordate orthologous gene**

Number and percentage of orthologues of the 178 *Drosophila melanogaster* genes exhibiting no orthologue genes in chordates were obtained for each species considered using the Ensembl metazoa gene and Ensembl gene databases and the BioMart tool, as described in Material and

Methods section. Comparison of orthologue percentages of the selected 178 genes and the *Drosophila* gene repertoire were performed using Chi square tests. \* indicates a significant difference ( $p < 0.0001$ ).

### **Figure 5: Biological process GO enrichment in ovarian genes**

Fold enrichment of biological process GO in mouse orthologue genes exhibiting an ovarian expression compared to the complete list of mouse orthologue genes (1556 genes). Fold enrichment of biological process GO in zebrafish orthologue genes exhibiting an ovarian expression compared to the complete list of zebrafish orthologue genes (1849 genes). GO analyses were performed using DAVID 6.8 (<https://david.ncifcrf.gov/home.jsp>). Only significantly enriched biological processes were represented ( $p < 0.05$ , raw p values).

### **Figure 6: Gene expression in mouse tissues**

Expression of 11 candidate genes was assessed in 10 mouse tissues (O, ovary; T, testis; M, muscle; U, uterus; Li, liver; S, spleen; B, brain; K, kidney; Bl, bladder and H, heart). After mRNA extraction and reverse transcription, qPCR was performed on tissues of three females or males. Expression of candidate genes is normalised by the expression of the *Rpl19* housekeeping gene and presented as mean  $\pm$  SEM of three replicates. Different letters indicate a significant difference ( $p < 0.05$ ). Therefore, if, at least, one letter is common between conditions, there is no significant difference between these 2 conditions (ex : a is different from b but ab is not different from b...).

### **Figure 7: Gene expression in medaka tissues**

Expression of 7 candidate genes was assessed in 10 medaka tissues (O, ovary; T, testis; M, muscle; L, liver; B, brain; K, kidney; I, intestine; E, eye; G, gills and F, fins). After mRNA

extraction and reverse transcription, qPCR was performed on tissues of eight females or males. Expression of candidate genes is normalised using exogenous luciferase transcript abundance in samples and presented as mean  $\pm$  SEM of eight replicates. Different letters indicate a significant difference ( $p < 0.05$ ). Therefore, if, at least, one letter is common between conditions, there is no significant difference between these 2 conditions (ex : a is different from b but ab is not different from b...).



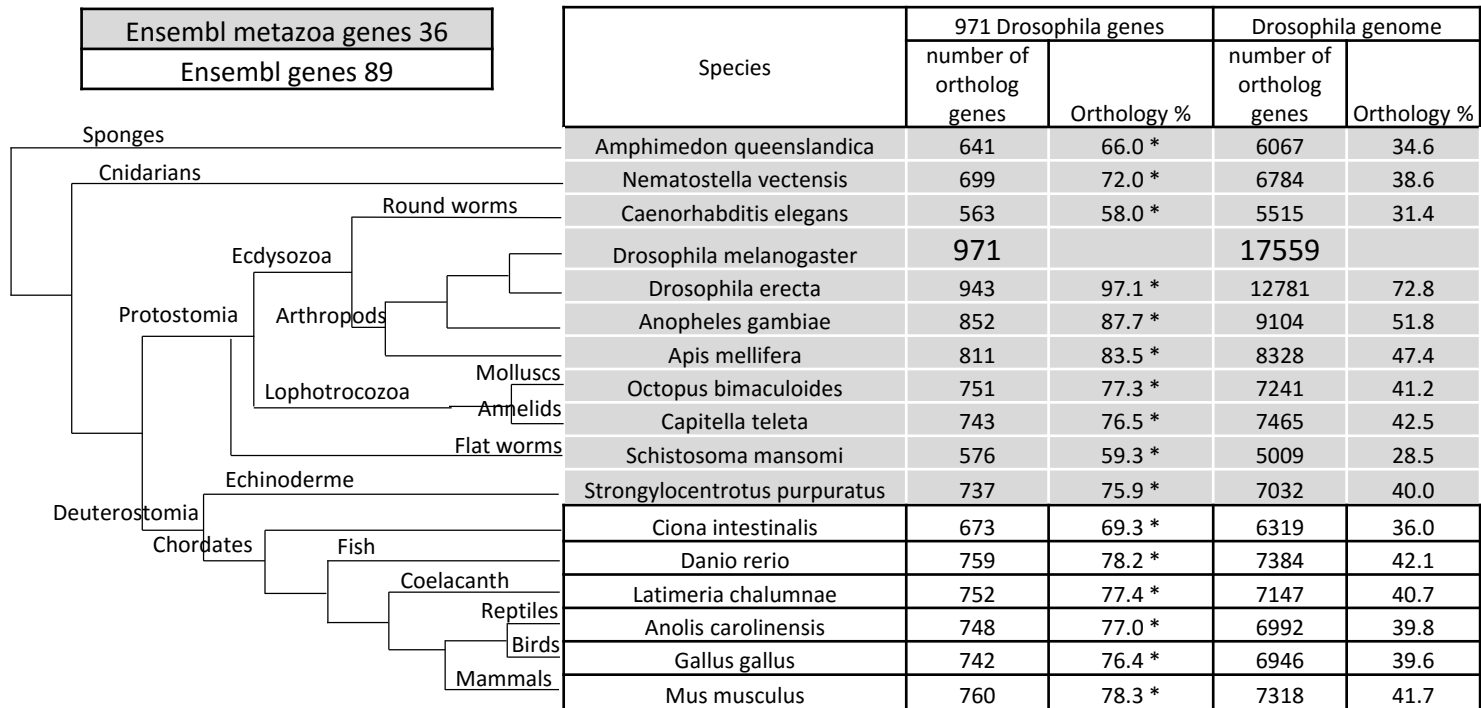


Figure 1

## Biological process GO enrichment

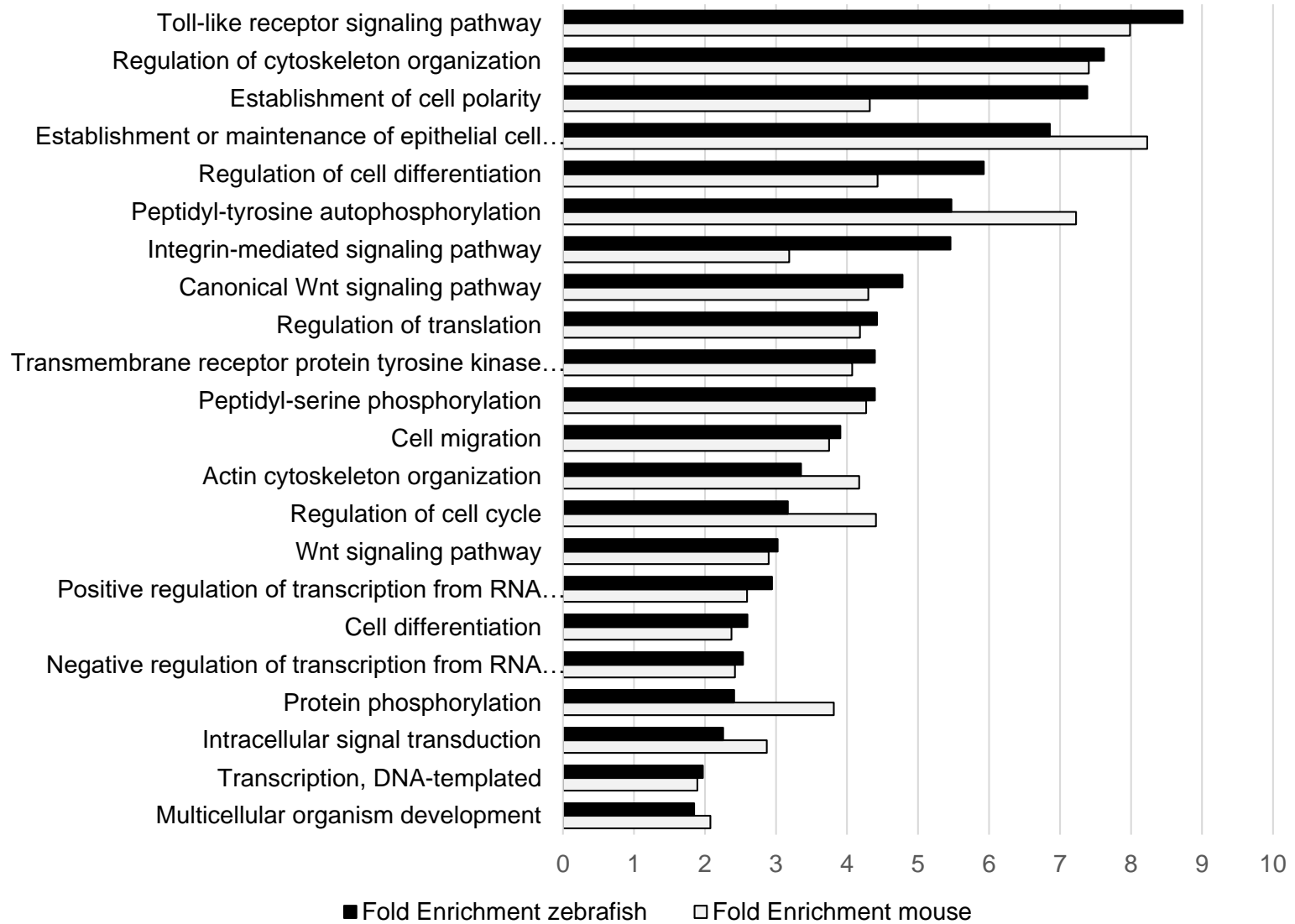


Figure 2

## Molecular function GO enrichment

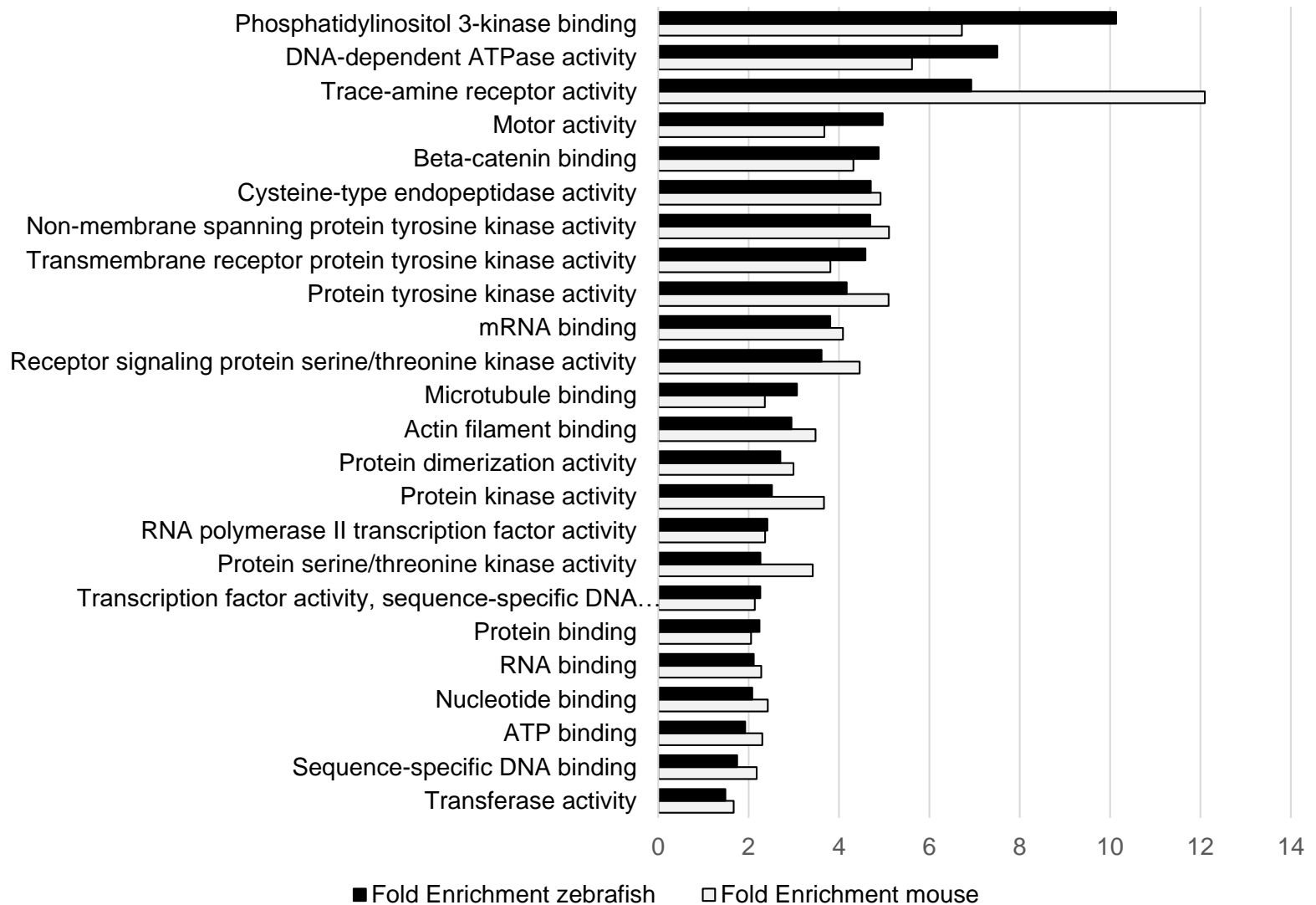
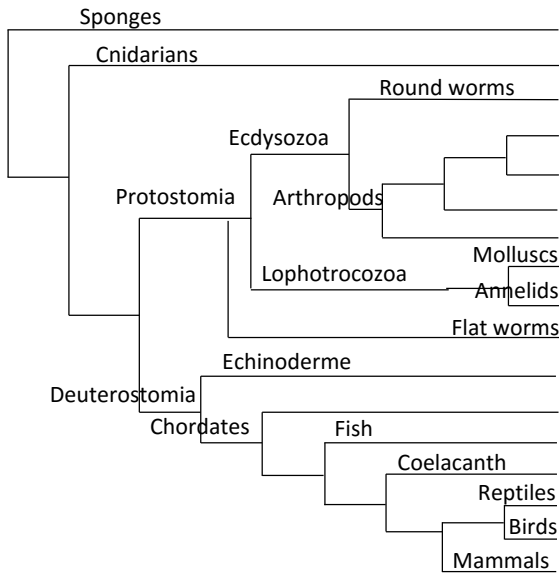


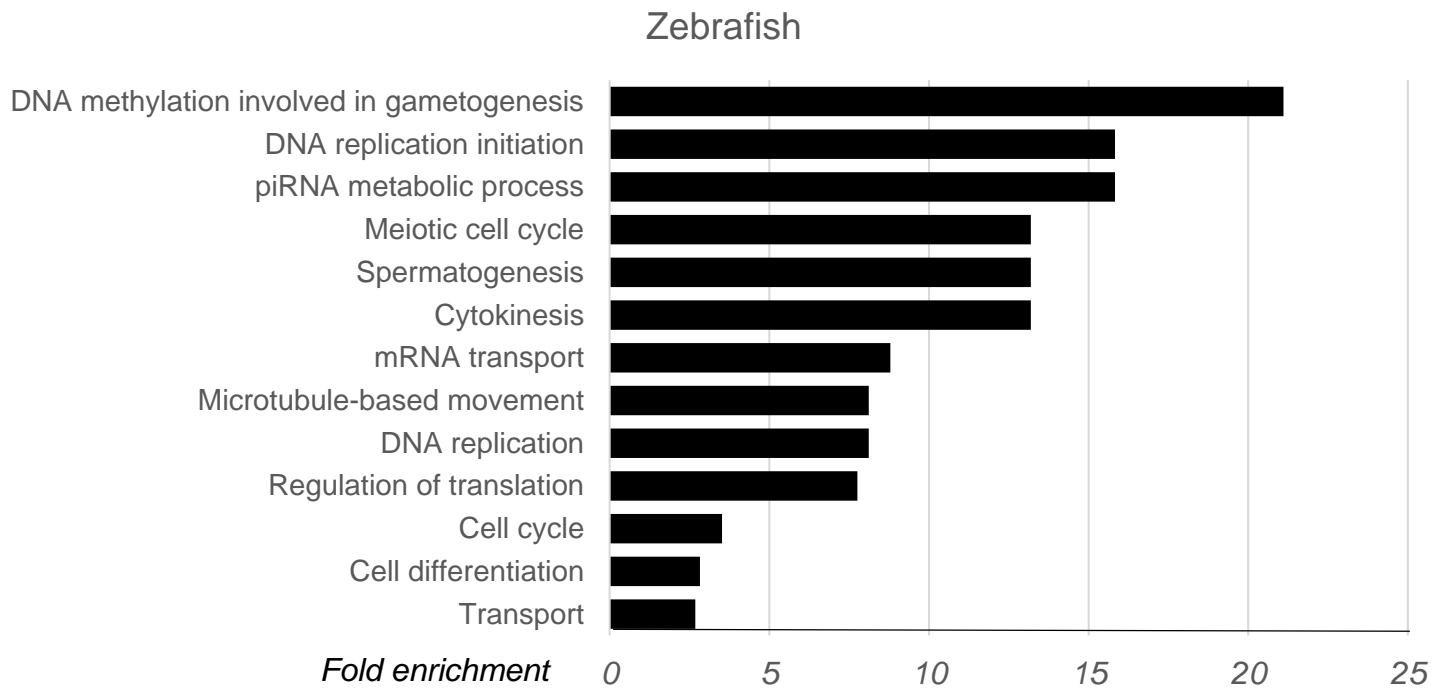
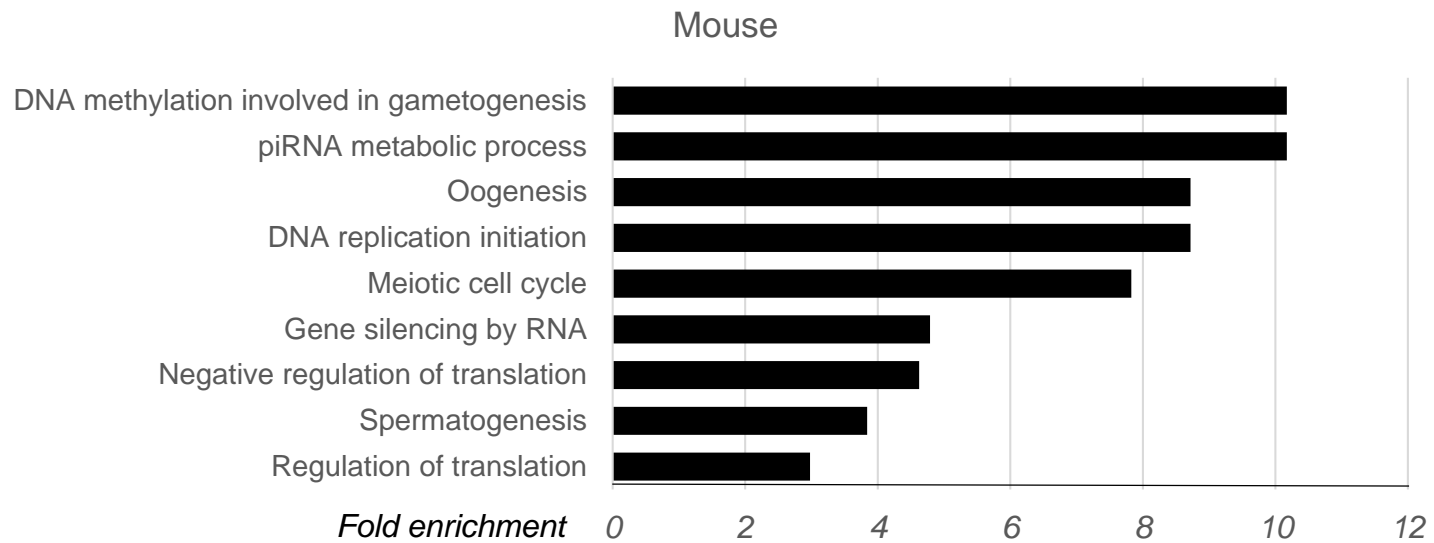
Figure 3

Ensembl metazoa genes 36
Ensembl genes 89



Species	971 Drosophila genes		Drosophila genome	
	number of ortholog genes	Orthology %	number of ortholog genes	Orthology %
Amphimedon queenslandica	26	14.6 *	6067	34.6
Nematostella vectensis	30	16.9 *	6784	38.6
Caenorhabditis elegans	22	12.4 *	5515	31.4
<b>Drosophila melanogaster</b>	<b>177</b>		<b>17559</b>	
Drosophila erecta	168	94.4 *	12781	72.8
Anopheles gambiae	103	57.9...	9104	51.8
Apis mellifera	94	52.8...	8328	47.4
Octopus bimaculoides	42	23.6 *	7241	41.2
Schistosoma mansoni	23	12.9 *	5009	28.5
Capitella teleta	49	27.5 *	7465	42.5
Strongylocentrotus purpuratus	35	19.7 *	7032	40.0
Ciona intestinalis	0	0.0 *	6319	36.0
Latimeria chalumnae	0	0.0 *	7147	40.7
Danio rerio	0	0.0 *	7384	42.1
Anolis carolinensis	0	0.0 *	6992	39.8
Gallus gallus	0	0.0 *	6946	39.6
Mus musculus	0	0.0 *	7318	41.7

Figure 4



**Figure 5**

