# Systematic Imaging Reveals Features of Localized mRNAs and Their Changing Subcellular Destinations in Development

**Helena Jambor[1*], Vineeth Surendranath[1], Alex T. Kalinka[1,2], Pavel Mejstrik[1], Stephan Saalfeld[1,3], Pavel Tomancak[1*]**

[1] Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

[2] present address: Institute of Population Genetics Department of Biomedical Sciences, University of Veterinary Medicine Vienna, Austria

[3] present address: Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA

[*] Correspondence: jambor@mpi-cbg.de, tomancak@mpi-cbg.de

**Abstract**

The asymmetric distribution of cytoplasmic components by mRNA localization is critical for eukaryotic cells and affects large numbers of transcripts. How such global subcellular localization of mRNAs is regulated is still unknown. We combined transcriptomics and systematic imaging to determine tissue-specific expression and subcellular localizations of 5862 mRNAs during *Drosophila* oogenesis. While the transcriptome is stable and alternative splicing and polyadenylation is rare, cytoplasmic localization of mRNAs is widespread. Localized mRNAs have distinct gene features and diverge in expression level, 3'UTR length and sequence conservation. We show that intracellular localization of mRNAs depends on an intact microtubule cytoskeleton and that specifically the posterior enrichment requires the localization of *oskar* mRNA to the posterior cortex. Using cross-tissue comparison we revealed that the localization landscape differs substantially between epithelial, germline and embryonic cells and the localization status of mRNAs also changes considerably within the oocyte over the course of oogenesis.

## Introduction

During development, cells differentiate to become highly specialized units of the metazoan organism. Cell differentiation is often accompanied by polarization and segregation of membranes, cytoplasm and organelles. A powerful mechanism to generate subcellular asymmetries is the localization of mRNAs and their controlled translation into a protein. Localization, which was initially described for *actin* mRNAs in ascidian eggs and chicken fibroblast cells (Jeffery et al., 1983; Lawrence and Singer, 1986), has been reported for developing, dividing and fully differentiated cell types in many organisms including unicellular eukaryotes and protozoans (reviewed in Medioni et al., 2012).

In metazoan cells, the long-range transport of mRNAs in the cytoplasm largely relies on the polarized cytoskeleton and the microtubule minus- and plus-end motor complexes. Chemical or genetic perturbation of the microtubule cytoskeleton strongly impairs mRNA localization. Correspondingly, mRNA enrichment at microtubule minus-ends is aberrant in mutants that affect the dynein motor complex, while plus-end directed

transport requires kinesin molecules (reviewed in Bullock, 2011; Medioni et al., 2012). For several mRNAs it was further shown that they locally enrich by trapping to a pre-localized anchoring activity (Forrest and Gavis, 2003; Sinsimer et al., 2011) or by hitch-hiking along with a localization-competent mRNA (Glotzer et al., 1997; Jambor et al., 2011).

Localization-competent mRNAs carry cis-regulatory sequences also termed mRNA zipcodes, which are often present in the 3'UTR of the transcript (reviewed in Jambhekar and Derisi, 2007). These signals are composed of few to several hundred nucleotides that usually form secondary or tertiary RNA structures (Bullock, 2011; Jambor et al., 2014; Serano and Cohen, 1995). The mRNA localization signals are specifically recognized by RNA binding proteins (Bullock et al., 2010; Chao et al., 2010; Dienstbier et al., 2009) that initiate the formation of transport competent ribonucleoproteins (RNPs) (Dienstbier et al., 2009; Dix et al., 2013). mRNAs can also harbor two antagonizing localization signals that act consecutively in cells and direct mRNAs sequentially to opposing microtubule ends (Ghosh et al., 2012; Jambor et al., 2014). Although no consensus zipcode for mRNA transport has been identified, it was shown for a few, well-characterized mRNA localization elements that they are active in several cell types (Bullock and Ish-Horowicz, 2001; Jambor et al., 2014; Kislauskis et al., 1994; Snee et al., 2005), suggesting that the mRNA transport machinery is widely expressed and mRNA localization elements function in a cell-type independent manner.

In the oocyte of *Drosophila melanogaster (D.melanogaster),* the embryonic axis is set up through localization and restricted translation of *oskar, bicoid* and *gurken* mRNAs that encode maternal determinants. Due to its large size and the abundance of genetic tools, the oocyte has become the key model system to study mRNA transport in cells. In early oogenesis, *bicoid, gurken* and *oskar* mRNAs enrich in the young oocyte, while in later stages *bicoid* and *gurken* are found anteriorly and *oskar* mRNA specifically enriches at the posterior pole (Berleth et al., 1988; Ephrussi et al., 1991; Neuman-Silberberg and Schüpbach, 1993; St Johnston et al., 1989).

Our detailed mechanistic understanding of mRNA localization comes from few but well studied examples, yet more recent work suggest that this phenomenon is a widespread cellular feature that affects a large proportion of expressed mRNAs (Blower et al., 2007; Cajigas et al., 2012; Lecuyer and Tomancak, 2008; Lecuyer et al., 2007; Shepard et al., 2003; Zivraj et al., 2010). How the cell distinguishes localized mRNAs from ubiquitous transcripts for their transport to specific subcellular destinations remains enigmatic. Co-packaging of several mRNAs, which could increase the efficiency of mRNA transport, has been shown in two cases (Jambor et al., 2011; Lange et al., 2008) but seems to be less common in neuronal and embryonic cell types where mRNA localization affects numerous mRNAs (Amrute-Nayak and Bullock, 2012; Mikl et al., 2011). Finally, we do not know whether mRNA localization is stable or differs between cell types.

Here we exploit the power of the *Drosophila* ovarian model to unravel the global landscape of mRNA localization in this tissue by systematic imaging and compare it with previously published embryo data (Lecuyer et al., 2007). By combining stage specific mRNA sequencing with genome-wide fluorescent in situ hybridizations (FISH) we find that the germline cells of the *Drosophila* ovary show only little transcriptional change but prevalent mRNA localization, suggesting that germline cells rely on posttranscriptional regulation rather than transcriptional control for differentiation.

Using a genome-wide, cross-tissue comparative approach we find that the cytoplasmic localization status of the majority of mRNAs differs in germline, embryonic and epithelial cells. Moreover, mRNA localization also diverges within one cell over time. The set of localized transcripts however shows characteristic gene level features such as longer and highly conserved 3'UTRs that clearly distinguish localized and ubiquitous mRNAs and is most pronounced in the class of posterior localized mRNAs. While all subcellular mRNAs tested require an intact microtubule cytoskeleton for their transport, posterior mRNAs further depend on initially localized *oskar* mRNA. We present here the entire transcriptomic and imaging data for over 6000 mRNAs as an open access resource to the scientific community through a rich, interactive web page interface - the Dresden Ovary Table (DOT).

## Results

### Subcellular mRNA localization is widespread in ovarian cells.

To systematically determine the subcellular distributions of mRNAs in ovarian cells we developed a mass-isolation protocol to isolate ovaries at large scale and used them for genome-wide fluorescent in situ hybridization (FISH) experiments. Using an established probe collection (Tomancak et al., 2007) we performed 6091 in situ experiments representing 5862 *D.melanogaster* genes (Figure 1A) and imaged cell specific gene expressions and subcellular localization patterns (Figure S1A). The expression patterns were described using a controlled vocabulary (CV) where recognizable anatomical and subcellular features are hierarchically linked[1] (Lecuyer et al., 2007; Tomancak et al., 2002).

The ovary is organized into ovarioles each containing egg-chambers at different developmental stages. At the ovariole tip, a cystoblast divides to produce sixteen germline cells of which one becomes the oocyte and 15 cells differentiate into nurse cells. Initially the germ cells maintain cytoplasmic connections that allow transport of RNAs and proteins, produced in the transcriptionally active nurse cells, into the oocyte. A single layer of somatic epithelial cells surrounds the germ cells and together they form the egg-chamber. We imaged the gene expression patterns in both germline and somatic cell types. The expression patterns were documented by acquiring z-stacks of multiple egg chambers at each stage of oogenesis.

Complementing the qualitative assessment of mRNA expression, we also analyzed the ovarian transcriptome at four time-points including the earliest stage of embryogenesis before the onset of zygotic transcription. Using two complementary deep sequencing methods (RNAseq and 3Pseq, see Experimental Procedures) we identified the expressed transcript isoforms, the 3'UTR ends and the expression levels. RNAseq and 3Pseq datasets were in good agreement and served as biological replicates (Pearson correlation 0.71, Figure S2A).

This entire dataset, including more than 30,000 3D images (Figure S1A), the CV description of the expression patterns, the associated RNAseq and 3Pseq expression quantification and cross-references to similar resources in other tissues, are presented in a publicly accessible database[2] (Figure S1B).

Of the annotated genes, 3624 mRNAs showed in situ signal (Figure 1A) and the vast majority of genes were also detectable

[1] http://tomancak-srv1.mpi-cbg.de/cgi-bin/ovary_annotation_hierarchy.pl

[2] http://tomancak-srv1.mpi-cbg.de/DOT/main

in transcriptomic experiments, indicating a high true-positive rate of the screen (Figure S4A). In 6% of the cases we detected FISH signal despite a below-cutoff deep sequencing signal. Since expression in small subset of cells of the ovary may evade detection by deep sequencing, we keep these potential false-positives in the downstream analysis. Among the mRNAs detected by FISH, 2357 (65%) were expressed ubiquitously at all time-points and 1290 mRNAs (35%) showed restricted expression patterns and/or localization in the ovary (Figure 1A). We broadly classified these into three mutually exclusive, high-level categories (Supplementary Table S1): 1. cell specific patterns ("cellular", n=309, 13%; Figure 2.1), 2. subcellular localization patterns ("subcellular", n=790, 22%; Figure 2.2), and 3. nuclear enriched mRNAs ("nuclear", n=191, 5%; Figure 2.3). These broad groups are biologically distinct from each other as shown by GO-term enrichment analysis (Figure 1B). While the ubiquitous set was enriched for GO-terms describing all major cellular processes, subcellular mRNAs were associated with reproductive processes, cytoskeleton organization and cell cycle regulation. The cellular gene set, being mostly expressed in the somatic cells, was enriched for GO-terms describing epithelial development, lipid trafficking and cuticle formation. Interestingly, the 191 nuclear RNAs were enriched for RNA regulatory processes.

Genes are typically expressed at several time-points in development. We therefore next asked for the ovary gene sets when and in which tissue they are expressed during embryonic development. For this we used the existing BDGP database of embryonic expression patterns (Tomancak et al., 2007). We observed that ubiquitous genes largely remain ubiquitous or exhibit the abundant, broad endo-mesodermal expression pattern in the embryo (Figure 1C, circle symbol). Genes patterning the somatic epithelium in the ovary (cellular pattern) are predominantly expressed in the various epithelial cell types of the late embryo (Figure 1C, plus symbol), whereas nuclear genes of the ovary adopt ubiquitous expression during early embryonic stages (Figure 1C). Interestingly, subcellular genes of the ovary are highly expressed in the nervous system throughout embryonic development (Figure 1C, star symbol; subcellular gene sets in detail are shown in Figure S3) suggesting a high degree of relatedness of these polarized tissues.

**Specific mRNA distributions of cellular, subcellular and nuclear gene sets.**

The imaging of the cellular genes revealed tremendous diversity of patterns in the somatic epithelium. These cells are heterogeneous and undergo dramatic morphological changes. Reflecting this, many mRNAs are specifically expressed in subgroups of epithelial cells and at specific time-points (Figure 1D, Figure 2.1A-O). The largest co-expressed groups are seen in the anterior and posterior somatic cells involved in setting up the anterior-posterior axis of the egg-chamber (Figure 1D, Figure 2.1E,F,K,M) and in the migratory border cells (Figure 1D, Figure 2.1I). We also found instances of cell cycle regulated genes in the dividing somatic cells (Figure 2.1 C,D) and in the endocycling germline nurse cells (Figure 2.1C). An interesting example is *Inx2* mRNA, which is required for formation and separation of the cyst (Mukai et al., 2011) and which we found expressed in late oogenesis specifically in the somatic cells separating nurse cells and the oocyte (centripetally migrating cells, Figure 2.1N).

In contrast to the diversity of cellular patterns, analysis of the subcellular enrichments revealed abundant mRNA localization that is limited to small number of subcellular domains (Figure 1E and Figure 2.2A-P). By far the largest group, 591 mRNAs, was enrichment in the oocyte

portion of the syncytial egg-chamber during early oogenesis (Figure 2.2.A-D). Such oocyte-enriched mRNAs are transported from the transcriptionally active nurse cells into the oocyte where they co-localize with the microtubule minus ends (Pokrywka and Stephenson, 1995). At mid-oogenesis the oocyte establishes its own polarized microtubule cytoskeleton; at this stage, we observed 106 mRNAs enriched towards the anterior and 119 mRNAs enriched at the posterior pole (Figure 1E), corresponding to where the microtubule minus and plus ends are enriched (Theurkauf et al., 1992). The quality of these localizations ranged from tight to diffuse association at the anterior-dorsal, the entire anterior or the posterior cortex (Figure 2.2.M-P). An example for a novel anterior mRNAs is *yemalpha*, which encodes an oocyte nuclear protein that is required for female meiosis (Meyer et al., 2010); among the novel posterior mRNAs is *fs(1)N,* which was identified in an early screen for female sterile mutations (Mohler, 1977). mRNAs were also detected, although less frequently, in subcellular domains of the nurse (Figure 1E, Figure 2.2.I-J) and epithelial cells (Figure 1E, Figure 2.2.D,G,K,L).

In addition to uncovering numerous mRNAs localized to known localization sites, we also observed previously unknown sites of mRNA accumulation in the ovary. We infrequently detected mRNAs that showed cytoplasmic granules (Figure 2.2F), were depleted from the oocyte (Figure 2.2D), showed cortical enrichment (Figure S4D) or formed ring-like structures in early egg-chambers (Figure 2.2G) or follicle cells (Figure S4D). 191 RNAs were enriched in temporally and spatially specific patterns in the nuclei of ovarian cells. Nuclear RNAs were predominantly detected in endocycling, polyploid nurse cells (Figure 2.3A-D), but also in the nuclei of epithelial cells and in 29 cases in the oocyte nucleus (Figure 2.3E-F, Figure S4D). In nurse cells, the nuclear patterns varied from a ring-like signal in the proximity of the nuclear membrane (Figure 2.3A), to numerous foci in a large area of the nucleus (Figure 2.3C) and even widespread distribution in the nucleoplasm (Figure 2.3D). These patterns were not correlated with chromosomal position (Figure S4B). The number of nuclear RNA foci increased over time, coinciding with the increase in nurse cell ploidy. Interestingly, also the precursors of micro RNAs (pre-miRNAs) and the long non-coding RNAs (lincRNAs) displayed varying degrees of nuclear enrichments (Figure S4C).

Taken together, by systematic 3D imaging of thousands of mRNAs we found that subcellular mRNA localization in ovaries is widespread and affects 22% of expressed transcripts and thus likely plays a major role in the orchestration of oogenesis progression and oocyte differentiation.

**The ovarian transcriptome is invariable over the course of oogenesis.**

We next asked whether there are global changes in the transcriptome that accompany the tissue specific diversity of subcellular mRNA enrichments. Analysis of our stage specific mRNA and 3' end sequencing datasets revealed that about half of the *D.melanogaster* genes were expressed at each sampled time point, ranging from 56% of expressed genes at early oogenesis to 41% at the onset of embryogenesis (Figure 3A and experimental procedures). This finding is in agreement with results from gene expression analyses of whole ovaries measured by microarray (Chintapalli et al., 2007) and RNAseq (Graveley et al., 2011). The vast majority of the expressed transcripts, 85%, were detectable at every time point from early oogenesis until embryogenesis (Figure 3B) and the expression levels across time points were highly correlated (Figure 3C), suggesting that the transcriptome remained

constant throughout oogenesis.

A significant up- or down regulation of gene expression levels was only observed for 626 transcripts (p-value adjusted for multiple testing < 0.1, Figure 3C red data points, Supplementary Table S2-4). GO-term enrichment analysis showed that differentially expressed genes were largely associated with regulation and formation of the extracellular matrix, vitelline membrane and cuticle (Figure 3C, Figure S2B), consistent with their expression in the somatic epithelial cells that are involved in formation of the protective layers of the egg. Indeed, down regulation of gene-expression most prominently occurred at the transition to embryogenesis, when the somatic cells have undergone regulated apoptosis (Nezis et al., 2002). Known germline genes were largely absent among differentially expressed genes with the notable exceptions of *exu* and *nanos* that showed up-regulation over the course of oogenesis (Supplementary Table 2-4, Figure S2C).

We did not detect shortening or lengthening of the 3'UTRs across oogenesis and at the transition to embryogenesis (Figure 3E, Supplementary Table S7). Also, the number of transcript ends remained constant (Figure S2E) and the vast majority (>99%) of genes showed no change in expression of alternatively spliced isoforms across the sampled period of development (Figure 3D, Figure S2D,F, Supplementary Table S5-6). In summary, the expressed mRNAs and their expression levels show little variation between egg chamber formation during early oogenesis and the onset of zygotic transcription in the embryo. Changes that do occur mostly affect the somatic epithelial cells. This data suggests that transcriptional changes, alternative splicing or 3' end usage alone cannot account for the changing subcellular mRNA localizations in germline cells.

**Subclasses of localized mRNAs share common features of gene architecture, expression level, evolutionary conservation and function.**

With a large and diverse collection of localized mRNAs at hand we next asked whether there are global gene features that could set localized mRNAs apart from ubiquitous ones. Ovarian expressed mRNAs differed in their expression levels over several orders of magnitude. We therefore compared the distributions of expression levels per annotation class as measured by 3Pseq counts. Consistent with the FISH data, "no signal" mRNAs had negligible expression, while ubiquitous and subcellular mRNAs were overall comparable to each other (Figure 4A and Figure S5A). Among the subcellular mRNAs however, the posterior class was significantly higher expressed than all other localized mRNA classes and this was consistent at all time-points of oogenesis (Figure 4A-A', Figure S5A). Even the related class, mRNAs localized to the anterior of the oocyte, show significantly lower expression levels compared to the posterior ones (p-value: 3.9e-6).

To investigate what distinguish ubiquitous and subcellular mRNAs on the genome level, we compared the gene-level variables of each localization class. mRNA localization signals are thought to reside primarily in the 3'UTR of genes (reviewed in Medioni et al., 2012). We therefore compared the 3'UTR lengths of the different gene sets defined by our 3P-seq data and found that subcellular genes have a significantly longer 3'UTR compared to ubiquitous genes (bootstrapped p-value: 0; Figure 4B, B', see Extended Experimental Procedures). Similarly to expression levels, the 3'UTRs of posterior genes were longer compared to anterior genes, even though both classes are localized (bootstrapped p-value: 0.0018; Figure 4B, B''). In fact, not only the 3'UTR, but also the total gene, 5'UTR, exon and

intron lengths were significantly longer among subcellular compared to ubiquitous genes and again, the posterior genes were in every aspect significantly longer than the anterior genes (Kolmogorov-Smirnov p-value: <0.01, Figure 4C-D, Figure S5B,C,E,I). In addition, the number of exons and introns was higher in posterior than in ubiquitous class genes (Figure S5D,G). The intron proportion was highest in the posterior class genes (Figure S5H), while the exon proportion did not increase and was instead highest in ubiquitous genes (Figure S5F). In summary, localized mRNAs have longer 3'UTRs and overall longer genes due to more non-coding features (introns), and this trend was most pronounced in posterior genes. Particularly the observation that posterior genes have the highest intron proportion and length is interesting considering that the posterior localization signal of *oskar* mRNA requires the assembly of an exon junction complex (EJC) through splicing (Ghosh et al., 2012) and that the stable deposition of this complex was correlated with a larger intron proportion (Ashton-Beaucage et al., 2010).

Not only were the 3'UTRs of posterior mRNAs longer, but in addition they also showed the highest degree of evolutionary conservation when comparing 24 *Drosophila* species (Figure 4E and Extended Experimental Procedures). Again, both the difference between ubiquitous and subcellular and between anterior and posterior mRNAs was significant (bootstrapped p-values: 0 and 0.0032 respectively; Figure 4E',E" and Extended Experimental Procedures).

We next analyzed whether the proteins encoded by co-localized mRNAs were more likely to interact with each other, as suggested by the close proximity of their transcripts in the cell. Oocyte enriched and posterior localized genes share significantly more protein-protein interactions (bootstrapped p-values: 3e-5 and 0.0099 respectively, see Extended Experimental Procedures) while anterior localized genes do not (bootstrapped p-values: 0.249), suggesting that there is a stronger functional relationship between posterior localized genes (Figure 4F).

To gain insight into potential biological functions of mRNA localization classes, we performed GO-enrichment analysis (Figure 4G). As expected oocyte enriched, anterior and posterior mRNAs were linked to reproductive and patterning processes. In addition, both oocyte enriched and anterior gene sets showed high enrichment for cytoskeleton regulation. While oocyte enriched mRNAs were associated with microtubule and actin cytoskeleton terms, anterior class mRNAs only contained terms describing microtubule association. Anterior mRNAs were also associated with the cellular component term chromosome and although with lower significance, also enriched for terms associated with cell cycle regulation. This is interesting considering that anterior mRNAs are in the vicinity of the oocyte nucleus that is undergoing meiosis. In contrast, the posterior mRNAs lacked enrichment of cytoskeletal terms but were strongly associated with signaling, cell fate commitment and membrane organization. This is consistent with known patterning processes involving communication between somatic and germline component of the ovary at the posterior pole (Roth et al., 1995) and with data showing that at the posterior pole the membrane organization changes dramatically as the germ plasm is being assembled (Tanaka and Nakamura, 2008; Vanzo et al., 2007).

Thus, irrespective of their specific subcellular destination, localized mRNAs tend to have longer genes with more introns/exons and longer 3'UTRs with higher conservation compared to ubiquitous mRNAs. Furthermore, compared to mRNAs localized to the anterior end of

the oocyte, posterior mRNAs are characterized by higher expression levels, longer, more complex gene models, a higher degree of evolutionary sequence conservation in their 3'UTRs and greater overlap of protein function.

**RNA localization depends on microtubules, the posterior pathway machinery and *oskar* mRNA.**

Transport of mRNAs towards the anterior and the posterior pole has been elucidated for a few exemplary mRNAs in great detail. The maternal determinants *oskar*, *bicoid* and *gurken* require polarized microtubules for their subcellular distribution (Cha et al., 2002; Pokrywka and Stephenson, 1995; Saunders and Cohen, 1999; Theurkauf et al., 1993). To address whether, apart from sharing global features, the new anterior and posterior mRNAs use the same cellular machinery for their cytoplasmic transport, we probed their distribution in colchicine treated egg-chambers. The localization of both anterior and posterior mRNAs was severely affected (Figure 5A-D', Figure S6A-B,F, Supplementary Table 8). In contrast, RNA foci in the nucleoplasm that lacks a microtubule cytoskeleton, were unaffected by the colchicine treatment (Figure S6E-E') while outer nuclear envelope-associated mRNAs do delocalize (Figure S6C-D'). We confirmed the necessity of an intact cytoskeleton by probing the distribution of selected novel posterior mRNAs in mutant egg-chambers that prematurely depolymerize microtubules (*Spire*$^{RP}$). In these egg-chambers localization of the maternal determinant mRNAs, including *oskar* mRNA, is abolished (Ephrussi et al., 1991; Manseau and Schupbach, 1989; Neuman-Silberberg and Schüpbach, 1993; Theurkauf, 1994) and also the novel posterior mRNAs fail to localize (Figure 5E-I). This suggests that the newly identified anterior and posterior mRNAs use microtubules for localization, and require the same transport machinery as *oskar, gurken* and *bicoid* mRNAs.

Posterior transport of *oskar* mRNA requires components of the EJC complex and the RNA binding protein Staufen while maintenance of *oskar* localization beyond stage 9 needs Oskar protein to anchor the mRNA (Ephrussi et al., 1991; Hachet and Ephrussi, 2004; Micklem et al., 2000; St Johnston et al., 1991; van Eeden et al., 2001; Vanzo and Ephrussi, 2002). The posterior enrichment of the selected candidate mRNAs was severely reduced in egg-chambers mutant for EJC components (*Btz*$^{1}$), Staufen (*Stau*$^{D3}$) and Oskar protein (*osk*$^{84}$*/Df(3R)p*$^{XT103}$) and strongly resembled the mis-localized *oskar* mRNA (Figure 5E-I). Thus the novel posterior mRNAs require the same proteins for their localization as *oskar* mRNA.

To investigate whether the novel candidate mRNAs use the posterior transport machinery independently of *oskar* mRNA we used a genetic combination in which egg-chambers lack any localized *oskar* mRNA (Jenny et al., 2006). None of the new posterior mRNAs showed localization in these egg-chambers (Figure 5E-I, Figure S6G-H) and instead some, for example *TwdlG* mRNA, were mislocalized to the anterior cortex of the oocyte. These results indicate that the novel posterior mRNAs depend on *oskar* mRNA for their localization and are mis-localized in any egg-chambers that do not support or perform *oskar* mRNA localization. It is possible that these mRNAs require localized *oskar* due to its function in recruiting and stabilizing microtubule plus-ends at the posterior pole (Zimyanin et al., 2007) or that they hitch-hike along with *oskar* mRNA possibly in a large transport granule.

**Comparison of the mRNA localization landscape across tissues reveals its variability during development.**

To address whether mRNA localization is

preserved across cell types we took advantage of the wealth of FISH data now available for *Drosophila* and combined our data for the somatic epithelial cells and the germline cells of the ovary with the FISH screen performed on embryonic cells (Lecuyer et al., 2007). Taken together these screens covered 9114 genes of which 5852 were expressed and 1674 mRNAs showed subcellular localization (Figure 6A). Those mRNAs that localized at least at one time-point either during oogenesis or embryogenesis were enriched for genes associated with cytoskeletal organization, cell cycle control, the membrane system and DNA/RNA/nucleotide binding (Figure S7A). The probe sets in the ovary and embryo datasets are not entirely overlapping. Therefore to specifically ask, which mRNAs showed localization in all three tissues, we filtered the datasets for mRNAs that were probed by FISH in all three cell-types and showed subcellular localization (n=720; Figure 6A). Across the three cell-types a meager five mRNAs were constitutively localized and only 111 mRNAs were localized in at least two cell types (Figure 6B). In each cell type the biggest localization class was mRNA enrichment at microtubule minus ends: 71 mRNAs localized at the apical membrane in embryos, nine transcripts localized apically in somatic epithelial cells and 291 mRNAs enriched in the early oocyte. However, of those only three mRNAs were constitutively localized (*Dok, Sdc, CG12006;* Figure 6C). This small overlap between mRNAs at equivalent subcellular sites was not a special feature of the mRNAs localized to microtubule minus ends, as also for example nuclear enrichment is not a constant feature across the cell types (Figure S7B). Thus, most subcellular localizations were transient during development and 1674 mRNAs therefore have "localization potential" but are not constitutively localized.

The expression of localized and non-localized mRNA isoforms of the same gene was previously shown to differentially regulate mRNA localization (Horne-Badovinac and Bilder, 2008; Whittaker et al., 1999). Among the subcellular mRNAs, 55% have more than one isoform (Figure S2F). Yet, during oogenesis we observed no change in alternative transcript expression (Figure 3D and Figure S2F) and the ubiquitous gene set showed similar transcript diversity as subcellular genes. Thus alternative isoform usage cannot fully account for diverging mRNA localizations across germline and soma cells.

Since mRNAs change localization so dramatically across cell types, we next wondered if mRNAs remained localized within one cytoplasm. To this end we compared mRNAs localized within the oocyte over time. We sorted mRNAs by microtubule association for each time point of oogenesis and clustered the resulting matrix to reveal temporal trends in mRNA localization status within the oocyte (Figure 6D) and also in the early embryo (Figure S7C). We made two major observations. Firstly, mRNA localization is not a stable property of mRNAs within one cell and only ~20% of the mRNAs were localized constitutively (Figure 6D). Instead, mRNAs change their specific subcellular localization over time. The majority of mRNAs were transported towards microtubule minus ends during early oogenesis (Figure 6E-G), however in total less than 100 mRNAs remained localized later (Figure 6E-F'). Most minus-end transcripts switched to a ubiquitous state at mid-oogenesis (Figure 6G); some mRNAs change to plus-ends in the stage 9 oocytes (Figure 6F) and resemble the localization pattern of *oskar* mRNA (Ephrussi et al., 1991). We also observed that mRNAs could adopt de-novo plus-end localization at stage 9 or stage 10 after having been entirely ubiquitously distributed during early oogenesis (Figure

6H-H'). The plus-end accumulation at stage 10 resembles *nanos* mRNA localization, which enriches at the posterior pole at stage 10 (Forrest and Gavis, 2003; Wang et al., 1994). Such de-novo accumulation of a previously ubiquitous or plus-end mRNA was not observed at minus-ends.

Secondly, minus-end mRNAs rapidly de-localized decreasing from 99 mRNAs at stage 8 to 39 mRNAs at stage 10. In contrast, plus-end mRNAs increased from 68 localized mRNAs at stage 9 to 109 mRNAs at stage 10. This trend was further accentuated when looking into early embryogenesis where only two mRNAs remained at the anterior pole (*bcd, lok*; Figure S7C). A rise in new minus-end accumulation was observed only after initiation of zygotic transcription and cellularization of the embryo. This could point to a possibly interesting link between transcription and minus-end localization of mRNAs, analogous to the known link between nuclear events and microtubule plus-end localization (Besse et al., 2009; Hachet and Ephrussi, 2004).

We observed only few changes in isoform expression or alternative poly-adenylation in subcellular localized transcripts (Figure 3D-E, Figure S2D-F). Thus, expression of different mRNA variants with or without a localization element cannot account for the large degree of mRNA re-localization in the oocyte. The global analysis of two genome wide localization datasets thus revealed that mRNA localization status is changing in different cell types but also within one cell, suggesting an intricate and cell-type and developmental time-point specific cytoplasmic regulation of mRNAs.

## Discussion

We combined stage specific mRNA sequencing with genome-wide in situ hybridization to comprehensively analyze mRNA expression and localization over the course of Drosophila oogenesis. The deep sequencing data and the database of more then 30,000 3D in situ images is available as a systematic and comprehensive resource, the Dresden Ovary Table (DOT). The results of the screen define large groups of mRNAs that exhibit similar localization behavior in the oocyte, dramatically expanding the spectrum of known RNA localization events. Using our resource in conjunction with previous screens that analyzed tissue-specific and subcellular gene expression in *Drosophila* embryos (Lecuyer et al., 2007; Tomancak et al., 2007), we found that similarly to embryonic cells, ovarian cells also show prevalent subcellular mRNA localizations. The specific mRNA distributions however changed in different cell types and also within one cell over time. Our analysis further revealed that localized mRNAs differ on the level of gene organization, conservation and expression level from ubiquitous mRNAs. Finally, genetic evidence suggests that the posterior localized RNAs require common molecular machineries for their subcellular distribution.

### *mRNA localization in ovaries*

During oogenesis gene expression varies only minimally, while mRNA enrichment at subcellular sites is widespread and affects many genes that have not been previously studied. In contrast to the embryo system (Lecuyer et al., 2007), ovarian cells displayed more homogenous subcellular enrichments. With few exceptions, the candidate mRNAs co-localized with microtubule ends and their cytoplasmic localizations were strongly affected upon microtubule depolymerization. This suggests that long-range transport and enrichment at the microtubule ends is the primary mechanism for mRNA localization during oogenesis. Interestingly, the ovarian localized mRNAs themselves are enriched for genes with cytoskeletal functions and anterior mRNAs further associated with

terms assigning a cell cycle regulating function. Similarly, mRNAs localized in the embryo were implicated in cell division/cytoskeletal organization (Lecuyer et al., 2007). Thus mRNA localization must affect other oogenesis processes beyond axis-formation. A local source of cytoskeletal proteins in the early oocyte could be beneficial to allow the rapid re-organization and growth of the cytoskeleton at the transition from early to mid-oogenesis. Similarly, anterior localized mRNAs could be instrumental for regulated meiosis in the oocyte probably in combination with restricted translation (Benoit et al., 2008; Cui et al., 2013; Kronja et al., 2014; Tadros et al., 2007).

We also identified mRNAs enriched in the nuclei of the nurse cells. As these mRNAs are not particularly highly expressed, we do not interpret the signal as an artifact of detecting abundant nascent transcript in the nucleus. Rather these mRNAs may be stalled in their nuclear processing, possibly to prevent their premature release into the ooplasm. Several mRNAs were shown to enter the oocyte during late oogenesis, when nurse cell dumping occurs (reviewed in Mahajan-Miklos and Cooley, 1994). 29 RNAs were also detectable in the meiotic oocyte nucleus at stages 9 and 10 that could either be made in the nurse cells and imported into the oocyte nucleus or reflect instances of transcription from the meiotic nucleus (Saunders and Cohen, 1999).

What controls the specificity of mRNA localization in various cell types? Our data show that the localized mRNAs have a more complex gene structure than ubiquitous mRNAs and in particular longer non-coding features such as the 3'UTR. This was most obvious for the class of posterior mRNAs that are also more conserved compared to the anterior localized mRNAs. Longer 3'UTRs could harbor several motifs that in sum lead to robust subcellular mRNA enrichments across time. In addition, posterior genes show higher expression levels than anterior mRNAs. Considering how seemingly inefficient the active transport towards the posterior is (Zimyanin et al., 2008), higher expression levels could act as an additional measure to ensure that enough mRNAs will eventually localize. It is also possible that for some mRNAs, high expression level in combination with an anchoring motif or a microtubule zipcode signal could result in localization. Indeed, we observed two phases of posterior localization, at stage 9 and later at stage 10. The late class resembles *nanos* mRNA that localizes through a posterior trapping mechanisms (Sinsimer et al., 2011), which could very well benefit from high expression levels. It will be interesting to analyze these gene feature variables in more cell types, possibly of other species, to test which of them could be used to predict localization status of mRNAs. Our data also show that the canonical examples of mRNA localization in the ovary (Berleth et al., 1988; Ephrussi et al., 1991; Neuman-Silberberg and Schüpbach, 1993; St Johnston et al., 1989) represent a broad class of functionally related mRNAs.

***Dynamic mRNA localization***

The integration of genome-wide FISH datasets from different tissues revealed that the mRNA localization profile changes dramatically during development. This is in contrast to the observations that mRNAs localize through sequence encoded mRNA zipcodes (reviewed in Medioni et al., 2012) and that the localization machinery is active in all cell types analyzed (Bullock and Ish-Horowicz, 2001; Jambor et al., 2014). The divergence of mRNA localization across different tissues could be due to alternative splicing producing forms with or without a zipcode. During oogenesis however, we could not detect alternative transcript expression that could explain the difference of localization over time.

Several explanations for the observed

variability of localization status remain. Firstly, if all localized mRNAs have a specific zipcode, then this RNA motif must be inactivated, for example blocked by a protein or melted by an RNA helicase, when the localization status of the mRNA changes (Figure 7). Since the localization status of mRNAs with respect to microtubule ends is continuously changing during oogenesis and into early embryogenesis (Figure 6D and Figure S7C) the remodeling of such a zipcode signal would have to be under temporal- and tissue-specific control. Secondly, it is possible that components of the localization machinery become limiting during oogenesis over time. In this scenario, a set of mRNAs with strong affinity to the localization protein complex would be constitutively localized. It is interesting to note that posterior mRNAs displayed the highest level of mRNA expression, which under limiting conditions could be beneficial to maintain their localization. However, this model cannot explain why some mRNAs but not others switch from microtubule minus- to plus-end transport. A third possibility is, that only few mRNAs have specific zipcode signals and the vast majority of localization events occur zipcode independently (Figure 7). Our observation that all examined posterior mRNAs fail to localize in the absence of *oskar* mRNAs lends support to the idea that these mRNAs are hitchhiking the *oskar* localization machinery and might not themselves have a "posterior"-zipcode. mRNA localization then could be a combinatorial effect of RNA protein interactions, RNA-RNA binding motifs, anchoring and trapping elements. Such a combinatorial code for mRNA localization could also explain the longer and more complex gene features we observed for the subcellular gene set.

To what extent widespread and dynamic localization of mRNAs is required for cellular function will have to be determined by examining the translation status of these mRNAs. Furthermore, combining large datasets from several cell types uncovered that mRNA localization is a phenomenon contingent on the cellular context and is most likely highly regulated. It will therefore be interesting to ask whether specificity of mRNA localization is based on a selective cell type specific mRNA regulation machinery, a zipcode signal, whether localized mRNAs travel in groups or are trapped by as yet unidentified physical properties of subcellular cytoplasmic domains. Our dataset enables the transition from deep mechanistic dissection of singular RNA localization events towards systemic examination of how RNAs transcribed in the nucleus distribute in cells and how this affects cellular architecture and cell behavior in development.

## Experimental Procedures

### Mass isolation of *Drosophila* egg-chambers

Flies were grown under standard laboratory conditions, fed for 2 days with fresh yeast at 21 and 25°C. For isolation of egg-chambers we developed a mass-solation protocol (see Extended Experimental Procedures) that allows us to enrich separated egg-chambers of all stages.

### RNA isolation, sequencing and analysis

We isolated total mRNA using TRIreagent (Sigma Aldrich) from stage 1-7 egg-chambers, including the germline stem cells, from stage 8-10 egg-chambers and from total ovaries containing mainly stage 11 and older egg-chambers. Additionally, RNA from 0-2h embryos was isolated. We used two complementary mRNA sequencing approaches; standard whole mRNA sequencing (RNAseq) and a sequencing method that captures specifically the sequence adjacent to the poly(A) tail and thus allows direct counting of transcripts (3Pseq, Surendranath and Dahl, manuscript in preparation). Of the

~50 million (3Pseq) and 100 million (RNAseq) Illumina reads we mapped 70% (3Pseq) and 90% (RNAseq) to the *D.melanogaster* release 5.52 genome with Bowtie. Quantification was done using HTSeq (Anders and Huber, 2010). Normalization and differential expression was done using DESeq (Anders and Huber, 2010). Noise thresholds of 70 and 50 counts, for RNAseq and 3Pseq respectively, were derived from observing the distributions of normalized counts. 3' UTR forms were assigned by overlaying annotated Flybase UTR forms with 3P-Seq reads lying within 200 nucleotides of the annotated 3' UTR end. Alternate Polyadenylation events were called by calculating the mean weighted UTR length (Ulitsky et al., 2012), a difference of 200 nucleotides in the mean weighted lengths corresponding to 2 biological stages resulted in the gene being considered as undergoing Alternate Polyadenylation.

**96-well Fluorescent in situ Hybridization**

We used an established protocol for in situ hybridization in 96-well plates (Tomancak et al., 2007) with minor adaptations (see Extended Experimental Procedures): we added an over-night wash step after hybridization, incubate the anti-DIG antibody over night and used fluorescent tyramides for probe detection. Each experiment was evaluated and imaged using a wide-field microscope (Zeiss Axioplan Imaging) equipped with an optical sectioning device (DSD1, Andor) to generate confocal-like z-stacks.

**Annotation and Database**

We developed a controlled vocabulary to describe the cell types and relevant subcellular structures for oogenesis for germline and somatic cells[3]. Experiments showing no detectable FISH signal were classified as "no signal at all stages", while experiments resulting in a homogeneous signal throughout oogenesis were classified as "ubiquitous signal at all stages". Gene expression patterns were imaged up to stage 10B of oogenesis after which cuticle deposition prevents probe penetration. Each pattern that did not fall in the above-mentioned classes was imaged at all stages of oogenesis in several individual egg-chambers per time-point. We collected 3D images and used custom scripts in FIJI (Schindelin et al., 2012) to manually select and orient representative 2D images that were uploaded to the Dresden Ovarian-expression Table[4] (DOT). The 2D images remain linked to the original image stacks and all the raw stacks that were used to create an exemplary 2D image are available for interactive inspection using a simple image browsing cgi script. Thus the record of each in situ experiment for a given gene consists of a set of 2D images assigned to a specific oogenesis stage and described using annotation terms selected from the controlled vocabulary (Figure S1B). For definition of broad classifications, subclass grouping and embryo annotation class definition see Supplementary Table S1.

**Binary matrix**

To facilitate access to the multidimensional annotation data and to integrate them with the quantitative measures of gene expression from the RNAseq and 3Pseq experiments, we developed a single matrix where rows represent FlyBase ID and columns stage specific expression levels and all available annotation terms. This, so-called, binary matrix is the input for all downstream analysis and is provided as a flat file for independent bioinformatics investigation of the dataset[5] (see Extended

[3] http://tomancak-srv1.mpi-cbg.de/cgi-bin/ovary_annotation_hierarchy.pl

[4] http://tomancak-srv1.mpi-cbg.de/DOT/main

[5] http://tomancak-srv1.mpi-cbg.de/cgi-bin/dump_binary_matrix_ovary.pl?db=insitu_ovaries

Experimental Procedures).

**GO-term analysis**

For GO-term enrichment of gene sets we used the DAVID web server (Huang da et al., 2009). Terms or features enriched at a false discovery rate (FDR) of ≤10% and/or a Benjamini p-value of <0.1 were considered significant. Two stringencies were applied: the standard FDR cut-off (≤10%) or the more stringent 'Benjamini' p-value (≤0.1). For the statistical analysis of gene features see Extended Experimental Procedures.

**Colchicine treatment and Mutant analysis**

Flies were fed for 15 hours at 25C with fresh yeast paste supplemented with 50ug/ml colchicine (Cha et al., 2002). The effect of colchicine on individual egg-chambers was determined by scoring the detachment of the oocyte nucleus from the anterior cortex and its migration towards the center of the oocyte. To test posterior localization in mutants that affect *oskar* mRNA localization we used ovaries from homozygous $Spire^{RP}$ (Manseau and Schupbach, 1989), $Stau^{D3}$ (St Johnston et al., 1991) and $Btz^{1}$ (van Eeden et al., 2001) flies. Further we analyzed egg-chambers from *osk84/* Df(3R)$p^{XT103}$ flies lacking functional Oskar protein (Lehmann and Nusslein-Volhard, 1986) and from *oskar*3'UTR/+;*oskA87*/ Df(3R)$p^{XT103}$ flies that entirely lack endogenous *oskar* mRNA but develop past the early oogenesis arrest characteristic for *oskar* RNA null flies due to a transgenic source of oskar 3'UTR (Jenny et al., 2006) that is incapable posterior localization.

**ACCESSION NUMBER**

The SRA accession number is SRP045258.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes seven figures, eight tables and extended Experimental Procedures and can be found with this article online.

**Author contributions**

H.J. designed and performed the experiments, participated in the bioinformatics analysis, and wrote the manuscript. V.S. designed and performed most bioinformatic analyses including mapping, quantification of RNAseq and 3'seq reads, and statistical analysis of gene architecture features. A.T.K. performed statistical analysis of 3'UTR length, evolutionary conservation, and protein interactome data. P.M. performed the 96-well in situ hybridization. S.S. designed an image processing script to enhance analysis of fluorescent image data. P.T. co-designed the study, performed cross-tissue analysis, developed database and website, and co-wrote the manuscript.

**Acknowledgement/Authors**

We thank Diana Selig, Jens Schmiedel and David Seniuk for technical assistance with image acquisition and processing. Holger Brandl/Bioinformatic services. Franziska Friedrich for drawing of the ovariole. Anne Starkloff for webpage coding/design. Andreas Dahl, Deep Sequencing Group at CRTD/BIOTECH, Dresden for generation of the 3Pseq data. Anne Ephrussi and Daniel St Johnston for fly lines. Michael Hiller for sharing the 24 *Drosophila* species alignment. Florence Besse, Simon Bullock, Carsten Hoege, Simone Reber, James Saenz and Vitaly Zimyanin for discussion and comments on the manuscript. V.S. received support from DIGS-BB. H.J. and P.T. were supported by FP7-EU. Project: GENCODYS. P.T. was additionally funded by HFSP Young Investigator Grant RGY0093/2012 and by The European Research Council Community's Seventh Framework Program (FP7/2007-2013) grant agreement 260746.

**References**

Amrute-Nayak, M., and Bullock, S.L. (2012). Single-molecule assays reveal that RNA

localization signals regulate dynein-dynactin copy number on individual transcript cargoes. Nature cell biology *14*, 416-423.
Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.
Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrancois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P.*, et al.* (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251-262.
Benoit, P., Papin, C., Kwak, J.E., Wickens, M., and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in Drosophila. Development *135*, 1969-1979.
Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nusslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. Embo J *7*, 1749-1756.
Besse, F., Lopez de Quinto, S., Marchand, V., Trucco, A., and Ephrussi, A. (2009). Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. Genes & development *23*, 195-207.
Blower, M.D., Feric, E., Weis, K., and Heald, R. (2007). Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. The Journal of cell biology *179*, 1365-1373.
Bullock, S.L. (2011). Messengers, motors and mysteries: sorting of eukaryotic mRNAs by cytoskeletal transport. Biochemical Society transactions *39*, 1161-1165.
Bullock, S.L., and Ish-Horowicz, D. (2001). Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. Nature *414*, 611-616.
Bullock, S.L., Ringel, I., Ish-Horowicz, D., and Lukavsky, P.J. (2010). A'-form RNA helices are required for cytoplasmic mRNA transport in Drosophila. Nature structural & molecular biology *17*, 703-709.
Cajigas, I.J., Tushev, G., Will, T.J., tom Dieck, S., Fuerst, N., and Schuman, E.M. (2012). The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron *74*, 453-466.
Cha, B.J., Serbus, L.R., Koppetsch, B.S., and Theurkauf, W.E. (2002). Kinesin I-dependent cortical exclusion restricts pole plasm to the oocyte posterior. Nat Cell Biol *4*, 592-598.
Chao, J.A., Patskovsky, Y., Patel, V., Levy, M., Almo, S.C., and Singer, R.H. (2010). ZBP1 recognition of beta-actin zipcode induces RNA looping. Genes & development *24*, 148-158.
Chintapalli, V.R., Wang, J., and Dow, J.A. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet *39*, 715-720.
Cui, J., Sartain, C.V., Pleiss, J.A., and Wolfner, M.F. (2013). Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in Drosophila. Developmental biology *383*, 121-131.
Dienstbier, M., Boehl, F., Li, X., and Bullock, S.L. (2009). Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. Genes & development *23*, 1546-1558.
Dix, C.I., Soundararajan, H.C., Dzhindzhev, N.S., Begum, F., Suter, B., Ohkura, H., Stephens, E., and Bullock, S.L. (2013). Lissencephaly-1 promotes the recruitment of dynein and dynactin to transported mRNAs. The Journal of cell biology *202*, 479-494.
Ephrussi, A., Dickinson, L.K., and Lehmann, R. (1991). Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. Cell *66*, 37-50.
Forrest, K.M., and Gavis, E.R. (2003). Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for

nanos mRNA localization in Drosophila. Curr Biol *13*, 1159-1168.
Ghosh, S., Marchand, V., Gaspar, I., and Ephrussi, A. (2012). Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. Nature structural & molecular biology *19*, 441-449.
Glotzer, J.B., Saffrich, R., Glotzer, M., and Ephrussi, A. (1997). Cytoplasmic flows localize injected oskar RNA in Drosophila oocytes. Curr Biol *7*, 326-337.
Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W.*, et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473-479.
Hachet, O., and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. Nature *428*, 959-963.
Horne-Badovinac, S., and Bilder, D. (2008). Dynein regulates epithelial polarity and the apical localization of stardust A mRNA. PLoS genetics *4*, e8.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.
Jambhekar, A., and Derisi, J.L. (2007). Cis-acting determinants of asymmetric, cytoplasmic RNA transport. RNA *13*, 625-642.
Jambor, H., Brunel, C., and Ephrussi, A. (2011). Dimerization of oskar 3' UTRs promotes hitchhiking for RNA localization in the Drosophila oocyte. RNA *17*, 2049-2057.
Jambor, H., Mueller, S., Bullock, S.L., and Ephrussi, A. (2014). A stem-loop structure directs oskar mRNA to microtubule minus ends. RNA *20*, 429-439.
Jeffery, W.R., Tomlinson, C.R., and Brodeur, R.D. (1983). Localization of actin messenger RNA during early ascidian development. Dev Biol *99*, 408-417.
Jenny, A., Hachet, O., Zavorszky, P., Cyrklaff, A., Weston, M.D., Johnston, D.S., Erdelyi, M., and Ephrussi, A. (2006). A translation-independent role of oskar RNA in early Drosophila oogenesis. Development *133*, 2827-2833.
Kislauskis, E.H., Zhu, X., and Singer, R.H. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. J Cell Biol *127*, 441-451.
Kronja, I., Yuan, B., Eichhorn, S.W., Dzeyk, K., Krijgsveld, J., Bartel, D.P., and Orr-Weaver, T.L. (2014). Widespread Changes in the Posttranscriptional Landscape at the Drosophila Oocyte-to-Embryo Transition. Cell reports *7*, 1495-1508.
Lange, S., Katayama, Y., Schmid, M., Burkacky, O., Brauchle, C., Lamb, D.C., and Jansen, R.P. (2008). Simultaneous transport of different localized mRNA species revealed by live-cell imaging. Traffic *9*, 1256-1267.
Lawrence, J.B., and Singer, R.H. (1986). Intracellular localization of messenger RNAs for cytoskeletal proteins. Cell *45*, 407-415.
Lecuyer, E., and Tomancak, P. (2008). Mapping the gene expression universe. Curr Opin Genet Dev *18*, 506-512.
Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174-187.
Lehmann, R., and Nusslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. Cell *47*, 141-152.
Mahajan-Miklos, S., and Cooley, L. (1994). Intercellular cytoplasm transport during Drosophila oogenesis. Developmental biology *165*, 336-351.
Manseau, L.J., and Schupbach, T. (1989). cappuccino and spire: two unique maternal-effect loci required for both the anteroposterior and dorsoventral patterns

of the Drosophila embryo. Genes Dev *3*, 1437-1452.
Medioni, C., Mowry, K., and Besse, F. (2012). Principles and roles of mRNA localization in animal development. Development *139*, 3263-3276.
Meyer, R.E., Delaage, M., Rosset, R., Capri, M., and Ait-Ahmed, O. (2010). A single mutation results in diploid gamete formation and parthenogenesis in a Drosophila yemanuclein-alpha meiosis I defective mutant. BMC genetics *11*, 104.
Micklem, D.R., Adams, J., Grunert, S., and St Johnston, D. (2000). Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. Embo J *19*, 1366-1377.
Mikl, M., Vendra, G., and Kiebler, M.A. (2011). Independent localization of MAP2, CaMKIIalpha and beta-actin RNAs in low copy numbers. EMBO reports *12*, 1077-1084.
Mohler, J.D. (1977). Developmental genetics of the Drosophila egg. I. Identification of 59 sex-linked cistrons with maternal effects on embryonic development. Genetics *85*, 259-272.
Mukai, M., Kato, H., Hira, S., Nakamura, K., Kita, H., and Kobayashi, S. (2011). Innexin2 gap junctions in somatic support cells are required for cyst formation and for egg chamber formation in Drosophila. Mechanisms of development *128*, 510-523.
Neuman-Silberberg, F.S., and Schüpbach, T. (1993). The *Drosophila* dorsoventral patterning gene *gurken* produces a dorsally localized RNA and encodes a TGFa-like protein. Cell *75*, 165-174.
Nezis, I.P., Stravopodis, D.J., Papassideri, I., Robert-Nicoud, M., and Margaritis, L.H. (2002). Dynamics of apoptosis in the ovarian follicle cells during the late stages of Drosophila oogenesis. Cell and tissue research *307*, 401-409.
Pokrywka, N.J., and Stephenson, E.C. (1995). Microtubules are a general component of mRNA localization systems in Drosophila oocytes. Dev Biol *167*, 363-370.
Roth, S., Neuman-Silberberg, F.S., Barcelo, G., and Schupbach, T. (1995). cornichon and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in Drosophila. Cell *81*, 967-978.
Saunders, C., and Cohen, R.S. (1999). The role of oocyte transcription, the 5'UTR, translation repression and derepression in *Drosophila gurken* mRNA and protein localization. Mol Cell *3*, 43-54.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nature methods *9*, 676-682.
Serano, T.L., and Cohen, R.S. (1995). A small predicted stem-loop structure mediates oocyte localization of Drosophila K10 mRNA. Development *121*, 3809-3818.
Shepard, K.A., Gerber, A.P., Jambhekar, A., Takizawa, P.A., Brown, P.O., Herschlag, D., DeRisi, J.L., and Vale, R.D. (2003). Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. Proceedings of the National Academy of Sciences of the United States of America *100*, 11429-11434.
Sinsimer, K.S., Jain, R.A., Chatterjee, S., and Gavis, E.R. (2011). A late phase of germ plasm accumulation during Drosophila oogenesis requires lost and rumpelstiltskin. Development *138*, 3431-3440.
Snee, M.J., Arn, E.A., Bullock, S.L., and Macdonald, P.M. (2005). Recognition of the bcd mRNA localization signal in Drosophila embryos and ovaries. Molecular and cellular biology *25*, 1501-1510.
St Johnston, D., Beuchle, D., and Nusslein-Volhard, C. (1991). Staufen, a gene required to localize maternal RNAs in the Drosophila egg. Cell *66*, 51-63.
St Johnston, D., Driever, W., Berleth, T., Richstein, S., and Nusslein-Volhard, C. (1989). Multiple steps in the localization of bicoid RNA to the anterior pole of the

Drosophila oocyte. Development *107 Suppl*, 13-19.
Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Developmental cell *12*, 143-155.
Tanaka, T., and Nakamura, A. (2008). The endocytic pathway acts downstream of Oskar in Drosophila germ plasm assembly. Development *135*, 1107-1117.
Theurkauf, W.E. (1994). Premature microtubule-dependent cytoplasmic streaming in cappuccino and spire mutant oocytes. Science *265*, 2093-2096.
Theurkauf, W.E., Alberts, B.M., Jan, Y.N., and Jongens, T.A. (1993). A central role for microtubules in the differentiation of Drosophila oocytes. Development *118*, 1169-1180.
Theurkauf, W.E., Smiley, S., Wong, M.L., and Alberts, B.M. (1992). Reorganization of the cytoskeleton during *Drosophila* oogenesis: implications for axis specification and intercellular transport. Development *115*, 923-936.
Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E.*, et al.* (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol *3*, RESEARCH0088.
Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol *8*, R145.
Ulitsky, I., Shkumatava, A., Jan, C.H., Subtelny, A.O., Koppstein, D., Bell, G.W., Sive, H., and Bartel, D.P. (2012). Extensive alternative polyadenylation during zebrafish development. Genome research *22*, 2054-2066.
van Eeden, F.J., Palacios, I.M., Petronczki, M., Weston, M.J., and St Johnston, D. (2001). Barentsz is essential for the posterior localization of oskar mRNA and colocalizes with it to the posterior pole. J Cell Biol *154*, 511-524.
Vanzo, N., Oprins, A., Xanthakis, D., Ephrussi, A., and Rabouille, C. (2007). Stimulation of endocytosis and actin dynamics by Oskar polarizes the Drosophila oocyte. Dev Cell *12*, 543-555.
Vanzo, N.F., and Ephrussi, A. (2002). Oskar anchoring restricts pole plasm formation to the posterior of the Drosophila oocyte. Development *129*, 3705-3714.
Wang, C., Dickinson, L.K., and Lehmann, R. (1994). Genetics of nanos localization in Drosophila. Developmental dynamics : an official publication of the American Association of Anatomists *199*, 103-115.
Whittaker, K.L., Ding, D., Fisher, W.W., and Lipshitz, H.D. (1999). Different 3' untranslated regions target alternatively processed hu-li tai shao (hts) transcripts to distinct cytoplasmic locations during Drosophila oogenesis. Journal of cell science *112 ( Pt 19)*, 3385-3398.
Zimyanin, V., Lowe, N., and St Johnston, D. (2007). An oskar-dependent positive feedback loop maintains the polarity of the Drosophila oocyte. Current biology : CB *17*, 353-359.
Zimyanin, V.L., Belaya, K., Pecreaux, J., Gilchrist, M.J., Clark, A., Davis, I., and St Johnston, D. (2008). In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. Cell *134*, 843-853.
Zivraj, K.H., Tung, Y.C., Piper, M., Gumy, L., Fawcett, J.W., Yeo, G.S., and Holt, C.E. (2010). Subcellular profiling reveals distinct and developmentally regulated repertoire of growth cone mRNAs. The Journal of neuroscience : the official journal of the Society for Neuroscience *30*, 15464-15478.

## Figure legends

### Figure 1. Summary of the FISH (FISH) screen on ovaries.

A. Summary of key numbers of the screen. For each of the 6091 FISH experiment we annotated the signal as “no signal”, “ubiquitous” or “specific”. Only specific and some ubiquitous signals were imaged.

B. GO-term enrichment analysis of ubiquitous, cellular, nuclear and subcellular gene sets.

C. Summary of gene expression patterns in the embryo for groups of genes defined by common expression pattern in the ovary. The graphs represent the embryo annotations as a linear hierarchy (Lecuyer et al.). Each color-coded bar represents organ systems of the embryo from its stage specific anlagen to primordia to final differentiated structures. Embryogenesis stages are separated by vertical black lines. The width of the bar is proportional to the frequency with which this annotation term was used in the embryo dataset, the height corresponds to a z-score of over- (above axis) or under-representation (below axis) of the term in the set of genes defined by ovary annotation. Genes expressed ubiquitously in the ovary mostly remained ubiquitous in the embryo and were additionally specifically expressed in meso- and endoderm (circle); genes that exhibit cellular patterns in the ovary are enriched in ectoderm/epidermis cells of the late embryo (plus); subcellular genes were highly expressed in the ectoderm and nervous system (star) of the embryo.

D. The cellular gene set was subcategorized according to the specific cellular expression pattern. Individual mRNAs can fall into several of these subgroups.

E. Subcategories of subcellular localized mRNAs. One mRNA can appear in more than one subgroup.

### Figure 2. Patterns observed by FISH (FISH) in ovarian cells.

Panels 1-3 show exemplary FISH experiments for the cellular (2.1), subcellular (2.2) and nuclear (2.3) expression sets. RNA is shown in green and the DNA (labeled with DAPI) is shown in magenta. Scale bars: 30µm.

Panel 1: cellular expression patters.

*tutl* is expressed in cap cells at the tip of the germarium (A), while *Ect3* mRNA is detectable in the somatic epithelial cells of the germarium (B). Several mRNAs are expressed in mosaic pattern, indicating cell cycle control in somatic epithelial cells (C,D) and in nurse cells (D). Expression in the anterior and posterior follicle cells is often seen simultaneously (E, F, M). Some mRNAs were expressed only in anterior follicle cells that become migratory border cells (I) or in posterior follicle cells (K). *CG8303* (G) is expressed in the follicle cells that become columnar, and many mRNAs are seen expressed at later stages (O). *aop* is exclusively seen in follicle cells that will give rise to the squamous epithelium (H) and several mRNAs are specifically expressed here at later stages (J, L). mRNAs are also expressed in cells forming the border of columnar and squamous epithelial cells (N).

Panel 2: subcellular expression patterns.

In the syncytial early egg-chamber, 590 mRNAs are transported from the site of transcription in the nurse cells into the developing oocyte. Here mRNAs are either restricted to a cortical domain (A) or detectable in the entire ooplasm (B). Occasionally mRNAs were specifically enriched in the oocyte portion of the syncytial egg-chamber and simultaneously enriched at the

apical membrane of the somatic epithelial cells (C). Five mRNAs were specifically excluded from the oocyte portion and enriched in the nurse cells (D). Few mRNAs were enriched anterior in stage 2-7 oocytes (E). mRNAs showed ubiquitous granules in the cytoplasm (F) or rarely ubiquitous ring-like staining patterns, arrow and 10x10µm inset showing only the RNA channel (G). mRNAs also enriched around the nucleus of the oocyte (H) and the nurse cells nuclei (I) and this varied from an entire ring around the nucleus to specific sub-areas of the perinuclear space (J). Apical enrichment was detected in late epithelial somatic cells (K) while basal localization in the follicle cells was relatively rare (L). Anterior and posterior RNA localization varied between diffuse (M, O) and tight cortical enrichments (N, P).

Panel 3: nuclear expression patterns.

Nuclei enrichments of RNAs in nurse cells can vary from a ring-like expression (A), foci in a discrete area (B) to widespread foci (C) or nucleoplasm signal (D). RNAs are also detectable in epithelial cell nuclei (E) and for 28 RNAs also in the oocyte nucleus. Grey-scale image shows the respective RNA staining only in a zoomed-in view.

**Figure 3. The transcriptome shows little variation over the course of oogenesis.**

A. Results of transcriptome-wide sequencing from stage specific oogenesis samples (stage 1-7 = early, stage 9-10 = late, full ovaries) and 0-2 hour embryos. Across oogenesis and early embryogenesis, ~5500 genes (red) were detected by both mRNA sequencing (RNAseq) and 3'prime end sequencing (3Pseq) at each stage while additional 1-2000 mRNAs were only captures with either RNAseq (grey) or 3Pseq (black) technique. Across all time-points about half of the *D.melanogaster* genome was expressed.

B. More than 85% of the genes were expressed at each time point of oogenesis as shown by a Venn diagram overlapping the early, late and full ovarian transcriptome determined by 3Pseq (red) and RNAseq (turquois).

C. Pair-wise correlation of early/late and late/full ovary datasets revealed that the stage-specific transcriptomes were highly similar (Pearson Correlation: 0.79/0.77) and only few genes, highlighted in red, were significantly up- or down-regulated (p-value adjusted for multiple testing < 0.1). GO-term analysis of the genes identified as up (arrow up) and down (arrow down) regulated during oogenesis/early embryogenesis revealed that particularly genes encoding components of the extra-embryonic layers (vitelline membrane, ECM, cuticle) changed their expression levels.

D. Correlation analysis of expressed transcript isoform (deduced from RNAseq data) revealed that from early to late and from late to full ovaries almost no transcript-isoforms significantly changed in their expression level. Transcripts with significant changes are shown in red.

E. Only ~300 genes (early-full: 298; late-full: 308; full-embryo: 346) changed their mean-weighted 3'UTR length that is indicative of an alternative polyadenylation. Alternative UTR form usage was found for 1-4 anterior (red) and 4-5 posterior (blue) mRNAs during oogenesis.

**Figure 4. Subclasses of localized mRNAs and their specific features at the level of gene architecture, mRNA expression, evolutionary conservation and function.**

A. Boxplots showing distributions of median mRNA expression levels for gene sets defined by annotations (see supplementary table S1). Shown are 3Pseq quantifications from late ovary mRNA (for early, full ovaries and early embryogenesis see Figure S5A). mRNAs of the posterior

group showed significantly higher expression than anterior mRNAs (A', Kolmogorov-Smirnov p-value: 3.9e-06).

B. Distributions of 3'UTR length for gene sets. B'-B''. Results of a non-parametric randomization test to show that ubiquitous and subcellular genes (B'; p-value = 0) and anterior and posterior genes (B"; p-value = 0.0018) have significantly different median 3'UTR lengths (i.e. no or little overlap of densities).

C-D. The median gene (C) and intron (D) length is significantly longer in anterior and again in posterior genes compared with ubiquitous transcripts. The boxplots show gene and intron length in nucleotides for the gene sets. The corresponding significance level tests are shown in Figure S5B,C.

E. Median conservation of the 3'UTR sequence for gene sets across 24 *Drosophila* species. E'-E'. Result of a non-parametric randomization test showing that ubiquitous genes are significantly less conserved in their 3'UTRs than subcellular genes (p-value: 0) and posterior genes show higher conservation than anterior genes (p-value: 0.0032).

F. Protein interaction analysis per gene set revealed that posterior genes, but not anterior genes, share significantly more protein-protein interactions than would be expected by chance.

G. GO-terms associated with oocyte enriched, anterior and posterior gene sets. Shown are the p-values for each GO-term calculated by the modified Fisher Exact test, which results in the EASE score p-value.

**Figure 5. mRNA localization requires the microtubule cytoskeleton and posterior enrichment is impaired in posterior localization pathway mutants.**

A-I: FISH experiments showing the RNA in green and DNA (labeled with DAPI) in magenta. Scale bars 30µm.

A-D. Localization of exemplary anterior (A-B) and posterior mRNAs (C-D) is lost upon microtubule depolymerization by colchicine (A'-D'). See summary of the results in Figure S6F and accompanying supplementary table S8.

E-I. Localization of the novel posterior candidate mRNAs *vkg* (F)*, TwdlG* (G), *PI3K21B* (H) and *zpg* (I) is lost in egg-chambers that prematurely depolymerize the microtubules (flies homozygous for *Spire*$^{RP}$), are mutant for the RNA binding protein Staufen (flies homozygous for *Stau*$^{D3}$) or mutant for the EJC protein Barentz (flies homozygous for *Btz*$^{1}$). The candidate mRNAs are mis-localized in a manner similar to *oskar* mRNA, whose localization is known to be disrupted in those mutant conditions (E). In *Btz*$^{1}$ egg chambers a weak enrichment of *vkg* mRNA remained that in rare instances is also observed for *oskar* mRNA. The localization of the tested novel posterior mRNAs was also lost at stage 10 in egg-chambers mutant for Oskar protein (*osk84/Df(3R)p*$^{XT103}$). All candidate mRNAs were lost in egg-chambers that do not express posterior *oskar* mRNA. Egg-chambers used were from *oskar* RNA null flies that express the *oskar* 3'UTR in order to rescue the early oogenesis arrest (*oskar 3'UTR/+; oskA87/Df(3R)p*$^{XT103}$; Jenny et al., 2006).

**Figure 6. mRNA localizations are highly variable across tissues and within cells.**

A. Comparison of the results of FISH screens in ovaries and Drosophila embryos (Lecuyer et al., 2007).

B. Venn-diagram showing the lack of overlap among the mRNAs probed in both embryonic and

ovarian cells that show instances of subcellular localization. Only 5 (<1%) showed localization in all three tissues and only 89 (14%) showed localization in two tissues.

C. Venn-diagram showing the lack of overlap among the mRNAs co-localizing with microtubule minus ends. Only 3 mRNAs (<1%) were constitutively found at minus ends in all three cell types sampled and localization in two tissues was seen for 29 mRNAs (9%).

D-H. mRNA localizations within one cell type, the oocyte, over time. Dendrogram (D) where each line represents the localization of single mRNAs in the oocyte from early to late oogenesis. Black colored lines indicate microtubule minus-end localization, red colored lines indicate plus-end localization and a ubiquitous phase of mRNA distribution is shown as a grey line. mRNAs initially localized at minus-ends in early oogenesis were observed to remain localized at minus ends in mid oogenesis (E-E'), switched to a plus-end accumulation at stage 9 or 10 (F-F') or, in the majority of cases, became ubiquitously distributed (G-G'). Initially ubiquitously distributed mRNAs could also adopt a plus-end accumulation at both stages 9 and 10 (H-H'). E-H'. FISH showing the RNA in green and DNA (labeled with DAPI) in magenta. Scale bar 30µm.

**Figure 7. How do cells mass-transport mRNAs dynamically over time?**

It is possible that all mRNAs with localization potential contain a zip code like cis-regulatory RNA signal that is responsible for the transport. In this case, switching of an mRNA from a localized to an unlocalized state would have to be regulated by RNA binding proteins and further cis-regulatory signals.

Alternatively, only some localized mRNAs could harbor an zip code signal, while most localized mRNAs would use indirect transport pathways, such as RNA hitch-hiking, formation of higher order transport granules (transport RNPs) or trapping/anchoring mechanisms for temporally regulated subcellular enrichments.

**Figure S1:**

A. Overview of the experimental procedure for transcriptome and genome-wide in situ hybridization experiments and evaluation.

B. Screenshot of the publicly available Dresden ovary table, DOT, and several of the key search and download functions.

**Figure S2:**

A. Scatterplot of RNAseq and 3Pseq gene expression showing correlation (Pearson Correlation 0.71) between these two RNA sequencing methods.

B. GO-term enrichment analysis for genes significantly up- and down-regulated during oogenesis for early, late and full ovaries. Late ovaries down-regulated genes and full ovaries up-regulated genes were not analyzed since they contained too few entries.

C. Example of the germline expressed *nanos* mRNA that shows a change in gene expression from early to full ovaries measured by RNAseq (green) and 3Pseq (red). The bottom part shows the *nanos* gene model with the position of introns and exons.

D. Among the genes showing alternative isoform expression during oogenesis (see Figure 3D), few are found among subcellular enriched mRNAs, for example as oocyte enriched and posteriorly localized RNAs. No mRNA localized at the anterior pole exhibited alternative isoform expression.

E. Boxplot showing that the vast majority of genes express only one 3'UTR form during oogenesis, suggesting low prevalence of alternative poly-adenylation.

F. Number of transcripts per gene for the ubiquitous and subcellular gene set; Highlighted in red and blue are the anterior and posterior localized among the subcellular genes. The prevalence of alternatively spliced mRNAs is not changing between early, late and full ovary samples.

**Figure S3:**

A. Linear hierarchy (Tomancak et al. 2007) plot showing at which embryonic stage and in which tissue are oogenesis gene sets re-expressed during embryogenesis. The following oogenesis gene sets are shown: "nosignal", "nurse cells perinuclear", "oocyte-enriched", "oocyte anterior" and "oocyte posterior". Most "no signal" genes are also underrepresented in almost all stages and tissues of embryogenesis, apart from the PNS and ectodermal derivatives in the late stages of embryogenesis. Perinuclear enriched genes are highly expressed in meso- and endoderm tissues. Oocyte enriched, oocyte anterior and oocyte posterior genes are overall very similarly expressed during embryogenesis, being high in the polarized CNS and ectoderm tissues.

**Figure S4:**

A. Estimate of false- positive/negative rate of the in situ screen using comparison with the independent transcriptomics data. A gene was classified as falsely positive if it was annotated as ubiquitous or specific by FISH but was not detectable by either 3Pseq or RNAseq at any time-point of oogenesis. In 20% of the experiments we failed to detect in situ signal ("nosignal") although the transcript was detected at least at one time point by at least one deep sequencing method. These may represent false negative results, possibly due to non-functional RNA probes,

however we nevertheless included them in the downstream analysis in the no signal category (Figure S4A).

B. Karyogram showing the chromosomal position of genes for nuclear, anterior and posterior localization classes. Neither nuclear RNA genes, which often appear in foci like enrichments, nor anterior or posterior class genes are clustered on the chromosome.

C. Examples of FISH experiment detecting distributions of non-coding RNA (in green). While pri-*miRNA-318* is enriched in somatic epithelial cell nuclei, pri-*miRNA-303,* pri-*miRNA-31-b* and the long non-coding RNA *CR42862* are restricted to nuclei of the germline nurse cells. Scale bar 30µm, DNA in magenta.

D. *CG9609* and *Doa* mRNAs detected in the oocyte nucleus showing the enrichment over time at stages 9, 10A and 10B. At stage 9 only few small mRNA foci are visible, at stage 10 the mRNAs were enriched in proximity of the DNA in two large foci. mRNA enrichments in the somatic epithelial cells overlaying the oocyte (*CG14639*) and at the cortex of nurse cells (*Actn*). RNA signal shown in green. DNA, labeled with DAPI, is shown in magenta. Scale bar 30µm.

**Figure S5:**

A. Boxplots showing the median mRNA expression measured by 3Pseq per gene set in early and full ovaries and in 0-2h embryos. At the onset of embryogenesis, the cellular mRNAs were almost as low as the "no signal" class, confirming their predominant expression in somatic cells that at this time-point have undergone apoptosis. Accompanying the boxplot is the matrix of statistical significance tests (Kolmogorov-Smirnov) of the null hypothesis that the distributions of median expression values across the subcellular gene sets are the same. Statistically significant differences (p-value <0.01) are shown in blue, while gene sets that did not differ significantly are shown in grey (p-value >0.01).

B-C. Kolmogorov-Smirnov (KS) tests showing significantly different (blue, p-value 0-0.01) or non-different (grey, p-value >0.01) distributions of gene length (Figure 4C) and intron length (Figure 4D) across gene sets.

D-I. Boxplot and the corresponding Kolmogorov-Smirnov (KS) tests showing the median exon number (D), length (E) and proportion (F), intron number (G) and proportion (H) and 5'UTR length (I) and significance levels of pair-wise comparisons of the gene sets. Statistically significant differences (p-value <0.01) are shown in blue, while gene sets that did not differ significantly are shown in grey (p-value >0.01).

**Figure S6:**

A-E. Localization of the ubiquitous mRNA *msl-2* (A) is unchanged upon microtubule depolymerization by colchicine (B). Localization in proximity to the nucleus (C) is lost upon in colchicine treated egg-chambers (C'), RNAs localized partially nuclear and partially perinuclear (D) loose the cytoplasmic localization (D') while strictly nuclear RNAs (E) are unaffected by microtubule depolymerization (E').

F. Summary of anterior and posterior mRNA distributions (round aggregates, tiny aggregates, dispersed and diffuse aggregates) upon microtubule depolymerization. The diffuse aggregates occurred for mRNAs that in wild type egg-chambers showed a diffuse posterior enrichment (e.g. Figure 2.2.M).

G-H. *Bsg* and *CG7777* mRNA distribution is impaired in in egg-chambers lacking posterior *oskar*

mRNA.

A-E', G-H. FISH showing the RNA in in green; DNA (labeled with DAPI) is shown in magenta (A-B, G-H) or blue (C-E') and the nuclear membrane is stained with WGA dye shown in red (C-E'). Scale bar 30µm.

*CG11076* mRNA forms intranuclear foci and perinuclear enrichment, *Scp2* mRNA forms nuclei-associated foci;

**Figure S7:**

A. 1674 mRNAs show either during oogenesis or embryogenesis instances of subcellular localization (are "localization competent"). These mRNAs are highly enriched for cytoskeletal/microtubule and cell fate/development biological functions.

B. Venn diagram of the mRNAs showing nuclear enrichment in either oogenesis or embryogenesis. Only three mRNAs are nuclear at both developmental time-points.

C. Expanded dendrogram from Figure 6D including the data for the first two time-points of embryogenesis (Lecuyer et al., 2007), showing that both microtubule minus-end (anterior) and microtubule plus-end (posterior) localization decreases with the onset of embryogenesis. Increase in minus-end localization is again observed at stage 3-5 of embryogenesis, when zygotic transcription is activated.

Figure 1

A

| in situ hybridization screen overview* | | |
|---|---|---|
| experiments | 6091 | |
| genes annotated | 5862 | |
| imaged experiments | 1541 | |
| stacks | 31618 | |
| no signal | 2366 | |
| signal: | 3647 | |
| ubiquitous | 2357 | (65%) |
| specific | 1290 | (35%) |
| cellular pattern | 309 | (13%) |
| subcellular pattern | 790 | (22%) |
| nuclear pattern | 191 | (5%) |

* = freeze September 2013

B

GO-term [BP] enrichment of RNA classes

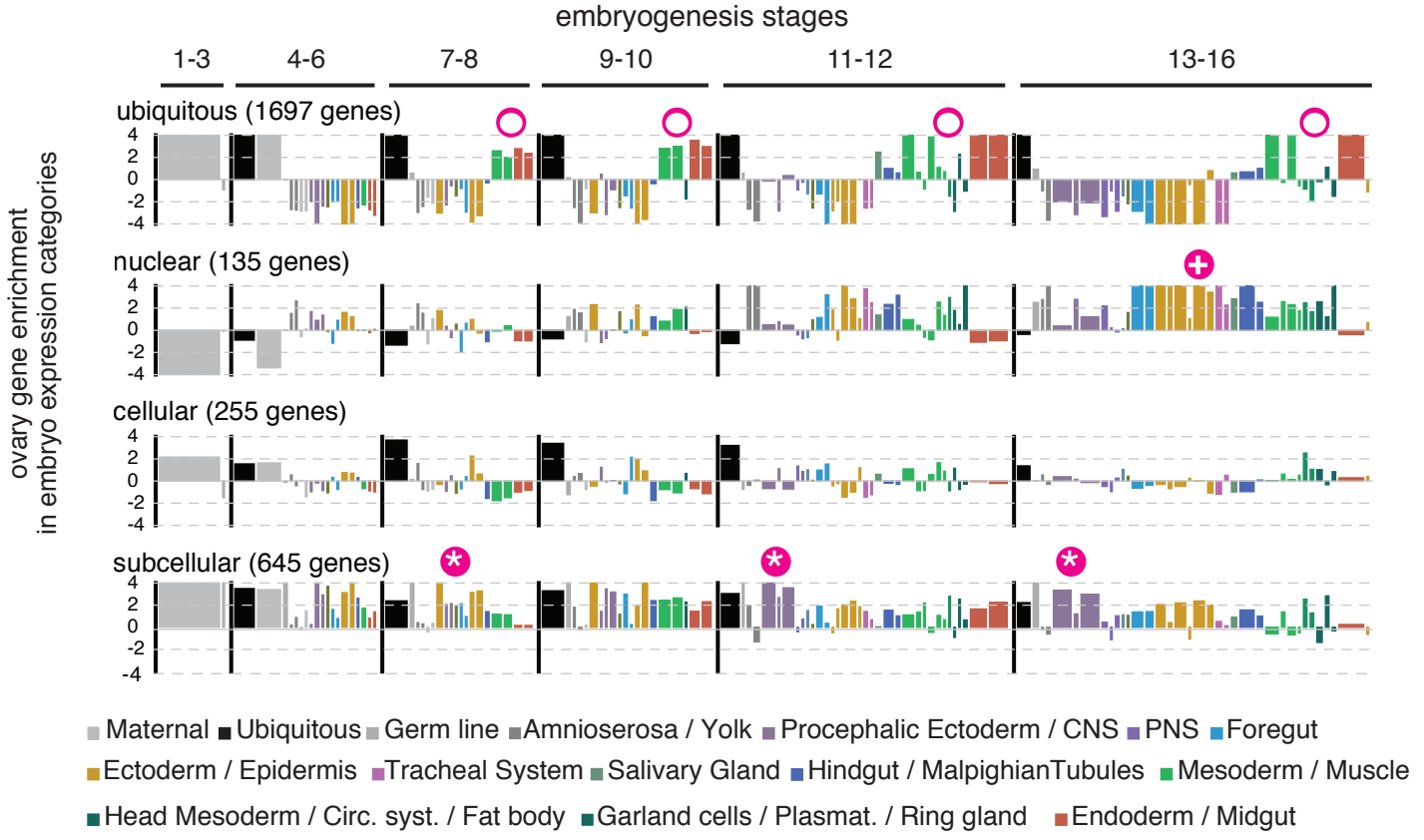
| ubiquitous genes | P-value |
|---|---|
| GO:0007052~mitotic spindle organization | 1.17E-12 |
| GO:0015031~protein transport/localization | 1.59E-12 |
| GO:0007010~cytoskeleton organization | 2.97E-11 |
| GO:0000226~microtubule cytoskeleton organization | 5.69E-11 |
| GO:0043933~macromolecular complex subunit organization | 2.17E-08 |
| GO:0030163~protein catabolic process | 2.17E-08 |
| GO:0006396~RNA processing | 1.61E-06 |
| GO:0006412~translation | 2.38E-06 |
| GO:0042254~ribosome biogenesis | 6.87E-06 |
| GO:0070271~protein complex biogenesis | 1.10E-05 |
| GO:0007049~cell cycle | 1.22E-05 |
| GO:0016192~vesicle-mediated transport | 2.00E-04 |

| subcellular genes | P-value |
|---|---|
| GO:0048610~reproductive process | 5.23E-10 |
| GO:0007010~cytoskeleton organization | 3.27E-09 |
| GO:0007017~microtubule-based process | 1.80E-08 |
| GO:0007389~pattern specification process | 2.81E-08 |
| GO:0007444~imaginal disc development | 3.00E-08 |
| GO:0051301~cell division | 4.76E-08 |
| GO:0019953~sexual reproduction | 6.24E-08 |

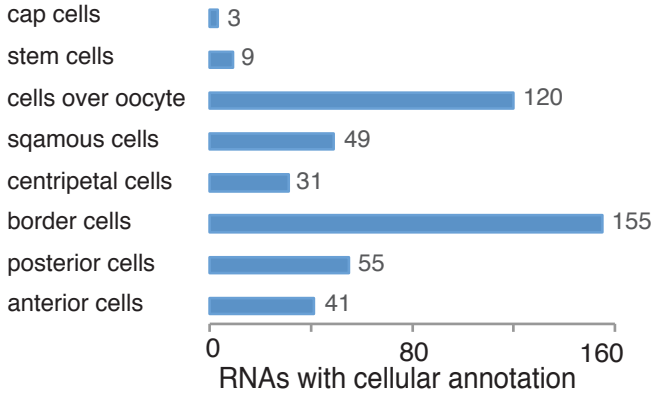
| cellular genes | P-value |
|---|---|
| GO:0009954~proximal/distal pattern formation | 1.43E-03 |
| GO:0008202~steroid metabolic process | 3.45E-03 |
| GO:0008610~lipid biosynthetic process | 4.51E-03 |
| GO:0035151~regulation of tube size | 8.97E-03 |
| GO:0007635~chemosensory behavior | 1.03E-02 |
| GO:0030030~cell projection organization | 1.40E-02 |
| GO:0060429~epithelium development | 1.41E-02 |

| nuclear genes | P-value |
|---|---|
| GO:0044265~cellular macromol. catabolic process | 3.64E-02 |
| GO:0000381~regulation alternative mRNA splicing | 4.42E-02 |
| GO:0050684~regulation of mRNA processing | 5.65E-02 |
| GO:0048024~regulation of nuclear mRNA splicing | 5.65E-02 |
| GO:0008202~steroid metabolic process | 6.06E-02 |
| GO:0043484~regulation of RNA splicing | 6.33E-02 |
| GO:0042026~protein refolding | 6.74E-02 |
| GO:0006468~protein amino acid phosphorylation | 6.78E-02 |
| GO:0009408~response to heat | 6.79E-02 |
| GO:0016310~phosphorylation | 6.96E-02 |

C




D




E

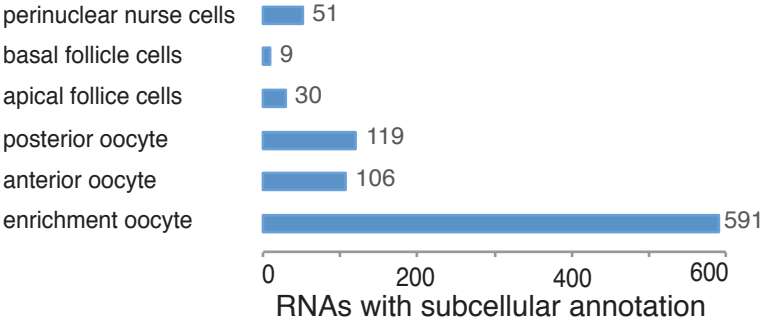
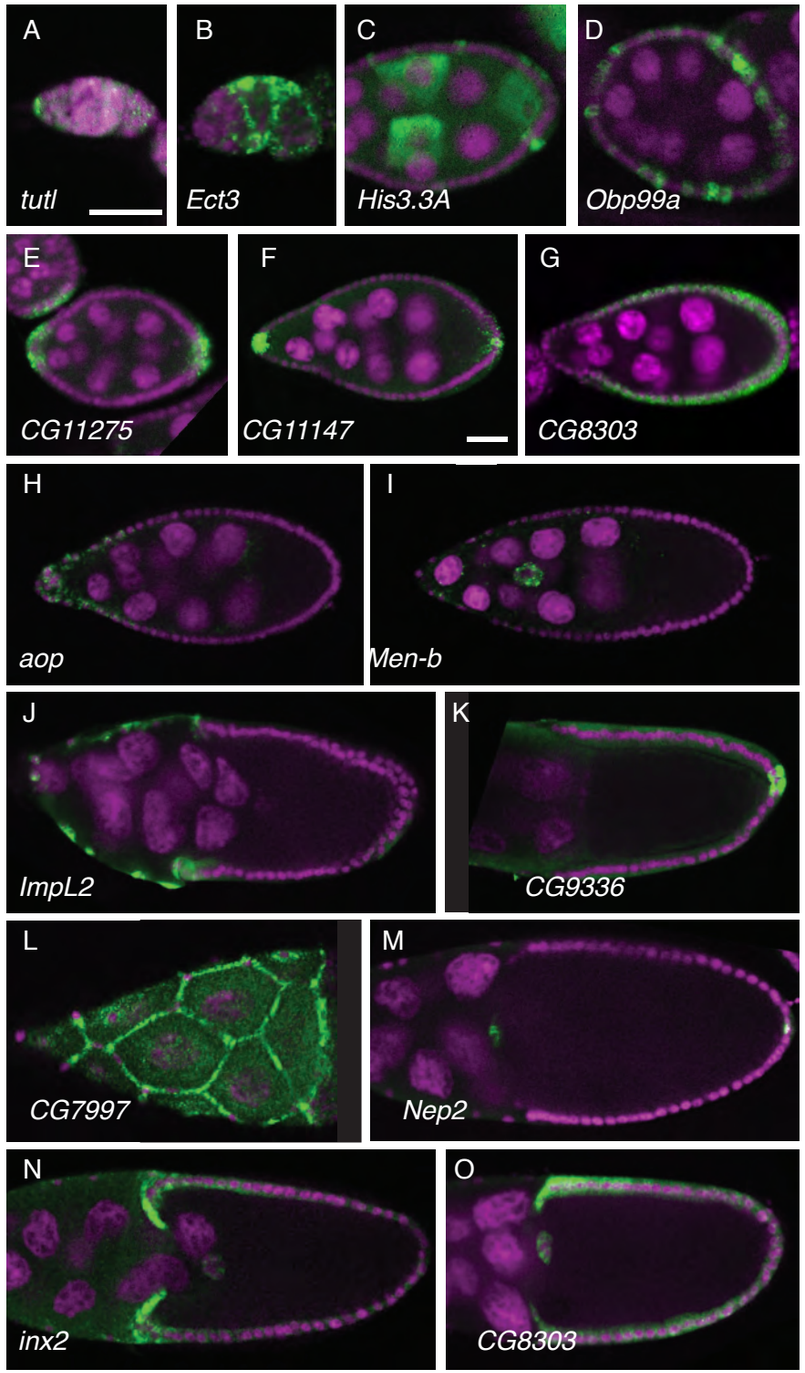
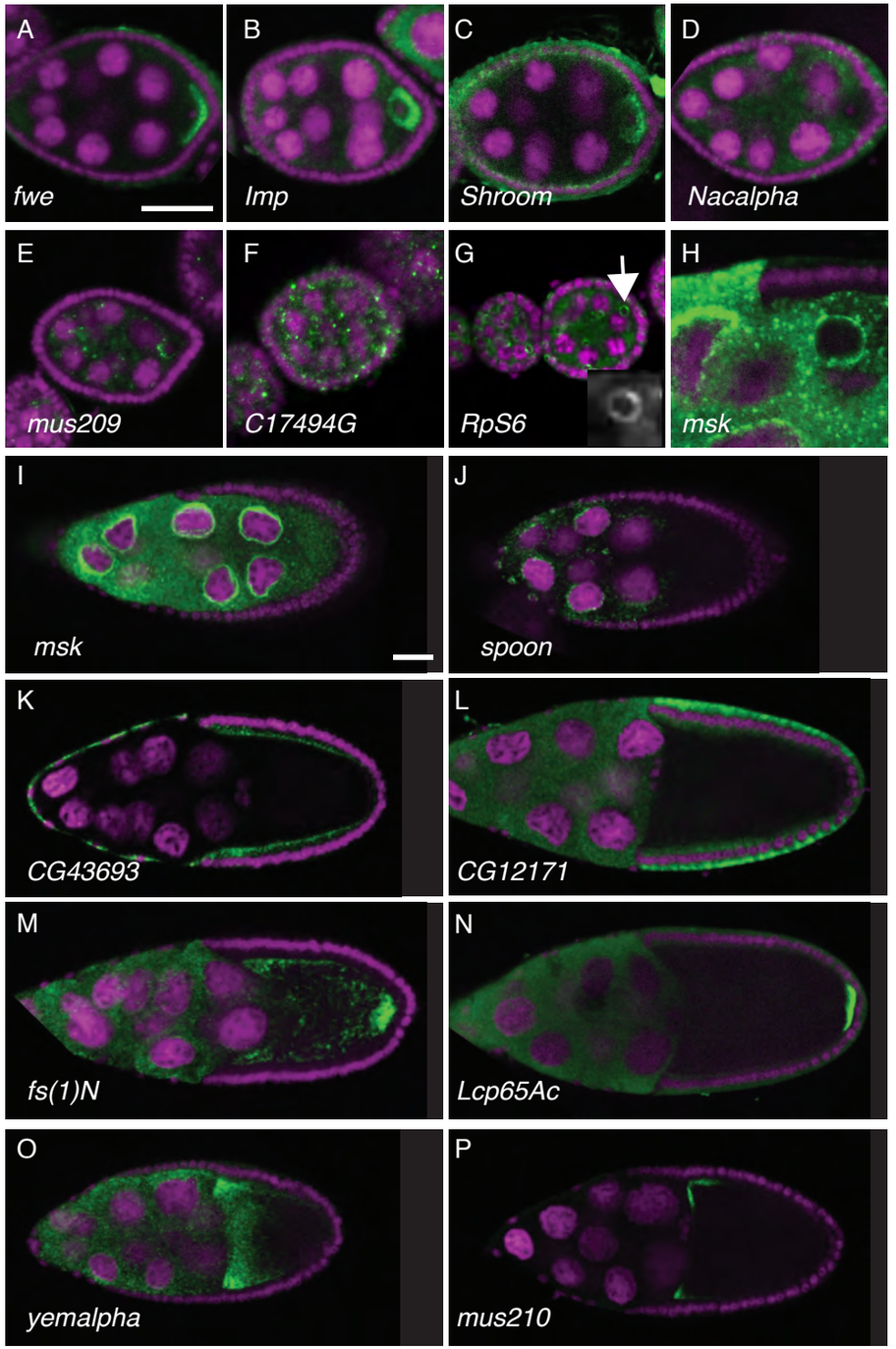


Figure 2

Panel 1: Cellular expression patterns




Panel 2: Subcellular expression patterns




Panel 3: Nuclear expression patterns

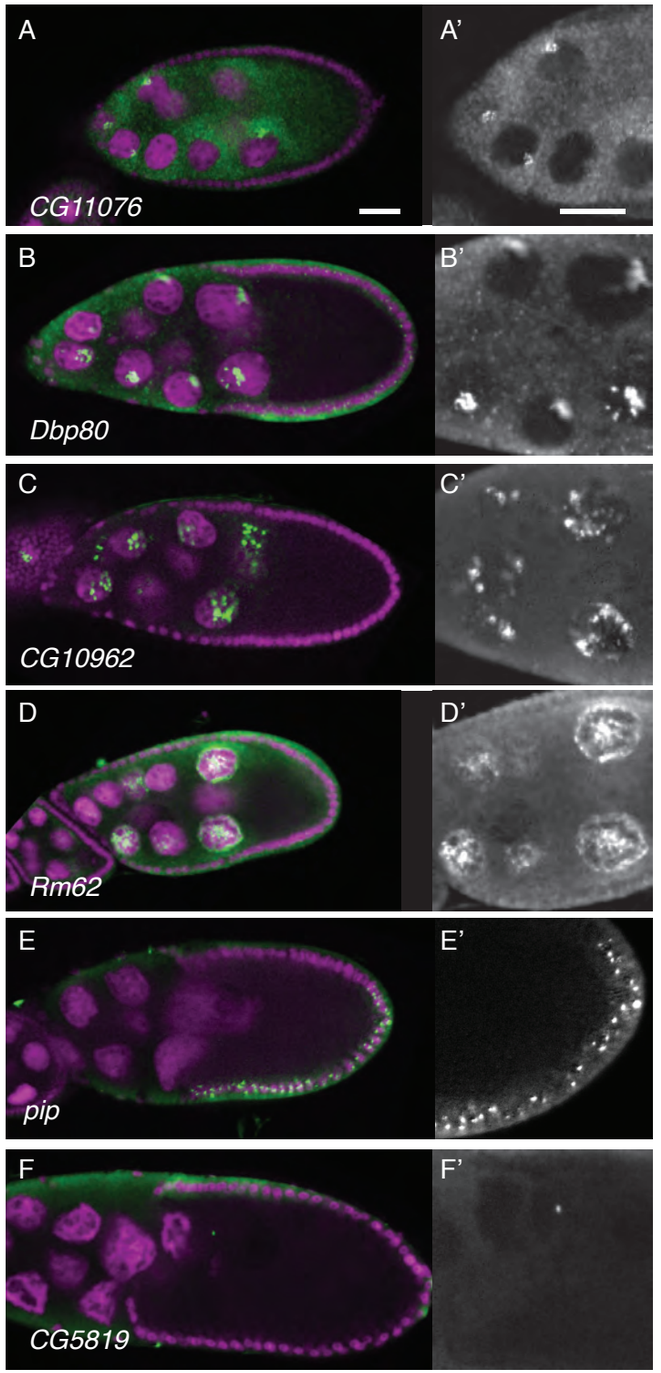


Figure 3

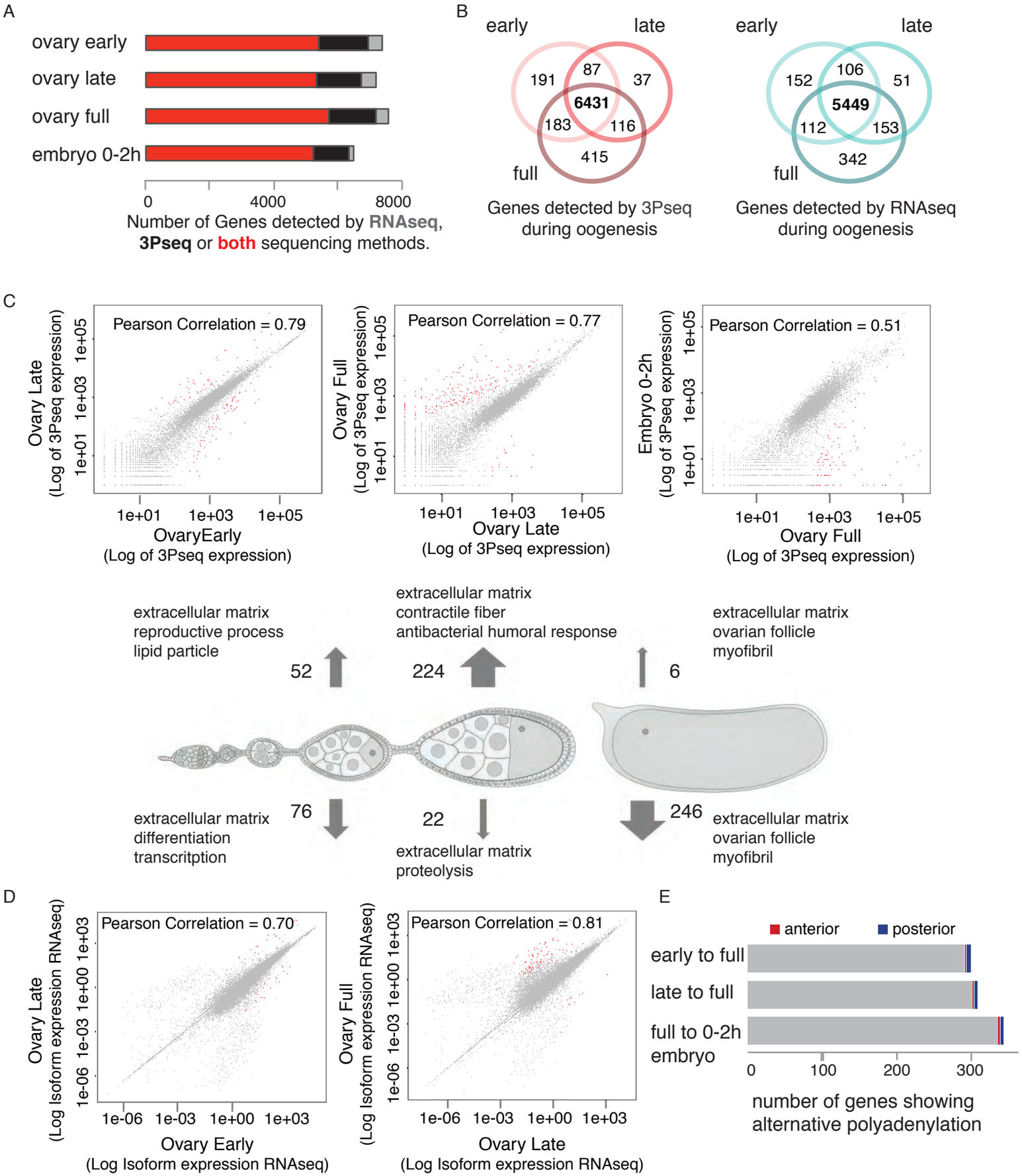
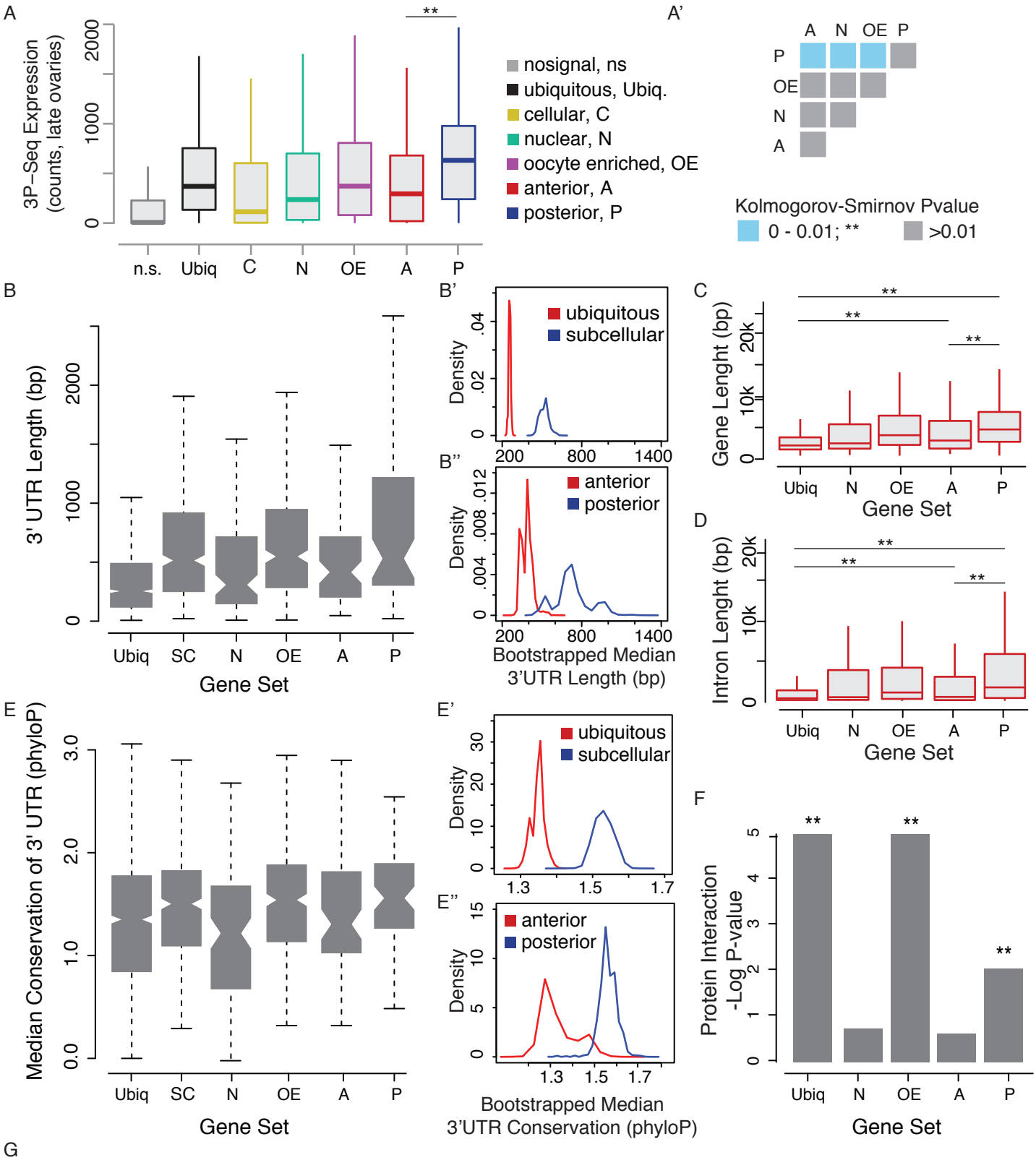

A
ovary early
ovary late
ovary full
embryo 0-2h
0
4000
8000
Number of Genes detected by RNAseq, 3Pseq or both sequencing methods.
B
early
late
full
191
87
37
6431
183
116
415
Genes detected by 3Pseq during oogenesis
early
late
full
152
106
51
5449
112
153
342
Genes detected by RNAseq during oogenesis
C
Pearson Correlation = 0.79
Ovary Late (Log of 3Pseq expression)
OvaryEarly (Log of 3Pseq expression)
Pearson Correlation = 0.77
Ovary Full (Log of 3Pseq expression)
Ovary Late (Log of 3Pseq expression)
Pearson Correlation = 0.51
Embryo 0-2h (Log of 3Pseq expression)
Ovary Full (Log of 3Pseq expression)
1e+01
1e+03
1e+05
extracellular matrix
reproductive process
lipid particle
52
extracellular matrix
contractile fiber
antibacterial humoral response
224
extracellular matrix
ovarian follicle
myofibril
6
76
extracellular matrix
differentiation
transcritption
22
extracellular matrix
proteolysis
246
extracellular matrix
ovarian follicle
myofibril
D
Pearson Correlation = 0.70
Ovary Late (Log Isoform expression RNAseq)
Ovary Early (Log Isoform expression RNAseq)
Pearson Correlation = 0.81
Ovary Full (Log Isoform expression RNAseq)
Ovary Late (Log Isoform expression RNAseq)
1e-06
1e-03
1e+00
1e+03
E
anterior
posterior
early to full
late to full
full to 0-2h embryo
0
100
200
300
number of genes showing alternative polyadenylation

Figure 4




G

| | oocyte enriched | P-value | anterior | P-value | posterior | P-value |
|---|---|---|---|---|---|---|
| GO-term BP | GO:0048610~reproductive cellular process | 3.36E-08 | GO:0030952~establ/maintenance of cytoskeleton polarity | 4.97E-05 | GO:0048610~reproductive cellular process | 9.46E-05 |
| | GO:0007389~pattern specification process | 6.54E-08 | GO:0000226~microtubule cytoskeleton organization | 2.42E-04 | GO:0009968~negative regulation of signal transduction | 2.62E-04 |
| | GO:0009968~negative regulation of signal transduction | 4.26E-07 | GO:0007281~germ cell development | 3.10E-04 | GO:0007276~gamete generation | 3.83E-04 |
| | GO:0007010~cytoskeleton organization | 4.56E-07 | GO:0009798~axis specification | 5.44E-04 | GO:0019953~sexual reproduction | 5.21E-04 |
| | GO:0007552~metamorphosis | 5.26E-07 | | | GO:0048477~oogenesis | 7.42E-04 |
| | GO:0003002~regionalization | 5.96E-07 | | | GO:0045165~cell fate commitment | 9.34E-04 |
| | GO:0007444~imaginal disc development | 7.52E-07 | | | | |
| GO-term CC | GO:0005856~cytoskeleton | 4.37E-06 | GO:0043232~intracellular non-membrane-bounded organelle | 3.72E-03 | GO:0005635~nuclear envelope | 2.64E-03 |
| | GO:0005938~cell cortex | 1.42E-05 | GO:0043228~non-membrane-bounded organelle | 3.72E-03 | GO:0012505~endomembrane system | 1.05E-02 |
| | GO:0044430~cytoskeletal part | 7.96E-05 | GO:0005856~cytoskeleton | 2.29E-02 | GO:0045495~pole plasm | 1.70E-02 |
| | | | GO:0044427~chromosomal part | 6.96E-02 | GO:0019908~nucl. cyclindep. protein kinase complex | 5.52E-02 |
| GO-term MF | GO:0003702~RNA polymerase II transcr. factor activity | 2.59E-07 | GO:0008017~microtubule binding | 1.04E-02 | GO:0032183~SUMO binding | 4.10E-03 |
| | GO:0030528~transcription regulator activity | 6.61E-06 | GO:0008092~cytoskeletal protein binding | 7.13E-02 | GO:0019207~kinase regulator activity | 1.79E-02 |
| | GO:0003677~DNA binding | 2.33E-05 | GO:0005158~insulin receptor binding | 7.64E-02 | GO:0003756~protein disulfide isomerase activity | 9.66E-02 |
| | GO:0008092~cytoskeletal protein binding | 2.35E-05 | GO:0060589~nucleoside-triphosphatase regulator activity | 8.62E-02 | | |
| | GO:0003700~transcription factor activity | 1.12E-04 | | | | |
| | GO:0004674~protein serine/threonine kinase activity | 4.48E-04 | | | | |
| | GO:0003779~actin binding | 4.87E-04 | | | | |

Figure 5

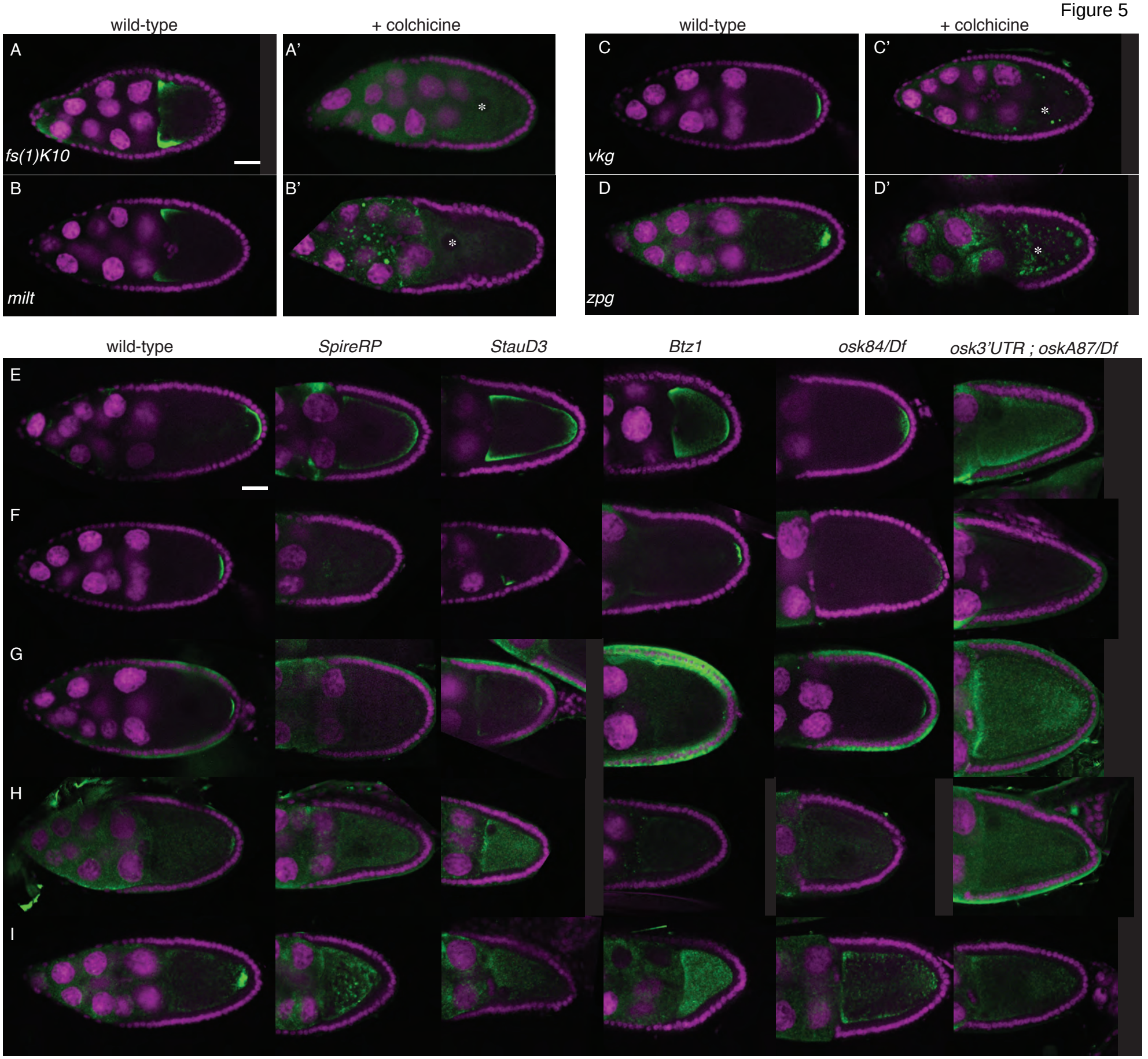

wild-type
+ colchicine
wild-type
+ colchicine
A
A'
C
C'
fs(1)K10
vkg
B
B'
D
D'
milt
zpg
wild-type
SpireRP
StauD3
Btz1
osk84/Df
osk3'UTR ; oskA87/Df
osk
E
vkg
F
TwdlG
G
PI3K21B
H
zpg
I

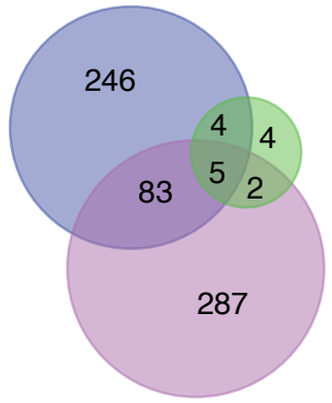
Figure 6

A

| | screened | expressed | localized |
|---|---|---|---|
| ovary | 5858 | 3621 | gc 669<br>fc 39 |
| embryo | 7199 | 3951 | 1081 |
| **intersect** | **3943** | **1720** | **5** |
| union | 9114 | 5852 | 1674 |

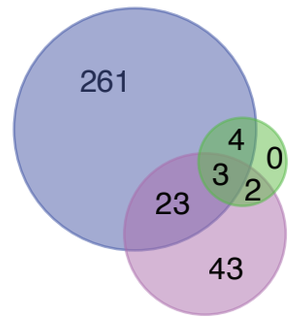
| probed in both datasets: | subcellular | minus-ends |
|---|---|---|
| gc | 338 | 291 |
| fc | 15 | 9 |
| embryo | 377 | 71 |

B




subcellular
RNA localization

C




microtuble minus-end
RNA localization

**cell type**

Ovary: germline | Ovary: somatic cell | Embryonic cells

D

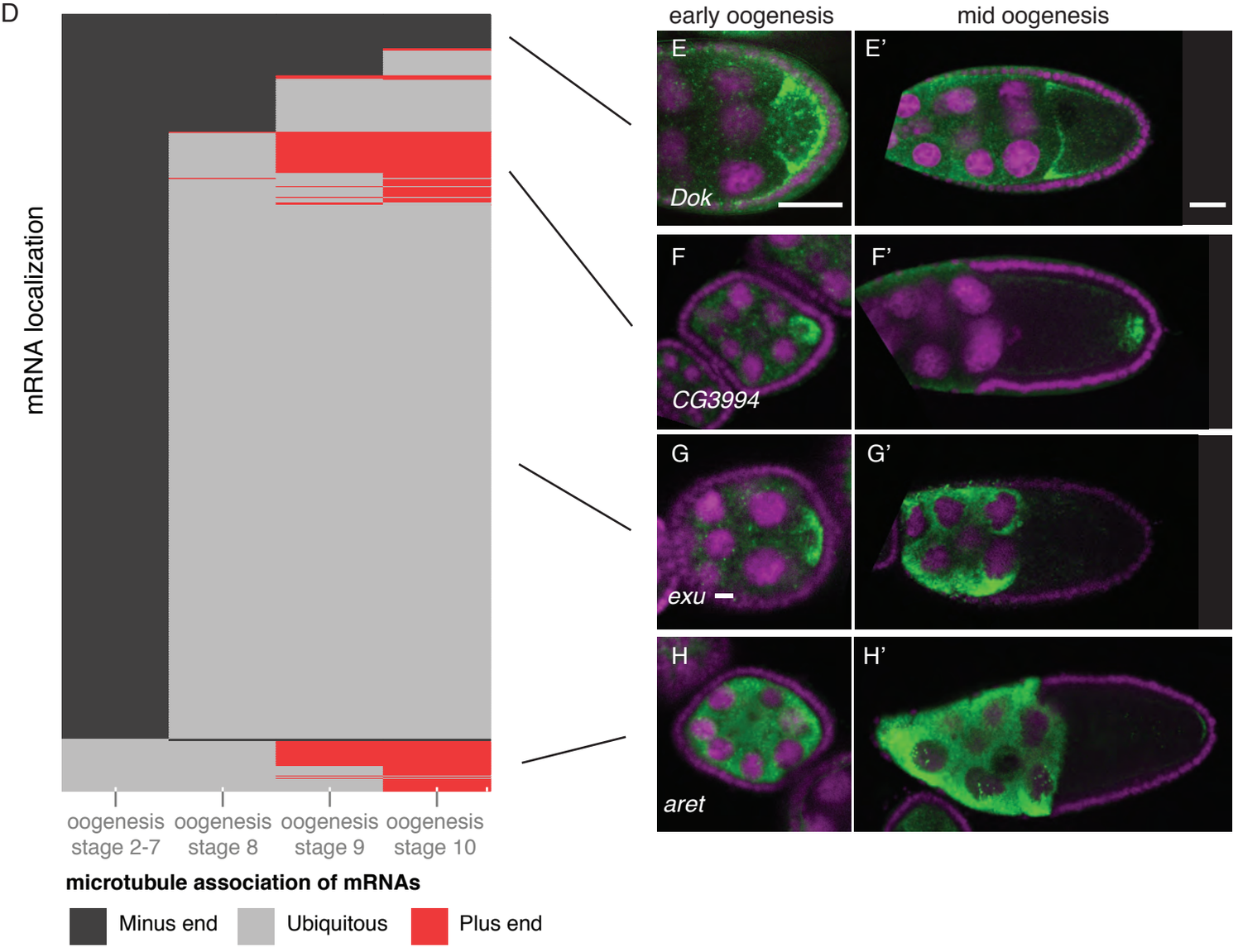
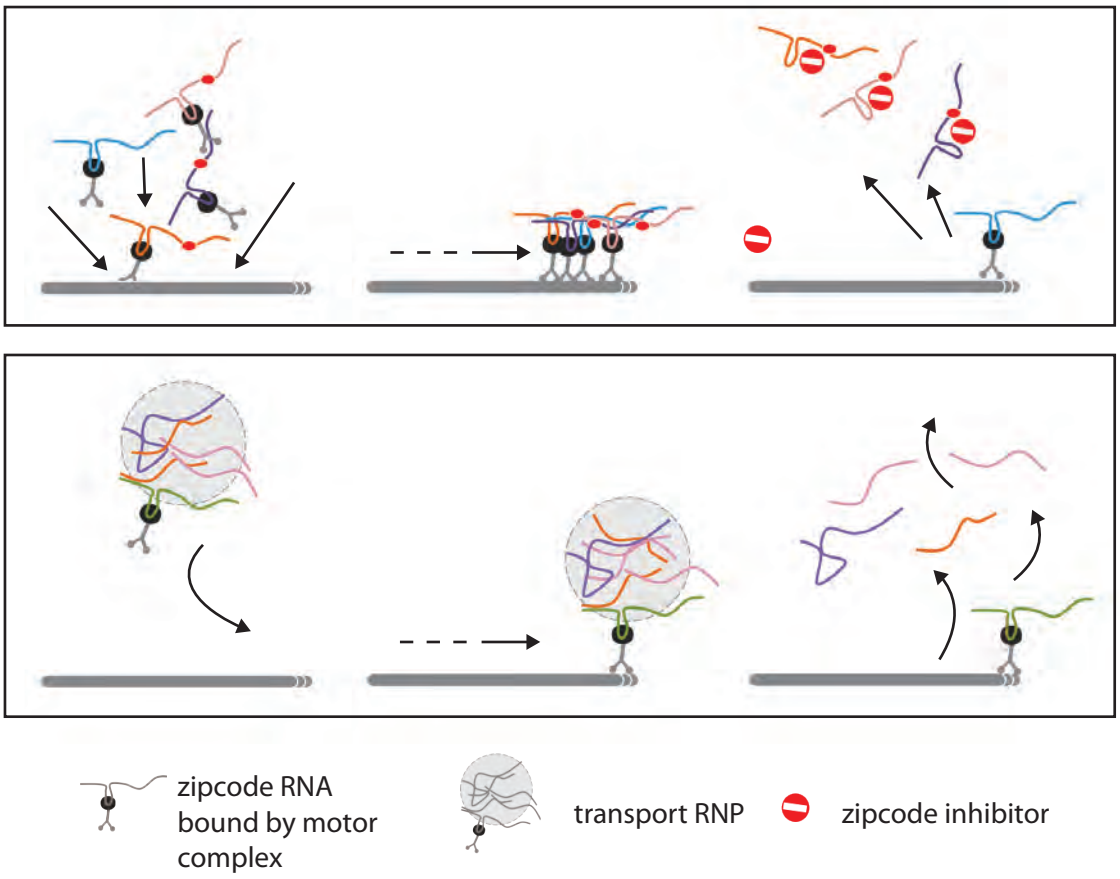


Figure 7


zipcode RNA bound by motor complex
transport RNP
zipcode inhibitor