Specialisation of ribosomes in gonads through paralog-switching

Tayah Hopes[1,2], Michaela Agapiou[1,2], Karl Norris[1,2,3], Charley G.P. McCarthy[4], Mary J O'Connell[4], Juan Fontana[1,3 *], Julie L Aspden[1,2 *]

1 School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, LS2 9JT, UK.

2 LeedsOmics, University of Leeds, UK

3 Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

4 School of Life Sciences, Faculty of Medicine and Health Sciences, The University of Nottingham, Nottingham, NG7 2UH, UK.

* For correspondence: j.aspden@leeds.ac.uk; j.fontana@leeds.ac.uk

Abstract

Ribosomes have long been thought of as homogeneous, macromolecular machines but recent evidence suggests they are heterogeneous and their specialisation can regulate translation. Here, we have characterised ribosomal protein heterogeneity across 5 tissues of *Drosophila melanogaster*. We find that testis and ovary contain the most heterogeneous ribosome populations, and that specialisation in these tissues occurs through paralog-switching. For the first time, we have solved structures of ribosomes purified from *in vivo* tissues by cryo-EM, revealing differences in precise ribosomal arrangement for testis and ovary 80S ribosomes. Differences in the amino acid composition of paralog pairs and their localisation on the ribosome exterior indicate paralog-switching could alter the ribosome surface, enabling different proteins to regulate translation. One testis-specific paralog-switching pair is also found in humans, suggesting this is a conserved site of ribosome specialisation. Overall, this work allows us to propose possible mechanisms by which ribosome specialisation can regulate translation.

Introduction

Protein synthesis is essential across the tree of life and undertaken by the highly conserved macromolecular complex of "the ribosome". mRNA translation is regulated at many levels, but until recently the ribosome itself was not thought to be part of this control system. Recent studies have suggested that ribosomes can contribute to gene expression regulation, through specific changes in their composition, i.e. specialisation [1-3]. These specialised ribosomes are thought to contribute to the translation of specific mRNA pools; but the mechanism by which this takes place is yet to be understood.

Previous analysis in a variety of organisms (mouse [1], yeast [4], and humans [5]) has shown that the composition of ribosomes is not homogeneous. In fact, specialisation of ribosomes is thought to be able to occur through a) additional protein components [6], b) substitution of ribosomal protein (RP) paralogs [7], c) post-translational modification of RPs [8], and d) rRNA modifications [9]. All these changes to the composition of ribosomes potentially contribute to ribosome specialisation.

Two significant factors have contributed to the logic behind the idea of 'specialised ribosomes'; a) prevalence of tissue specific RP expression and b) distinctive phenotypes when RP genes are disrupted [10]. Many RPs exhibit differences in expression levels across various tissues in mammals [1, 5, 11], plants [12], and insects [7]. For example, RpS5A and RpS5B are expressed in different cell types during early *Arabidopsis thaliana* development [13]. Disrupted RP genes result in varied, distinctive phenotypes suggesting that not all components are equally important all the time. For example, RpL38 mouse mutants exhibit a homeotic transformation phenotype with few other effects [1], whilst RpL38 mutants in *D. melanogaster* exhibit large wings, small bristles, delayed development and disorganised wing hair polarity [14].

Human cytoplasmic ribosomes usually comprise of 80 RPs and 4 rRNAs. This is similar across the majority of multicellular eukaryotes including *D. melanogaster* with 80 RPs and 5 rRNAs. However, annotated in FlyBase there are 93 cytoplasmic RP genes, including 39 small subunit proteins and 54 large subunit proteins [15]. These additional genes code for 13 paralogs in *D. melanogaster*. In fact, across eukaryotes many RP genes possess paralogs, for example human RpL3 and RpL3L [11] and *Arabidopsis* RpS8A and RpS8B [13]. In total, there are 19 pairs of paralogs in humans [4] and all 80 RPs in *Arabidopsis thaliana* have paralogs [16].

To dissect the function of ribosome heterogeneity it is necessary to understand biological importance within context of whole organisms. Within the developmental biology field, a large proportion of research focuses on the contribution of transcription to gene expression control. However, during development a variety of processes and key time points are highly dependent on the regulation of mRNA translation (oogenesis in *Xenopus* [17], early embryo development in *Drosophila* [18] and mammalian erythropoiesis [19]). The balance between self-renewal and differentiation at the stem cell niche is highly dependent on translation in both the ovary and the testis [20]. This is exemplified by disruptions to the stem cell niche in the testis when RPs are knocked down e.g. RpL19 RNAi results in over-proliferation of early germ cells in *D. melanogaster* [21]. During the meiotic phase of gametogenesis, transcription does not occur [22]; therefore meiotic cells rely on post-transcriptional gene regulation [23]. The translational machinery has evolved to become specialised within the testis with various testis specific components e.g. eIF4E-3 in *D. melanogaster* [24]. Many of the RP mutants associated with the *Minute* phenotypes have impaired fertility in both males and females [25, 26]. Moreover, mutations in 64 RPs in *D. melanogaster* result in *Minute* phenotypes of some sort [27].

Several human diseases have been attributed to mutations in RP genes. These diseases are called ribosomopathies, and they result from impaired translation and/or extra-ribosomal RP functions. The clinical symptoms vary between different RPs, suggesting human RPs also possess specialised functions, likely with respect to their contribution to the translation of specific mRNA pools. For example, mutations in RpS19 result in Diamond-Blackfan anaemia (DBA): a condition that presents with pure red cell aplasia [28].

Here we hypothesise that specialised ribosomes exist in the *D. melanogaster* testis to provide an additional level of mRNA translational regulation during spermatogenesis. Thus, we set out to determine potential changes to the ribosome and its function by probing the protein composition in 3 tissues (head, testis and ovary), during development in the embryo and embryo derived tissue culture S2 cells in *D. melanogaster*. Using quantitative mass spectrometry we identified heterogeneous ribosome populations, especially in the gonads. The main source of this variation in ribosome composition is paralog-switching, occurring in up to 50% of ribosomes in the testis and the ovary for specific paralogs. We found little difference in composition between single 80S ribosomes and, the more translationally active, polysome ribosomes from the same tissue, apart from in the ovary, where 2 paralogs are more abundant in 80S ribosomes. We solved structures of different ribosome populations to understand the potential mechanistic impact of these paralog-switching events. The resultant structures suggest potential mechanisms of translational regulation by different paralogs within the ribosome. To understand the broader importance of specialisation through paralog switching events we analysed the levels of conservation between paralog pairs. RpL22 has a duplicate RpL22L in mammals (including humans), and RpL22-like in *Drosophila,* these duplication events have occurred independently suggesting that it may represent a common mechanism of specialisation across a range of organisms and ribosomes.

Results

Heterogeneous ribosome populations exist in different tissues

Many eukaryotic genomes contain numerous RP paralogs and their contribution to ribosomal function is poorly understood. In *D. melanogaster* there are 93 RP genes (FlyBase), which includes 13 pairs of paralogs, normally resulting in 80 proteins in each ribosome [29]. The expression of RPs and

specifically RP paralogs has been reported to vary in a tissue specific manner. To profile potential differences in expression in *D. melanogaster* we analysed publicly available RNA-Seq data across various developmental time points and tissues. Hierarchical clustering of RP mRNA abundances across these different biological samples reveals variations in expression of RP mRNAs between tissues, with a cluster of RPs with much higher expression in the testis compared to other tissues (Fig 1A). This includes RpL22-like, a paralog of RpL22 previously reported as a testis specific ribosomal protein [7]. These results suggest the presence of testis-specific translational machinery.

To determine whether these different RPs are translated and incorporated into ribosomes we assessed the protein composition of ribosomes from these same tissues and cells; testes, ovaries, heads (mixture of male and female), embryos (0-2hr) and S2 cells (derived from embryo). Ribosomal complexes were purified using sucrose gradients and ultracentrifugation (Fig 1B). Both 80S and polysome complexes were isolated. The relative amounts of ribosomes existing as 80S or polysome complexes varied substantially across the samples (Sup 1A-E). Both monosome (80S) and polysome fractions were isolated for each tissue/cell type in two independent experiments before being subjected to quantitative mass spectrometry (Tandem Mass Tag; TMT). Overall correlation between the two biological replicates is high as the global protein content in testis 80S samples had a Pearson's correlation coefficient of 0.93 (Fig 1C). Similar results are obtained when considering only ribosomal proteins (Sup 1F & G) and across samples (Sup 1H-K).

To understand differences in ribosome composition between the tissues, protein abundances of ribosomal proteins were subject to hierarchical clustering (Fig 1D). A cluster of proteins emerged, which were enriched in the testis 80S ribosomes compared to 80S ribosomes from other tissues. This cluster included RpL22-like, RpL37b, RpS19b, RpS10a and RpS28a, RpS15Ab. There was also an ovary 80S enriched cluster of ribosomal proteins, RpL24-like, RpL7-like and RpL0-like (Fig 1D). PCA of protein abundances by ribosomal protein revealed that the majority of RPs (75/93) form a group together, suggesting they are incorporated in all ribosomes. The expression of RpL22-like, RpL37b, RpS19b, RpS10a and RpS28a clusters together, as their incorporation pattern across the different tissues is similar and this is driven mainly by their differential presence in testis 80S (Fig 1E, inset). The same can also be seen in the ovary enriched proteins (Fig 1E, inset).

When ribosomal protein abundances are plotted between different 80S complexes, we report that the largest differences are from paralogs rather than canonical RPs (Fig 1F & G). Comparison of testis 80S and head 80S shows 6 paralogs (RpL22-like, RpL37b, RpS19b, RpS10a, RpS28a and RpS15Ab) are highly enriched in the testis 80S compared to head (Fig 1F), whilst RpS11 is enriched in the head 80S. Comparison of testis 80S and ovary 80S reveals that whilst the majority of ribosomal proteins correlate between the two gonads, the same paralogs enriched in testis 80S compared to head were also enriched compared to ovary (Fig 1G). RpL24-like, RpL7-like, RpL0-like and RpS5b are all far more abundant in ovary 80S ribosomes than in the testis (Fig 1G). Overall, specialisation seems most common in the gonads and we identify both testis- and ovary-enriched paralogs.

Ribosomal protein paralogs contribute to ribosome heterogeneity

There are 13 pairs of RP paralogs in the *D. melanogaster* genome and from our TMT data we can see the majority are both expressed and incorporated into 80S ribosome in at least one of the analysed tissues or the developmental time point of the embryo. Hierarchical clustering of these paralogs re-emphasises the existence of gonad specific ribosomal complexes (Fig 2A). To understand the relationship between each of the two paralogs we used the mass spectrometry data to quantify relative abundances of the two paralogs with the matched pairs within the various tissues. Interestingly, for the majority of these proteins one of the paralogs seems to be dominant in terms of its presence in 80S ribosomes (Fig 2B). Strikingly the testis differs in composition the most when compared to the other samples (Fig 2A & B). In total, we find ~60% of testis 80S ribosomes contain RpL22-like rather than RpL22. These patterns were seen with both TMT experiments (Sup 2A). For 5 paralog pairs the second paralog is most abundant in the testis, and low in other samples (RpL22-like, RpL37b, RpS19b, RpS10a, RpS28a), we term these 'testis-enriched paralogs'. A similar situation

is seen for 4 paralog pairs where the second paralog is most abundant in the ovary (RpL24-like, RpL7-like, RpL0-like and RpS5b), 'ovary-enriched paralogs'. Interestingly RpS5b is present in ~50% of ovary 80S ribosomes, 45% of embryo 80S ribosomes and 30% of testis 80S ribosomes. Thus, RpS5b has an unusually broad incorporation across the different sampled ribosomes.

Differences in ribosome composition are mainly the result of selective protein incorporation

To understand the expression of RP paralogs, we analysed mRNA-Seq levels of each of the paralog pairs (Sup 2B). When relative paralog pair expression is profiled as a percentage on the basis of RNA-Seq, it is clear that differences in protein composition of ribosomes is not purely driven by transcriptional control of paralog genes (Sup 2B). We directly compared RP RNA expression (RNA-Seq) and RP protein incorporation (ribosome-TMT) identifying when RP incorporation into the ribosome does not correlate with mRNA expression level (Fig 2C). Specifically, RpL24-like is transcribed across all tissues at substantial levels (Sup 2B) and there is no difference in mRNA level between ovary and testis. However, RpL24-like is far more abundant in ovary 80S than testis 80S (Fig 2C) and is only represented in ovary 80S ribosomes (Fig 2B). RpL34a is very lowly incorporated into all ribosomes (Fig 2B) but its mRNA is expressed across tissues at substantial levels (Sup 2B). The opposite is true for RpS15Ab, whose RNA levels are similarly low between testis and ovary but is preferentially incorporated into testis 80S (Fig 2C). RpL7-like is expressed at the RNA level broadly in substantial amounts (Sup 2B) but is only incorporated into ribosomes at very low levels compared to RpL7 (<10%). Of note, the differential incorporation of RpL22-like into testes ribosomes compared to ovaries is driven by a transcriptional difference between the two tissues (Fig 2C, Sup 2B).

Composition of 80S ribosomes and polysomal ribosomes is similar

There is conflicting evidence as to the functionality or translational activity of monosomes (80S ribosomes), some suggest that these ribosomes are actively translating [30] whilst others suggest that not all 80S ribosomes are engaged in active translation [31]. To determine if there was any difference in ribosome composition between monosomes and polysome complexes, we compared the two by TMT. In general, there is very little difference in RP composition between 80S ribosomes and polysomes, e.g. testis (Fig 3A), head (Sup 3A). However, there are two paralogs enriched in the ovary 80S compared to the ovary polysome, RpL7-like and RpL24-like (Fig 3B). Such a large enrichment of these two paralogs in 80S complexes suggests that they potentially represent ribosome complexes whose activity is being regulated. Therefore, these 80S complexes may not be as translationally active as the polysome complexes. When the composition of ovary and testis polysomes are compared we identify 6 testis-enriched RPs, which are all paralogs; RpL22-like, RpL37b, RpS19b, RpS10a, RpS28a and RpL15Ab (Fig 3C). In fact, these are the same proteins enriched in testis 80S compared to ovary 80S (Fig 1G). In this comparison we also identify a group of proteins slightly enriched in the ovary polysomes; RpL37a, RpL22, RpS5b, RpL0-like and RpL40 (Fig 3C). Compared to the testis paralogs this fits well with the paralog switching between RpL37a/b, RpS5a/b and RpL22/RpL22-like. When the relative composition of polysomes for paralog pairs was determined the overall pattern was similar to 80S (Sup 3B). Differential incorporation within paralog pairs (Fig 3D) highlights the main differences between 80S and polysomes are associated with ovaries, and are RpL24/24-like, RpL7/7-like.

Cryo-electron microscopy of testis and ovary ribosomes reveals a mechanism for inactivation of testis 80S ribosomes

To understand the molecular implications of the paralog switching events we identified by mass spectrometry, we sought to solve structures of different ribosome populations. Ribosomal complexes were isolated in the same way as was previously described for TMT by sucrose gradient centrifugation (Fig 1B), with an additional step to concentrate purified samples (see Methods). Imaging the sample by cryo-electron microscopy (cryo-EM) confirmed that the ribosome complexes were highly pure and concentrated (Sup 4A). Testis 80S ribosomes were applied to grids and a

dataset containing ~47,000 particles was collected. Three-dimensional classification of this testis 80S dataset identified a single structurally distinct class of 80S ribosomes, which was refined to an average at 3.5 Å resolution (Fig 4A and Sup 4B). This provided a substantial improvement to the only other *D. melanogaster* ribosome cryo-EM average at 6 Å resolution, from embryos [29]. We performed a similar experiment with ovary 80S ribosome preparations, collecting a dataset containing ~200,000 particles, and resulting in an average at 3.0 Å resolution (Fig 4B; Sup 4C & D). These averages allowed us to generate atomic models for testis and ovary 80S ribosomal complexes (Sup Table 1).

Comparison of the testis and ovary averages revealed that the main difference between them was at the P/E tRNA site (Fig 4A and B). While the ovary 80S average did not contain any densities in this region, the testis 80S average contained densities that did not correspond to a tRNA (Fig 4A, circle). As a comparison, the previously published *D. melanogaster* average contained densities for an E-tRNA and for elongation factor 2, both of which are not present in our averages. By combining information from the testis 80S structure and the corresponding TMT data, we identified this density to be CG31694-PA (Fig 4C), which is highly abundant in the testis 80S complexes (10,451 normalised abundance, see Methods; 54th most abundant protein in testis 80S). CG31694-PA is an ortholog of IFRD2, identified in translationally inactive rabbit ribosomes as being bound to P/E sites of ~20% 80S isolated from rabbit reticulocytes [32]. Strikingly, in the reticulocytes the presence of IFRD2 is always accompanied by a tRNA in a noncanonical position (termed Z site), in the testis 80S average no tRNA was found in this region. In mammals IFRD2 is thought to have a role in translational regulation during differentiation. Differentiation is a key process during spermatogenesis within the testis, and in this context it is unsurprising to have found this protein in the testis 80S. CG31694-PA has considerable amino acid sequence conservation with IFRD2, 32% identity (Sup 4E & F). The presence of CG31694-PA suggests that a significant proportion of the testis 80S ribosomes is in fact not actively engaged in translation. CG31694-PA density was not present in the ovary 80S structure suggesting far fewer ribosomes are inactive by this mechanism in the ovary (5,105 normalised abundance in ovary 80S TMT compared with 10,451 in testis 80S). The presence of CG31694-PA does not affect the paralog switching events because these events were identical between the testis 80S and testis polysome ribosomes. To verify this, we solved the structure of ribosomes isolated from testis polysomes (cryo-EM average resolution was 4.9 Å) (Fig 4D and Sup 4G-I). It is clear from the density map that CG31694-PA is not present in the P/E sites; rather there is density for the E-tRNA in these actively translating ribosomes (Fig 4C & D). The TMT data indicates that levels of CG31694-PA are higher in the testis 80S than the testis polysomal complexes (10,451 normalised abundance in 80S compared to 6,144 in polysomes, see methods).

Functional implications of paralog switching event in gonads

By mapping the paralog switching events onto our ribosome structures we identified three clusters of paralogs undergoing switching. 1) Paralogs within the small subunit, including RpS19a/b and RpS5a/b, map to the head of the 40S near the mRNA channel (Fig 5A & B). 2) Paralogs within the large subunit tend be surface-exposed. Specifically, RpL22/RpL22-like and RpL24/RpL24-like locate towards the back of the ribosome (Fig 5C & D). 3) Paralogs that are located in ribosome stalks, RpLP0 and RpL10A, potentially interacting with the mRNA during translation (Fig 5E). Of note, the small subunit paralogs are close to the mRNA channel, pointing towards functional differences in mRNA selectivity of the ribosome.

By comparing the atomic models for testis 80S and ovary 80S, we identified differences between switched paralogs (Table 1). Specifically, the three paralogs with the greatest proportion (RpL22-like, 60% abundant in testis 80S; RpS19b, close to 50% abundant in testis 80S; and RpS5b, over 50% abundant in ovary 80S; Fig 2B) showed the largest differences in their atomic models (Fig 6A-F). Additionally, of the paralogs that do not switch between testis 80S and ovary 80S, RpS28b showed the largest differences (Fig 6G & H). This is probably due to its proximity to CG31694-PA (Fig 6I).

Comparing the amino acid sequences of each paralog pair it is possible to predict that they might contribute different functionality to the ribosome (Table 2, Sup 6A-J and Sup 7A-H). RpL22 and RpL22-like are only 45% identical, even though they are very similar in length (Fig 7A, Sup 7A). Unfortunately, the most different region between RpL22 and RpL22-like (i.e., the N-terminal region; Fig 7A), faces the exterior of the ribosome and is not resolved in the cryo-EM density (Fig 7A shows in bold the regions of RpL22 and RpL22-like present in the ovary 80S and testis 80S reconstructions, respectively). It is possible to imagine that given the majority of these paralogs are localised to the exterior of the ribosome, by switching one for the other might provide a difference exterior surface with which other associated factors might bind and change.

Conservation of paralog switching and implications for human disease

To probe how widespread paralog switching events might be to facilitate ribosome specialisation we determined the level of conservation of RpL22 and RpL22-like in other animal genomes. Orthologs of RpL22 were identified across a range of animals including Drosophilids. We determined that the paralogous pair RpL22 and RpL22-like present in *D. melanogaster* evolved by 3 independent duplication events across the animal clade (Fig 7B). A duplication event unique to the drosophila clade produced the paralogous pair RpL22 and RpL22-like that are identifiable in 6 out of the 12 Drosophila species sampled. The additional 2 duplication events present in the vertebrate clade may be the result of whole genome duplication rather than individual gene duplication events. The first of these vertebrate duplications produced the paralog pair RpL22 and RpL22L we observe in humans for example. The second vertebrate RPL22 duplication specific event occurred amongst teleost fishes and the most parsimonious explanation of pattern of distribution of duplicate copies would suggest subsequent lost in some lineages (Fig 7B). Thus, RPL22 has undergone multiple independent duplication events, generating a complex array of paralogous pairs.

Discussion

We have characterised the heterogeneity of ribosome composition across *Drosophila melanogaster* tissues and the developmental time point of embryos. For the first time we have identified differences in 80S ribosome composition purified from *in vivo* tissues. The main source of heterogeneity we discovered were paralog-switching events in the gonads. We have identified five testis-specific paralogs (RpL22-like, RpL37b, RpS19b, RpS10a, Rp28a) and four ovary-enriched paralogs (RpL24-like, RpL7-like, RpL0-like and RpS5b), which includes paralog, RpS5b, which is also to a lesser extent present in embryo and testis. There are very few differences between the composition of 80S and polysome ribosomes across all tissues. The exception to this is an enrichment of RpL24-like and RpL7-like in ovary 80S ribosomes compared to polysome ribosomes. These results are, in general, not just the consequence of transcriptional regulation of these paralogous genes. Rather there is modulation at the level of the translation of these proteins or incorporation into the ribosome. Regulation of the composition of these gonad ribosomes suggests the generation of specialised ribosomes for specific functions.

For the first time we have purified ribosomes from complex *in vivo* tissues. We have solved the cryo-EM structures of three different ribosome complexes; 80S ribosomes from the testis (3.5 Å), 80S ribosomes from the ovary (3.0 Å) and polysomal ribosomes from the testis (4.9 Å), improving the resolution from the only other previous ribosome structure from *D. melanogaster* [29]. One key difference was the testis 80S structure contains the *Drosophila* ortholog of IFRD2. Its presence indicates there is functional homology between CG31694 and IFRD2 in inhibiting mRNA translation through the ribosome, during differentiation. In mammals IFRD2 was seen in differentiating reticulocytes [32], whilst in our work we found CG31694 in the testis 80S (but not in the ovary 80S) where it could be involved in regulation of translation during the differentiation of spermatocytes, which is central to the function of the testis.

The paralogs we find switching in the gonads are localised in three clusters; a) the head of the 40S near the mRNA channel, b) the surface-exposed back of the large subunit and c) ribosome

stalks, potentially interacting with the mRNA during translation. The position of these three clusters provides potential explanations of how specialisation is achieved, mechanistically. Differences in amino acid sequence and precise position of the testis and ovary specialised paralogs (Fig 6C-F) can potentially affect the interaction of the mRNA and the ribosome, specifically during initiation when 40S ribosomes are recruited to the 5' end of mRNAs. The back of the 60S where RpL22 and RpL22-like are located, would provide an ideal site for additional protein factors to differentially bind to ribosomes containing these proteins. This is particularly true for this paralog pair, which has the lowest sequence identity between each other, 45%. The termini of these proteins are likely to be dynamic given the lack of density for them in our structures. Our phylogenomic analysis suggests that the modulation of this part of the exterior ribosome surface is in common across many organisms, and that the generation of paralogs has occurred independently three times for RpL22. Therefore, this potential mechanism might regulate the ribosome across many eukaryotes. Although paralogs are not conserved across a range of organisms, and many are limited to Drosophilids, there are many organisms with many RP paralog pairs, including human (19 pairs) and *Arabidopsis* (80 pairs). Therefore, these potential mechanisms of ribosome regulation could be conserved, if not the precise details.

The result we find here, that the gonads are important sites of ribosome heterogeneity and specialisation, further indicates how important mRNA translational regulation is in the testis and ovary. Many other testis-specific translation components exist to enable tight regulation such as eIF4-3 [24] and it is now clear that RP paralog switching also plays a part in this regulation.

The importance of the paralog-switching event between RpS5a and Rp5b has recently been functionally characterised in the *Drosophila* ovary [33]. Females without RpS5b produce ovaries with developmental and fertility defects, whilst those without RpS5a have no defects. RpS5b specifically binds to mRNAs encoding proteins with functions enriched for mitochondrial and metabolic GO terms in the ovary, suggesting ovary RpS5b containing ribosomes translate this specific pool of mRNAs [33]. It will be interesting to see how widespread this finding is for RpS5b, since this is a frequently switched paralog: we find that 50% of ovary 80S ribosomes contain RpS5b, whilst 45% of embryo 80S and 30% of testis 80S also contain RpS5b. It has been known for some time that mutations in RpS5a produce a *Minute* phenotype (including infertility), so it seems likely that these two paralogs both have biologically important roles in the fly. RpS5a and RpS5b have also been seen to exhibit tissue-specific expression in *A. thaliana*, in a developmentally regulated manner[13]. atRpS5a was suggested to be more important than atRpS5b during differentiation, because of its expression pattern, but the regulation mechanism remains elusive in *A. thaliana*.

The function of the RpL22 and RpL22-like paralog pair in *Drosophila* testis has been explored and it has been suggested that the two proteins are not functionally redundant in development or spermatogenesis. However, knockdown of RpL22 is partially rescued by RpL22-like and vice versa [34, 35]. Further work is needed to directly link effects on ribosome composition and mRNA translational output, as the two paralogs may interact with different pools of mRNA in the testis [35].

Interestingly, we found little differences between 80S and polysomal ribosome composition apart from an enrichment of RpL24-like and RpL7-like in 80S ribosomes in the ovary. RpL24-like is thought to have a role in the formation and processing pre-60S complexes (by similarity), with RpL24 replacing RpL24-like at the very end of processing [36]. Given that we saw enrichment of RpL24-like in 80S compared to polysomes in the ovary, it suggests that a proportion of these 80S complexes could represent the final stage of testing 80S competency in the ovaries. It is not clear why this would be the case in the ovary and not in other tissues. RpL24-like is present in other insects and some non-insect arthropods (FlyBase). A paralog switching event between RpL24 and RpL24-like could be important in translation initiation or indeed provide a platform for additional proteins to bind to the ribosome, given RpL24/RpL24-like is located close to RpL22/RpL22-like.

Several of the RPs that have gonad specific paralog pairs (including RpS19, RpS5, RpS10, RpS28 and RpL22 [37, 38]) have been linked with human diseases, specifically Diamond-Blackfan anemia

and cancer (Table 2). Thus, it will be important to uncover their contribution to mRNA translation regulation and work *in vivo* using *Drosophila* could help understand how they contribute to the translation of specific mRNAs.

One of the few canonical RPs we found to be differentially incorporated was RpS11 in the head 80S ribosomes. RpS11 phosphorylation, in humans, has been found to be linked to Parkinson's disease [39] and higher levels of RpS11 correlate with poorer prognosis in glioblastoma patients [40]. Therefore, understanding RpS11 levels in *Drosophila* head could provide a mechanism of future exploration for dissecting the molecular mechanisms by which RP mutations result in human disease.

Altogether our data reveal ribosome heterogeneity occurs in a tissue specific manner. Paralog-switching events are most abundant in the gonads and our structural analysis has provided insights into how this switch might regulate translation mechanistically. Additionally, our evolutionary data suggest specialisation may represent a conserved mechanism of translation regulation across eukaryotes.

Materials and Methods

Growth conditions

Drosophila melanogaster wild type (Dahomey) were raised on standard sugar–yeast agar (SYA) [41] . Flies were kept at 25°C and 50% humidity with a 12:12 hr light:dark cycle in 6 oz Square Bottom Bottles (Flystuff). Semi-adherent S2 cells were maintained in Schneider's medium containing L-glutamine (Sigma) supplemented with 10% FBS (Sigma), 100 U/mL penicillin, 100 µg/mL streptomycin, 25 µg/mL amphotericin B (GE Healthcare) and maintained at 26°C in non-vented, adherent flasks (Sarstedt).

Tissue harvest

~300 pairs of ovaries were harvested from 3-6 day old females in 1X PBS (Lonza) with 1 mM DTT (Sigma) and 1 U/µL RNAsin Plus (Promega) and flash frozen in liquid nitrogen. ~500 (rep 1) and ~1000 (rep 2) pairs of testes were harvested from 1-4 day old males in 1X PBS with 4 mM DTT and 1 U/µL RNAsin Plus and flash frozen in groups of ~10 pairs. ~500 heads (50:50 female:male, 0-4 days old) per gradient were isolated by flash freezing whole flies and subjecting them to mechanical shock to detach heads. Heads were passed through 1 mm mesh filter with liquid nitrogen and transferred to Dounce homogeniser for lysis. ~500 µL of 0-2 hour embryos/gradient were obtained from cages after pre-clearing for 2 hours. Laying plates comprised of 3.3% agar, 37.5% medium red grape juice compound (Young's Brew) and 0.3% methyl 4-hydroxybenzoate, supplemented with yeast paste of active dried yeast (DCL) and dH20. Embryos were washed in dH20 and embryo wash buffer (102.5 mM NaCl (Sigma), 0.04% TritonX-100 (Sigma) and then flash frozen with minimal liquid. ~120 $x10^6$ cells/gradient were treated with 100 µg/mL cycloheximide (Sigma) for 3 minutes before harvesting. Cells were pelleted at 800 xg for 8 minutes, washed in ice-cold 1X PBS supplemented with 100 µg/mL cycloheximide.

Ribosome purification

All stages were performed on ice or at 4°C wherever possible. Ovaries and testes were disrupted using RNase-free 1.5mL pestles (SLS) in lysis buffer A (50 mM Tris-HCl pH 8 (Sigma), 150 mM NaCl, 10 mM $MgCl_2$ (Fluka), 1% IGEPAL CA-630 (Sigma), 1 mM DTT, 100 µg/mL cycloheximide, 2 U/µL Turbo DNase (Thermo Fisher), 0.2 U/µL RNasin Plus, 1X EDTA-free protease inhibitor cocktail (Roche)). Ovaries, testes and S2 cells were lysed in 500 µL lysis buffer A. Heads were lysed using 8 mL Dounce homogeniser with loose pestle in 1.5mL lysis buffer B (10 mM Tris-HCl pH 7.5 (Gibco), 150 mM NaCl, 10 mM MgCl2, 1% IGEPAL CA-630, 1% Triton X-100, 0.5% sodium deoxycholate (Sigma), 2 mM DTT, 200 µg/mL cycloheximide, 2 U/µL Turbo DNase, 40 U/mL RNAsin Plus, 1X EDTA-free protease inhibitor cocktail). Then 500 µL aliquots were transferred to 2 ml

Dounce with tight pestle and further lysed for approximately 30 strokes. Embryos were ground in liquid nitrogen using pestle and mortar and added to lysis buffer B. All lysates were lysed for ≥30 mins with occasional agitation, then centrifuged for 5 minutes at 17,000 xg to remove nuclei. Head and embryo cytoplasmic supernatants were obtained by avoiding both floating fat and insoluble pellet and repeatedly centrifuged until free of debris.

Cytoplasmic lysates were loaded onto 18 – 60% sucrose gradients (50mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM $MgCl_2$, 100 µg/mL cycloheximide, 1 mM DTT, 1X EDTA-free protease inhibitor cocktail) and ultra-centrifuged in SW40Ti rotor (Beckman) for 3.5 h at 170,920 xg at 4°C. Fractions were collected using a Gradient Station (Biocomp) equipped with a fraction collector (Gilson) and Econo UV monitor (BioRad). Fractions containing 80S were combined, and same with polysomes. Fractions were concentrated using 30 kDa column (Amicon Ultra-4 or Ultra-15) at 4°C and buffer exchanged (50 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM $MgCl_2$) until final sucrose ≥0.1%. Samples were quantified using Qubit Protein Assay Kit.

TMT mass spectrometry

40 µg purified protein per sample was subject to tandem mass tag mass spectrometry using Orbitrap Fusion Mass Spec machine by University of Bristol Proteomics Facility. Sequest search was performed against the UniProt *Drosophila* database plus 'Common Contaminants' database and filtered using a 5% FDR cut-off [42].

TMT analysis

To be confident of protein identity and presence results were filtered to only include protein IDs where 30% of each protein was covered my mass spec peptides and based on 2 or more unique peptide identities. Only peptide IDs corresponding to *D. melanogaster* proteins were considered. For TMT1 this resulted in a list of 836 proteins and TMT 2, 836 proteins. The full list of *D. melanogaster* ribosomal proteins was extracted from FlyBase (April 2019). Abundances are the sum of the S/N values for the TMT reporter groups for all PSMs matched to the protein. Normalised abundances of these values are then normalised to Total Peptide Amount in each sample such that the total signal from each TMT tag is the same. Normalised abundances were used to quantify levels of proteins. To quantify relative incorporation of paralogs into ribosomes normalised abundances were used to generate percentages, assuming the sum of paralog 1 and paralog 2 were 100%. Several paralogs were not detected and therefore calculated to be 0%, several failed to pass our standard thresholds but were included in this analysis for completeness. Analysis of TMT data and hierarchical clustering was performed in R.

Source of RNA-Seq data

RNA-Seq data was extracted from ModMine (intermine.modencode.org) with data from modENCODE project.

Cryo-EM

For cryo-EM, 400 mesh copper grids with a supporting carbon lacey film coated with an ultra-thin carbon support film < 3 nm thick (Agar Scientific, UK) were employed. Grids were glow-discharged for 30 seconds (easiGlow, Ted Pella) prior to applying 3 µL of purified ribosomes, and vitrification was performed by plunge-freezing in liquid ethane cooled by liquid nitrogen using a Leica EM GP device (Leica Microsystems). Samples were diluted using the buffer exchange buffer (50 mM Tris pH 8, 150 mM NaCl, 10 mM $MgCl_2$) as required. Cryo-EM data was collected on a FEI Titan Krios (Astbury Biostructure Laboratory, University of Leeds) EM at 300 kV, using a total electron dose of 80 $e^-/Å^2$ and a magnification of 75,000 × at -2 to -4 µm defocus. Movies were recorded using the EPU automated acquisition software on a FEI Falcon III direct electron detector, with a final pixel size of 1.065 Å/pixel (Sup Table 1).

Image processing

Initial pre-processing and on-the-fly analysis of data was performed as previously described [43]. Image processing was carried out using RELION 2.0/2.1 or 3.0 [44]. MOTIONCOR2 [45] was used to correct for beam-induced motion and calculate averages of each movie. gCTF [46] was used to contrast transfer function determination. Particles were automatically picked using the Laplacian of Gaussian function from RELION [47]. Particles were classified using a reference-free 2D classification. Particles contributing to the best 2D class averages were then used to generate an initial 3D model. This 3D model was used for 3D classification, and the best 3D classes/class were 3D refined, followed by per-particle CTF correction and Bayesian polishing [47]. Post-processing was employed to mask the model, and to estimate and correct for the B-factor of the maps [48]. The testis 80S map was further processed by multi-body refinement, as previously described [49]. The final resolutions were determined using the 'gold standard' Fourier shell correlation (FSC = 0.143) criterion (Sup Table 1). Local resolution was estimated using the local resolution feature in RELION.

Atomic modelling

D. melanogaster embryo ribosome (pdb code 4v6w) was used as a model to calculate the structures of the testis and ovary ribosomes. First, the full atomic model was fitted into the testis 80S cryo-EM average using the 'fit in map' tool from Chimera [50]. Then, fitting was refined by rigid-body fitting individual protein and RNA pdbs into the maps using Chimera. The 18S and 28S ribosomal RNAs were split into two separate rigid bodies each. Proteins and RNAs not present in our averages (i.e. elongation factor 2 and Vig2 for all models, and E-tRNA for the 80S ribosome models) and proteins and RNA with poor densities (i.e. RpLP0 and RpL12, and some regions of the 18S and 28S ribosomal RNAs) were removed at this stage. The paralog proteins used for each ribosome are listed in Table 1. For the testis 80S atomic model, CG31694-PA was modelled using SWISS-MODEL [51]. For the testis polysome model, the mRNA was based on pdb model 6HCJ, and the E-tRNA on pdb model 4V6W. The full atomic models were refined using Phenix [52], and the paralogs listed in Fig 2A were manually inspected and corrected using COOT [53] (except Rp10Ab, which was not manually inspected due to the low resolution of that area in the average maps, and RpLP0, which was not present in the model). This cycle was repeated at least three times per ribosome model. The quality of the atomic models was assessed using the validation server from the pdb website (https://validate-pdbe.wwpdb.org/). As the 60S acidic ribosomal protein P0 deposited in the pdb (4v6w) is from *Homo sapiens*, we generated a homology model using SWISS-MODEL. This protein was rigid-body fitted using Chimera after the atomic model refinement and is displayed in Fig 5 for relative position and size comparison purposes only. Figures were generated using Chimera.

Vertebrate dataset construction

Coding DNA sequence (CDS) data for 207 vertebrate animals and 4 non-vertebrates (*D. melanogaster*, two *Caenorhabditis* species and *S. cerevisiae*) was obtained from Ensembl (release 97, [54]). We performed homology searches using two human Rpl22 family proteins (RPL22 and RPL22L) were searched against 6,922,005 protein sequences using BLASTp (e^{-5}) [55]. We identified 1,082 potential RpL22 proteins from 185 vertebrates and 4 non-vertebrates, which were homologous to one or both human RpL22 proteins. As an initial step to reduce the amount of redundancy in the vertebrate dataset, 181 potential RpL22 proteins from 42 selected vertebrates (including humans) were retained to represent as broad a taxonomic sampling of the group. All non-vertebrate sequences, with the exception of two *S. cerevisiae* Rpl22 proteins (RPL22A and RPL22B), were also removed from the dataset. 92 alternative transcripts and spurious hits were removed from the dataset through manual cross-validation with Ensembl Genome Browser to give total of 87 vertebrate and 2 yeast RpL22 family proteins.

Invertebrate dataset construction

CDS data data for 78 invertebrate animals was obtained from Ensembl Metazoa (release 44, [54]). The sequence homology search was performed using two *D. melanogaster* Rpl22 family proteins (RPL22 and RPL22-like) were searched against 1,618,385 protein sequences using BLASTp (e^{-5}) [55]. BLASTp identified 90 potential Rpl22 family proteins across 70 invertebrates, which were homologous to one or both *D. melanogaster* Rpl22 proteins. 15 alternative transcripts and spurious hits were removed from the dataset through manual cross-validation with Ensembl Genome Browser to give total of 75 invertebrate RpL22 family proteins. Together with 87 vertebrate and 2 outgroup proteins, our final dataset consisted of 164 RpL22 family proteins sampled across the metazoan tree of life.

Phylogenetic reconstructions of metazoan RpL22 family

Initial phylogenetic reconstruction of the metazoan RpL22 family was performed using the full dataset of 164 sequences (87 invertebrate sequences, 75 vertebrate sequences and two yeast sequences). All sequences were aligned using three different alignment algorithms: MUSCLE [56], MAFFT [57] and PRANK [58]. MUSCLE was run with the default parameters, and MAFFT was run with the automatically-selected most-appropriate alignment strategy (in this case, L-INS-I). PRANK was run with both the default parameters and the PRANK+F method with "permanent" insertions. All four resultant alignments were compared against each other using MetAl [59], and were all judged to be mutually discordant based on differences of 20-25% between each pair of alignments. Column-based similarity scores were calculated for each alignment using the norMD statistic [60]. The MUSCLE alignment had the highest column-based similarity score (1.281) and was selected for further analysis. This alignment was trimmed using TrimAl's gappyout method [61]. Maximum-likelihood phylogenetic reconstruction was performed on the trimmed alignment using IQTREE [62], with a WAG+R6 model selected by ModelFinder Plus [63] and 100 bootstrap replicates.

A reduced sampling of the metazoan Rpl22 family was used to generate a phylogeny was performed using taxonomically-representative dataset containing 50 Rpl22 genes from 30 animals and *S. cerevisiae*. This dataset was aligned using the same four methods described above, and all alignments were judged to be mutually discordant (differences of 19-37%) using MetAl [59]. The MUSCLE alignment had the highest column-based similarity score assigned by norMD (0.702) and was selected for further analysis. As above, this alignment was trimmed using TrimAl's gappyout method. Maximum-likelihood phylogenetic reconstruction was performed on the trimmed alignment using IQTREE [62], with a DCMut+R3 model selected by ModelFinder Plus [63] and 100 bootstrap replicates.

Data deposition

The EM-density maps for testis 80S, testis polysomes and ovaries 80S are deposited in the EMDB under the accession numbers EMD-10622, EMD-10623 and EMD-10624. The refined models are deposited in the PDB under accession codes 6XU7, pdb 6XU7 and pdb 6XU8.

Acknowledgements

We thank the Astbury Biostructure Laboratory (ABSL) Facility Staff for assisting with cryo-EM data collection. All Electron Microscopy was performed at ABSL which was funded by the University of Leeds and the Wellcome Trust (108466/Z/15/Z). Electron microscopy image processing was partially undertaken on ARC3, part of the High Performance Computing (HPC) facilities at the University of Leeds. CGPM and MJOC would like to thank The University of Nottingham for HPC facilities and staffing. Mass spectrometry was performed by Bristol University Proteomics Facility. Julie Aspden and Juan Fontana are funded by the University of Leeds (University Academic Fellow scheme). This work was funded by Royal Society (RSG\R1\180102), BBSRC (BB/S007407/1), Wellcome Trust ISSF (105615/Z/14/Z), White Rose University Consortium- Collaborative Grant and MRC (MR/N000471/1). MA was funded from BBSRC DTP BB/M011151/1.

References

1. Kondrashov, N., et al., *Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning.* Cell, 2011. **145**(3): p. 383-97.
2. Shi, Z., et al., *Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide.* Molecular Cell, 2017.
3. Segev, N. and J.E. Gerst, *Specialized ribosomes and specific ribosomal protein paralogs control translation of mitochondrial proteins.* J Cell Biol, 2018. **217**(1): p. 117-126.
4. García-Marcos, A., et al., *Yeast ribosomal stalk heterogeneity in vivo shown by two-photon FCS and molecular brightness analysis.* Biophys J, 2008. **94**(7): p. 2884-90.
5. Bortoluzzi, S., et al., *Differential expression of genes coding for ribosomal proteins in different human tissues.* Bioinformatics, 2001. **17**(12): p. 1152-7.
6. Simsek, D., et al., *The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity.* Cell, 2017. **169**(6): p. 1051-1065 e18.
7. Kearse, M.G., A.S. Chen, and V.C. Ware, *Expression of ribosomal protein L22e family members in Drosophila melanogaster: rpL22-like is differentially expressed and alternatively spliced.* Nucleic Acids Res, 2011. **39**(7): p. 2701-16.
8. Carroll, A.J., et al., *Analysis of the Arabidopsis cytosolic ribosome proteome provides detailed insights into its components and their post-translational modification.* Mol Cell Proteomics, 2008. **7**(2): p. 347-69.
9. Natchiar, S.K., et al., *Visualization of chemical modifications in the human 80S ribosome structure.* Nature, 2017. **551**(7681): p. 472-477.
10. Dinman, J.D., *Pathways to Specialized Ribosomes: The Brussels Lecture.* J Mol Biol, 2016. **428**(10 Pt B): p. 2186-94.
11. Guimaraes, J.C. and M. Zavolan, *Patterns of ribosomal protein expression specify normal and malignant human cells.* Genome Biol, 2016. **17**(1): p. 236.
12. Szakonyi, D. and M.E. Byrne, *Ribosomal protein L27a is required for growth and patterning in Arabidopsis thaliana.* Plant J, 2011. **65**(2): p. 269-81.
13. Weijers, D., et al., *An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene.* Development, 2001. **128**(21): p. 4289-99.
14. Marygold, S.J., C.M. Coelho, and S.J. Leevers, *Genetic analysis of RpL38 and RpL5, two minute genes located in the centric heterochromatin of chromosome 2 of Drosophila melanogaster.* Genetics, 2005. **169**(2): p. 683-95.
15. Thurmond, J., et al., *FlyBase 2.0: the next generation.* Nucleic Acids Res, 2019. **47**(D1): p. D759-D765.
16. Barakat, A., et al., *The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome.* Plant Physiol, 2001. **127**(2): p. 398-415.
17. Richter, J.D. and P. Lasko, *Translational control in oocyte development.* Cold Spring Harb Perspect Biol, 2011. **3**(9): p. a002758.
18. Qin, X., et al., *Global analyses of mRNA translational control during early Drosophila embryogenesis.* Genome Biol, 2007. **8**(4): p. R63.
19. Boria, I., et al., *The ribosomal basis of Diamond-Blackfan Anemia: mutation and database update.* Hum Mutat, 2010. **31**(12): p. 1269-79.
20. Slaidina, M. and R. Lehmann, *Translational control in germline stem cell development.* J Cell Biol, 2014. **207**(1): p. 13-21.
21. Yu, J., et al., *Protein synthesis and degradation are essential to regulate germline stem cell homeostasis in Drosophila testes.* Development, 2016. **143**(16): p. 2930-45.
22. Brengues, M., L. Pintard, and B. Lapeyre, *mRNA decay is rapidly induced after spore germination of Saccharomyces cerevisiae.* J Biol Chem, 2002. **277**(43): p. 40505-12.
23. Vedelek, V., et al., *Analysis of Drosophila melanogaster testis transcriptome.* BMC Genomics, 2018. **19**(1): p. 697.

24. Ghosh, S. and P. Lasko, *Loss-of-function analysis reveals distinct requirements of the translation initiation factors eIF4E, eIF4E-3, eIF4G and eIF4G2 in Drosophila spermatogenesis.* PLoS One, 2015. **10**(4): p. e0122519.
25. Lambertsson, A., *The minute genes in Drosophila and their molecular functions.* Adv Genet, 1998. **38**: p. 69-134.
26. Kongsuwan, K., et al., *A Drosophila Minute gene encodes a ribosomal protein.* Nature, 1985. **317**(6037): p. 555-8.
27. Marygold, S.J., et al., *The ribosomal protein genes and Minute loci of Drosophila melanogaster.* Genome Biol, 2007. **8**(10): p. R216.
28. Martinez Barrio, A., et al., *Targeted resequencing and analysis of the Diamond-Blackfan anemia disease locus RPS19.* PLoS One, 2009. **4**(7): p. e6172.
29. Anger, A.M., et al., *Structures of the human and Drosophila 80S ribosome.* Nature, 2013. **497**(7447): p. 80-5.
30. Heyer, E.E. and M.J. Moore, *Redefining the Translational Status of 80S Monosomes.* Cell, 2016. **164**(4): p. 757-69.
31. Branco-Price, C., et al., *Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation.* Ann Bot, 2005. **96**(4): p. 647-60.
32. Brown, A., et al., *Structures of translationally inactive mammalian ribosomes.* Elife, 2018. **7**.
33. Kong, J., et al., *A ribosomal protein S5 isoform is essential for oogenesis and interacts with distinct RNAs in Drosophila melanogaster.* Sci Rep, 2019. **9**(1): p. 13779.
34. Mageeney, C.M., et al., *Functional interplay between ribosomal protein paralogues in the eRpL22 family in Drosophila melanogaster.* Fly (Austin), 2018. **12**(3-4): p. 143-163.
35. Mageeney, C.M. and V.C. Ware, *Specialized eRpL22 paralogue-specific ribosomes regulate specific mRNA translation in spermatogenesis in.* Mol Biol Cell, 2019. **30**(17): p. 2240-2253.
36. Saveanu, C., et al., *Sequential protein association with nascent 60S ribosomal particles.* Mol Cell Biol, 2003. **23**(13): p. 4449-60.
37. Campagnoli, M.F., et al., *RPS19 mutations in patients with Diamond-Blackfan anemia.* Hum Mutat, 2008. **29**(7): p. 911-20.
38. Cao, B., et al., *Cancer-mutated ribosome protein L22 (RPL22/eL22) suppresses cancer cell survival by blocking p53-MDM2 circuit.* Oncotarget, 2017. **8**(53): p. 90651-90661.
39. Martin, I., V.L. Dawson, and T.M. Dawson, *Recent advances in the genetics of Parkinson's disease.* Annu Rev Genomics Hum Genet, 2011. **12**: p. 301-25.
40. Yong, W.H., et al., *Ribosomal Proteins RPS11 and RPS20, Two Stress-Response Markers of Glioblastoma Stem Cells, Are Novel Predictors of Poor Prognosis in Glioblastoma Patients.* PLoS One, 2015. **10**(10): p. e0141334.
41. Bass, T.M., et al., *Optimization of dietary restriction protocols in Drosophila.* J Gerontol A Biol Sci Med Sci, 2007. **62**(10): p. 1071-81.
42. Consortium, U., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.
43. Thompson, R.F., et al., *Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy.* Nat Protoc, 2019. **14**(1): p. 100-118.
44. Fernandez-Leiro, R. and S.H.W. Scheres, *A pipeline approach to single-particle processing in RELION.* Acta Crystallogr D Struct Biol, 2017. **73**(Pt 6): p. 496-502.
45. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nat Methods, 2017. **14**(4): p. 331-332.
46. Zhang, K., *Gctf: Real-time CTF determination and correction.* J Struct Biol, 2016. **193**(1): p. 1-12.
47. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.
48. Scheres, S.H. and S. Chen, *Prevention of overfitting in cryo-EM structure determination.* Nat Methods, 2012. **9**(9): p. 853-4.

49. Nakane, T., et al., *Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION.* Elife, 2018. **7**.
50. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
51. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.
52. Afonine, P.V., et al., *New tools for the analysis and validation of cryo-EM maps and atomic models.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 9): p. 814-840.
53. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
54. Cunningham, F., et al., *Ensembl 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D745-D751.
55. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
56. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
57. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80.
58. Loytynoja, A., *Phylogeny-aware alignment with PRANK.* Methods Mol Biol, 2014. **1079**: p. 155-70.
59. Blackburne, B.P. and S. Whelan, *Measuring the distance between multiple sequence alignments.* Bioinformatics, 2012. **28**(4): p. 495-502.
60. Thompson, J.D., et al., *Towards a reliable objective function for multiple sequence alignments.* J Mol Biol, 2001. **314**(4): p. 937-51.
61. Capella-Gutierrez, S., J.M. Silla-Martinez, and T. Gabaldon, *trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.* Bioinformatics, 2009. **25**(15): p. 1972-3.
62. Nguyen, L.T., et al., *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.* Mol Biol Evol, 2015. **32**(1): p. 268-74.
63. Kalyaanamoorthy, S., et al., *ModelFinder: fast model selection for accurate phylogenetic estimates.* Nat Methods, 2017. **14**(6): p. 587-589.

Figure 1

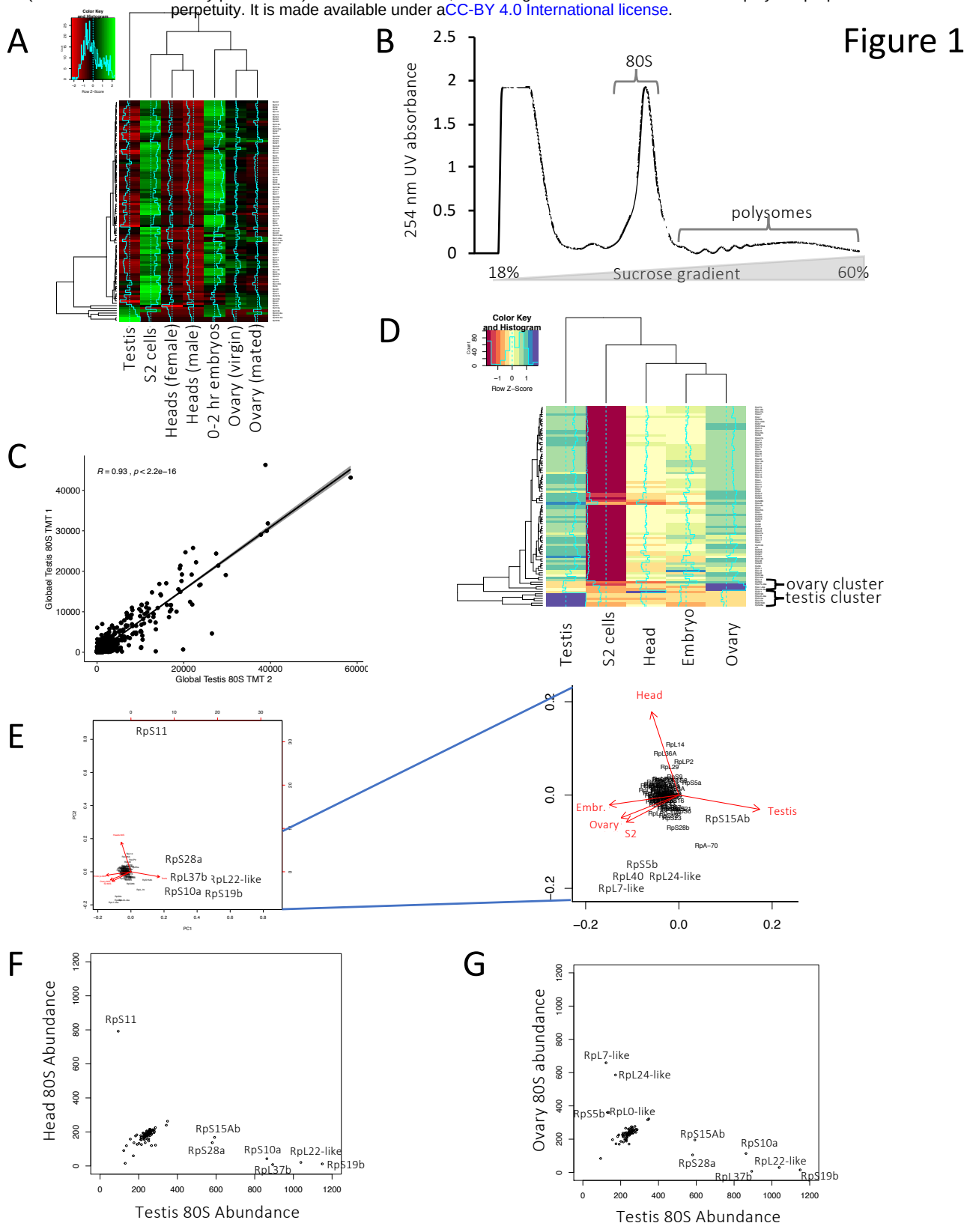
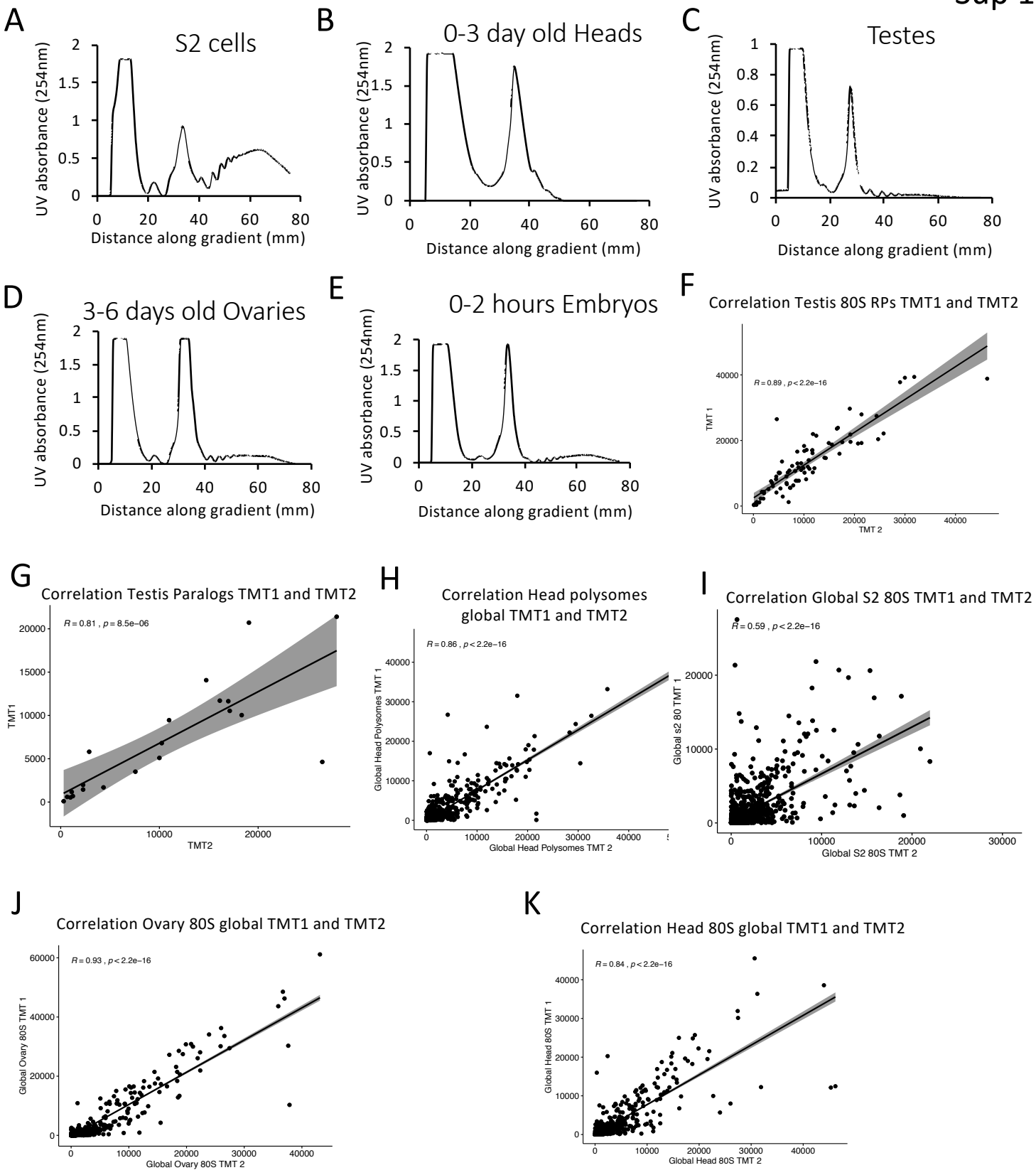


Figure 1: Heterogeneous ribosome populations exist in different tissues

(A) Hierarchical clustering of modENCODE RNA-Seq data for Ribosomal Proteins (RPs) across testis, ovary, head, embryo and S2 cells reveals differences in RP expression. RPKMs are clustered by row. (B) Schematic of strategy used to isolate and compare ribosomal complex composition. (C) Correlation of two biological repeats of TMT mass spec experiment for protein abundances within 80S ribosomes isolated from testis, with Pearson's correlation calculated, shows replicates are reproducible. (D) Identification of heterogeneity of ribosome protein composition across samples. Hierarchical clustering of normalised protein abundances from replicate 2, clustered according to row i.e. ribosomal protein. (E) PCA of RPs, showing most cluster together and behave similarly across tissues. Two groups of proteins are different from majority. (F) Scatter plot of scaled abundances for testis 80S and head 80S for ribosomal proteins. Identification of testis enriched components. Proteins enriched in one sample over the other are labelled. (G) Scatter plot of scaled abundances for testis 80S and ovary 80S for ribosomal proteins. Identification of both testis and ovary enriched components. Proteins enriched in one sample over the other are labelled.



Sup 1: Determined ribosomal composition in tissues and during development

254 nm UV plots across sucrose gradients with 80S and polysomal complexes isolated from (A) S2 cells, (B) 50:50 mixture of female:male 0-3 day old Heads. (C) ~500 pairs of 1-4 day old adult testes, (D) ~500 pairs of 3-6 day old adult ovaries, (E) 0-2 hour embryos. Correlation of two TMT mass spec experiments for protein abundances within 80S ribosomes isolated with Pearson's correlation calculated, shows replicates are reproducible from (F) RPs in testis 80S, (G) RP paralogs in testis 80S, (H) head polysome, (I) S2 80S, (J) ovary 80S, (K) head 80S.

Figure 2

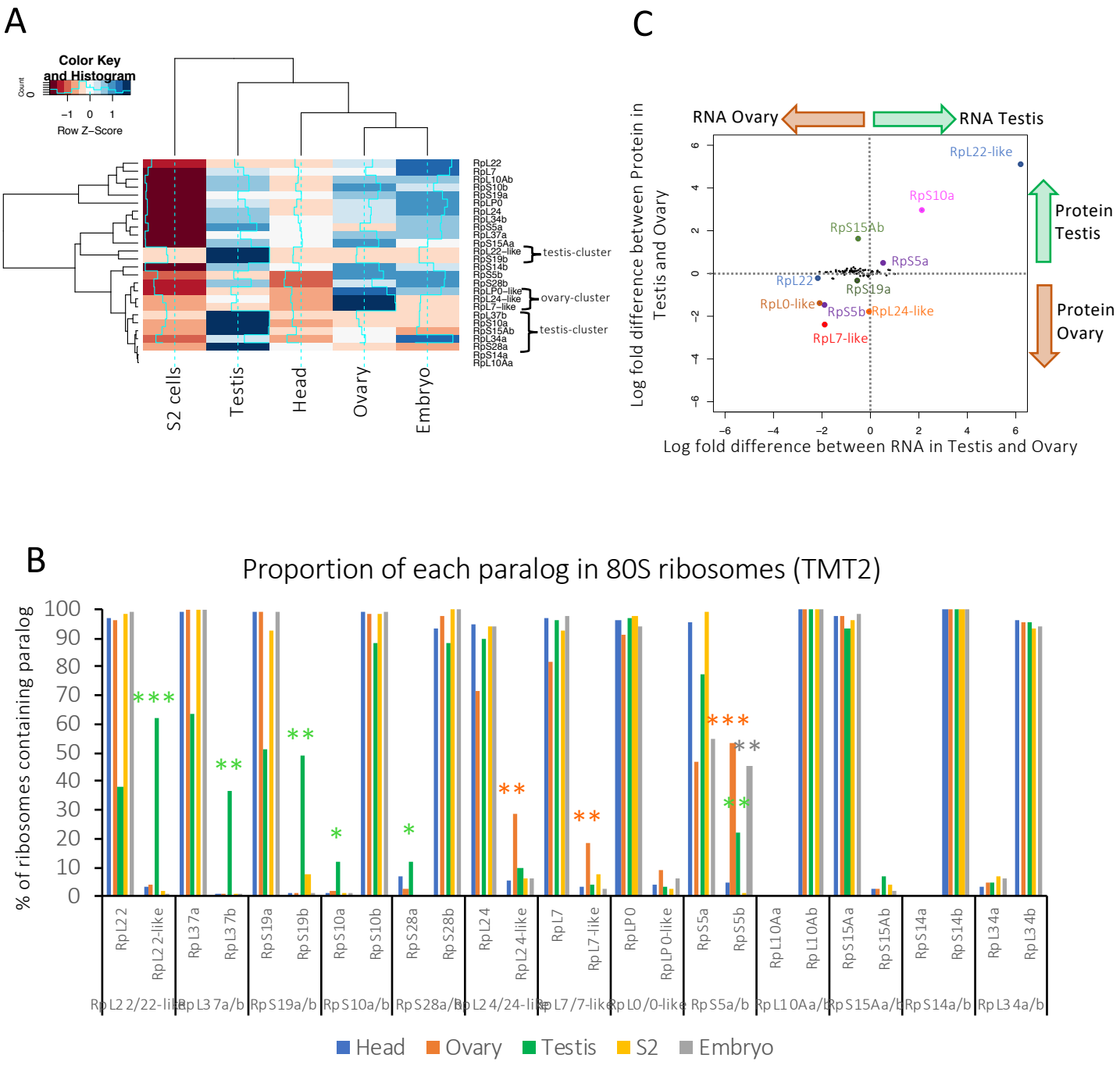
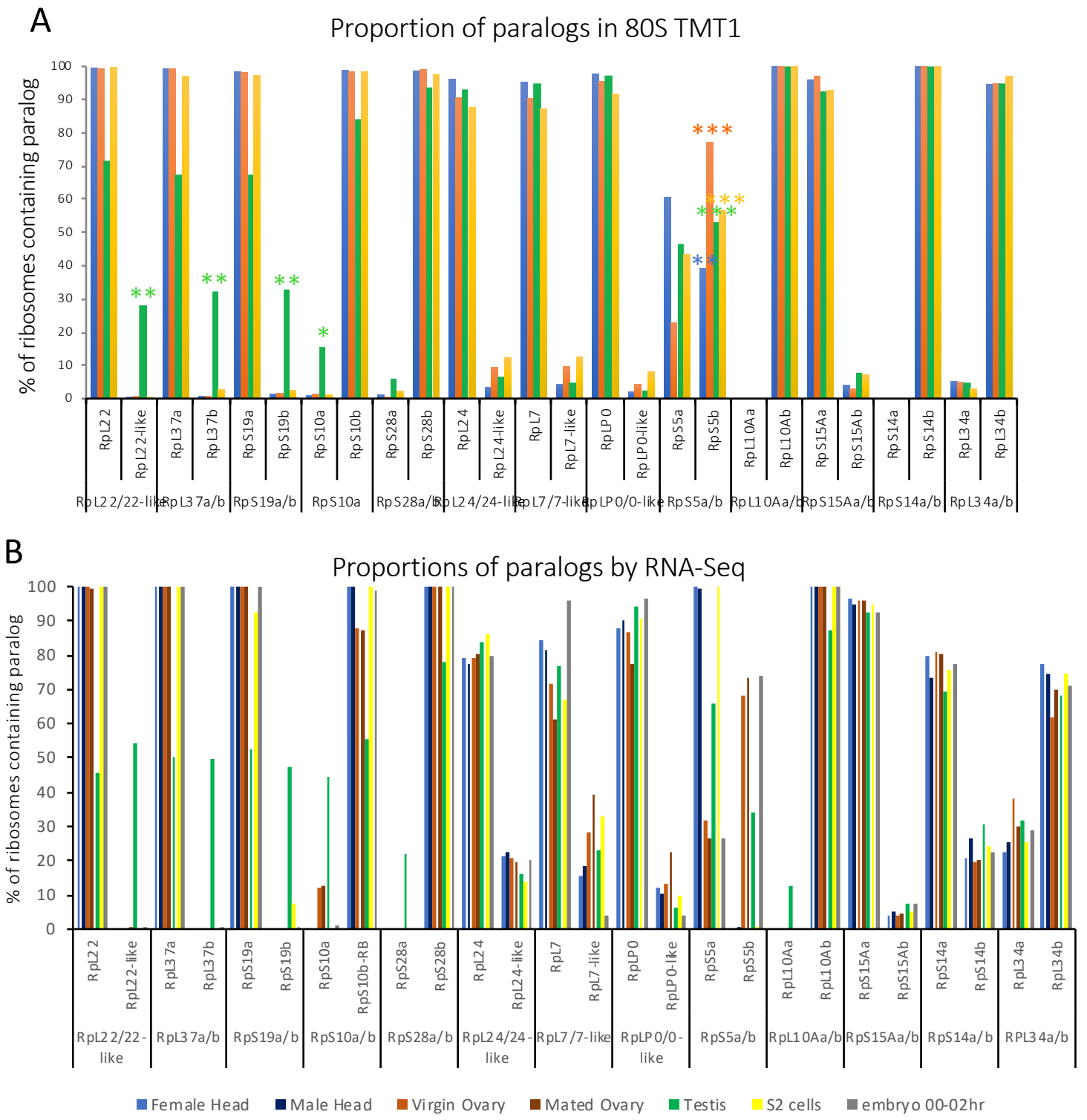


Figure 2: Specialisation of gonad ribosomes through paralog switching

(A) Hierarchical clustering of RP paralog proteins across the 5 samples scaled by row. (B) Proportion of ribosomes containing either of two ribosomal protein paralogs (expressed at percentage). 5 pairs have testis specificity and 4 ovary specificity. *** indicates if second paralog is >50%; with ** if >25% and with * if >10%. (C) Log fold difference plot showing differences between testis and ovary at both RNA (ovary RNA, mated females) and protein levels.



Sup 2: Specialisation of gonad ribosomes through paralog switching

(A) Bar chart of paralog pair % incorporation into 80S ribosomes across different tissues from TMT experiment 1 (head, ovary, testis, S2 cells and 0-2hr embryo). (B) Bar chart of paralog pair % relative expression by ModENCODE RNA-Seq across different tissues (female head, male head, virgin ovary, mated ovary, testis, S2 cells and 0-2hr embryo).

Figure 3

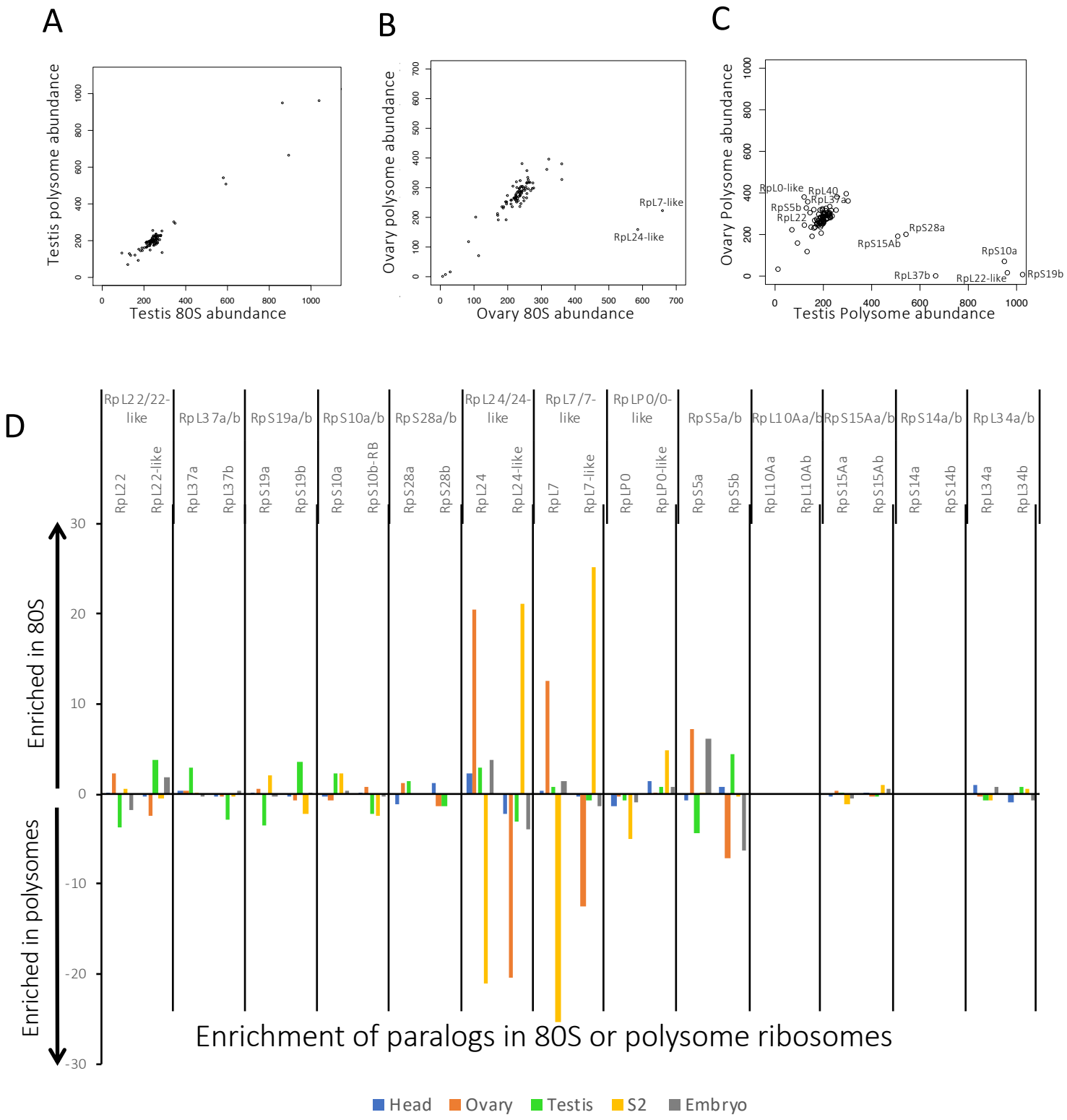
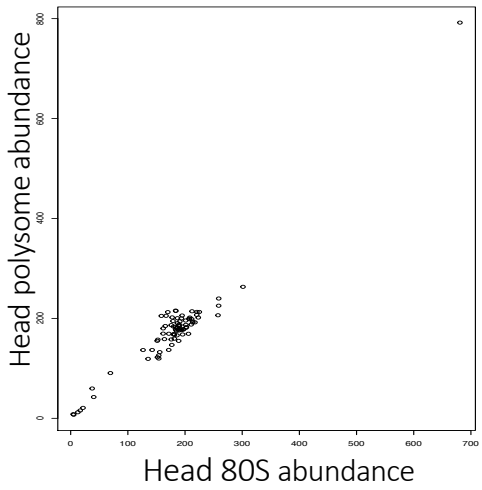


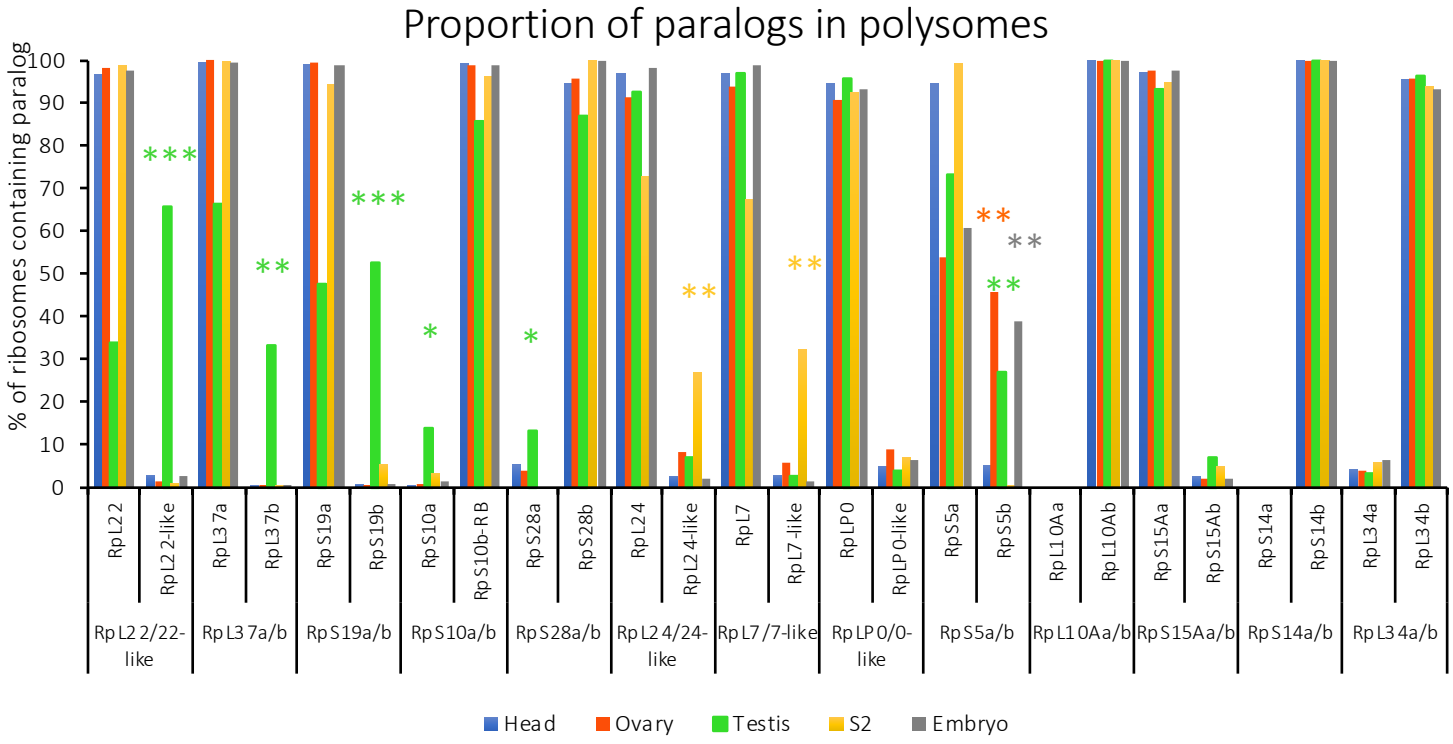
Figure 3: Little difference between composition of 80S and polysome ribosomes

(A) Scatter plot of scaled abundances for testis 80S and testis polysomes for ribosomal proteins shows little difference. (B) Scatter plot of scaled abundances for ovary 80S and ovary polysomes for ribosomal proteins reveals 80S enrichment of RpL7-like and RpL24-like in the 80S. (C) Scatter plot of scaled abundances for testis polysomes and ovary polysomes for ribosomal proteins reveals enrichment of similar proteins differentially incorporated in 80S; Testis RpD15Ab, RpS28a, RpL37b, RpS10a, RpS19b and RpL22-like; Ovary RpL37a, RpL22, RpL40, RpL0-like and RpS5b. (D) Comparison of overall difference in composition between 80S and polysomes across all paralogs and all tissues; showing main difference between 80S and polysome composition is with ovary for RpL24/24-like and RpL7/7-like.

A



B



Sup 3: Little difference between composition of 80S and polysome ribosomes

(A) Scatter plot of scaled abundances for head 80S and head polysomes for ribosomal proteins reveals little difference in ribosomes composition. (B) Bar chart of paralog pair % incorporation into polysome ribosomes across different tissues from TMT experiment 2 (head, ovary, testis, S2 cells and 0-2hr embryo).

Figure 4

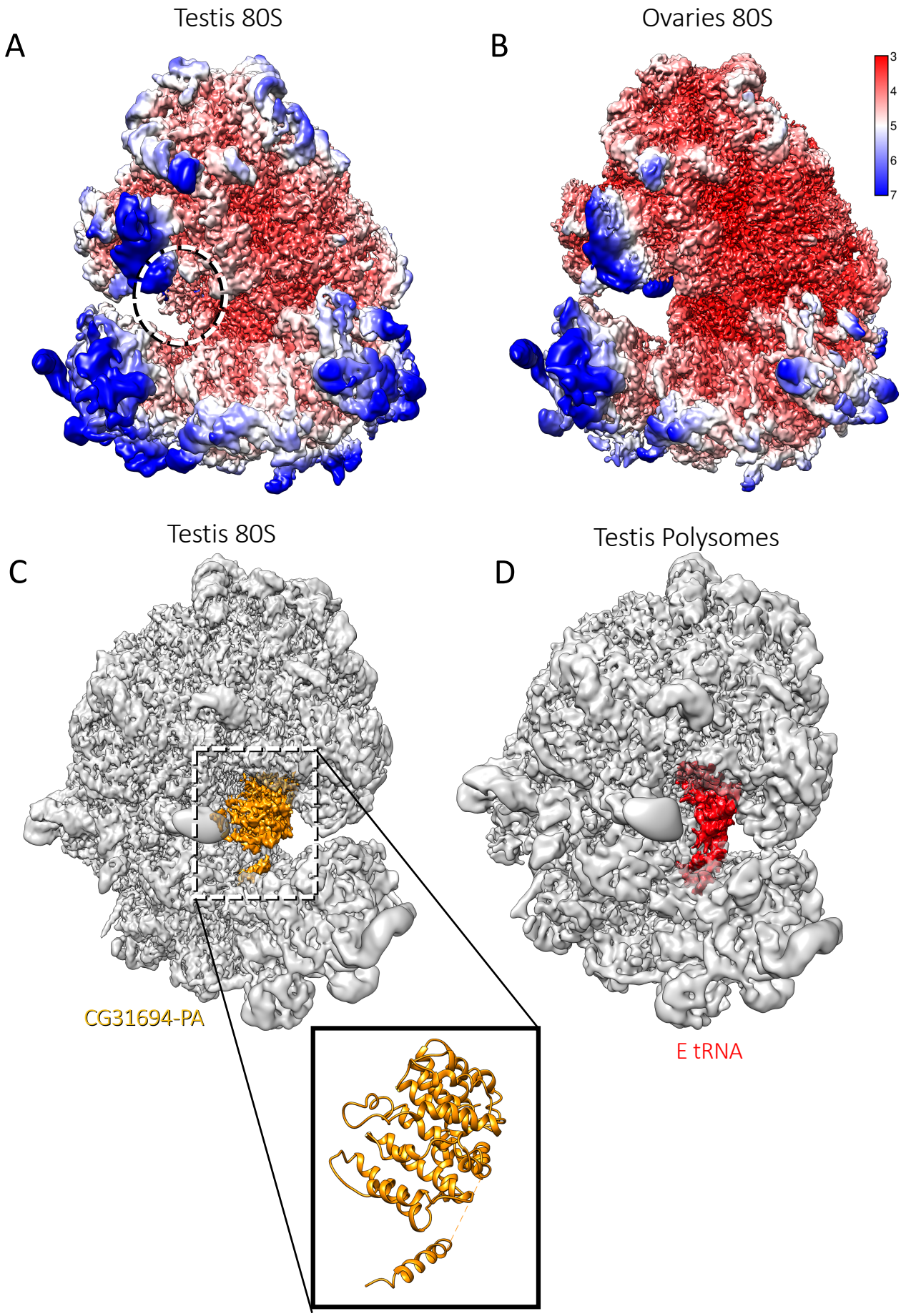
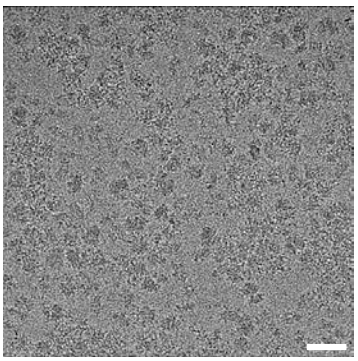


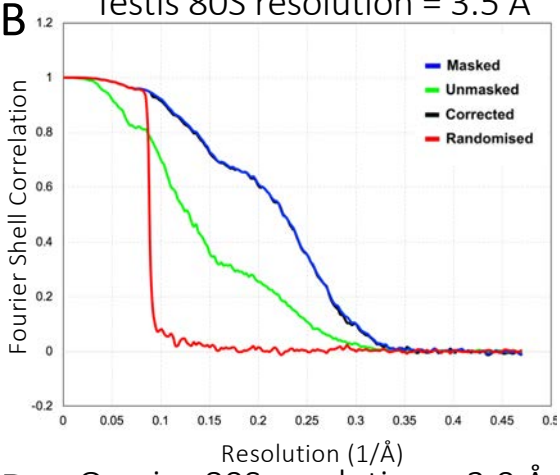
Figure 4: Cryo-electron microscopy of testis and ovary ribosomes

A and B) Electron microscopy averages of testis 80S (A) and ovaries 80S (B) ribosomes, color-coded according to their local resolution. C and D) cryo-EM averages of 80S (C) and polysome (D) ribosomes purified from testes, with segmented densities corresponding to CG31694-PA and E-tRNA colored in orange and red respectively. The atomic model of CG31694-PA is shown in the inset.

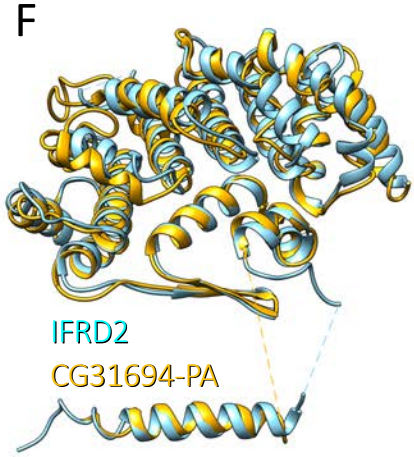
A



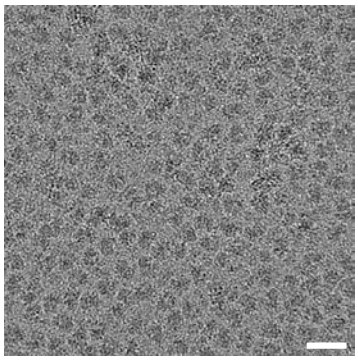
B



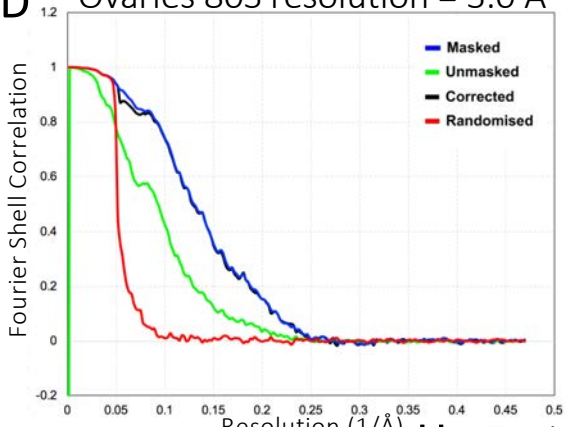
F



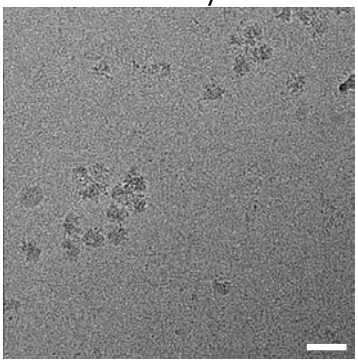
C



D



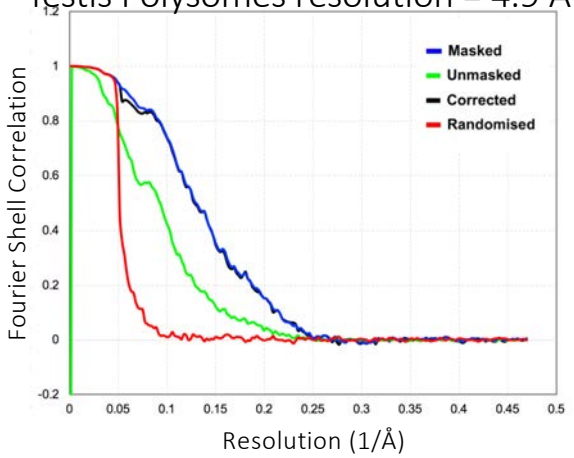
G



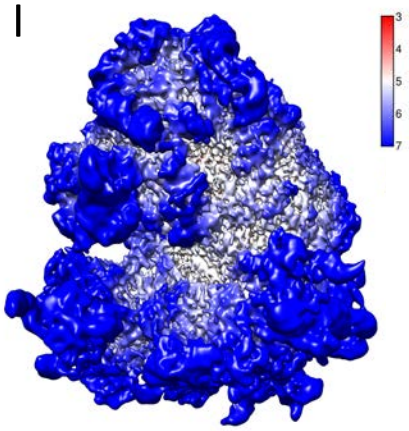
E

```
CLUSTAL O(1.2.4) multiple sequence alignment


CG31694-PA      ------------------------------------------------------------    0
IFRD2           MAWNSPSRRPVWQGGAPREDGGARGVWLPSSGQVSAQRTGRRLVGLEPTPTGSLTPRPPR   60


CG31694-PA      ----MPRRNKKSAAGRG-------RT----NDSN-SEDESFD-NVSVYSHMSEVASSEAT   43
IFRD2           PVPGMPRARKGNTLRKGGQRRGGGARSSAQADSGSSDDEAASEARSTASECPSLLSTTAE.  120
                    *** .* .:  :*              **. *:**: .   *. *.  .: *: *


CG31694-PA      ---------DELANERFEEKFEKALEQATEKSAQTRVQALQAICELLMHRYMPDFVEDRK   94
IFRD2           DSLGGDVVDEQGQQEDLEEKLKEYVDCLTDKSAKTRQGALESLRLALASRLLPDFLLERR.  180
                         ::  :* :***::: ::  *:***:**  **:::   *  * :***: :*:


CG31694-PA      MTLMDFVEKSIRRGKGQEQVWGARLAPLLVLQMGGDE---GISKAMNQFLLNTVQDKSVG   151
IFRD2           LTLADALEKCLKKGKGEEQALAAAVLGLLCVQLGPGPKGEELFHSLQPLLVSVLSDSTAS.  240
                :** * :**.:::***:**. .* :  ** :*:* .     : :::: :*:..:.*:..


CG31694-PA      FDARAKCCTAVGLLSFLGCEDVGELVHLMQSFEAIFAGSYLRGDDKTPVSVTAEAGTFHA   211
IFRD2           PAARLHCASALGLGCYVAAADIQDLVSCLACLESVFSRFYGLGGSSTSPVVPASLHGLLS   300
                  ** :*.:*:** .:::. *: :**  : .:*::*:  *  *...*   * *.   : :


CG31694-PA      EALNAWGLLLTLIPSGDFVSLMTTGQNMFPSIKKFLGLLQSTHLDVRMAAGETIALILES   271
IFRD2           AALQAWALLLTICPSTQISHILDRQL------PRLPQLLSSESVNLRIAAGETIALLFEL   354
                 **:**.****: ** ::  ::           ::  **.*  :::*:********::*


CG31694-PA      GRAHEEDFLEDDIAELSEAVKQLATDSHKYRAKRDRKAQRATFRDVLRYLEEDISPEISI   331
IFRD2           ARDLEEEFVYEDMEALCSVLRTLATDSNKYRAKADRRRQRSTFRAVLHSVEGGECEEEIV   414
                .*  **:*: :*:  *...:: *****:***** **: **:*** **: :* . . *  :


CG31694-PA      RFGTESLTLDSWSIHHQYSAMCTVMGPGMTSQLQENEFIRDIFQLGPRPTNTGINGNAKV   391
IFRD2           RFGFEVLYMDSWARHRIYAAFKEVLGSGMHHHLQNNELLRDIFGLGPVLLLDATA-LKAC   473
                *** * * :***: *: *:*:  *:* **  :**:**::**** ***     .


CG31694-PA      KPTKLERHLVNAAAFKARSITRGKNRDKRSAVVT   425
IFRD2           KVPRFEKHLYNAAAFKARTKARSRVRDKRADIL-   506
                *  ::*:** ********: :*.: ****: ::
```

H



I



Sup 4: Cryo-electron microscopy of testis and ovaries ribosomes

Cryo-electron micrographs (A, C and G) and FSC curves (B, D and H) for the cryo-EM averages of testis 80S (A & B), ovaries 80S (C & D) and testis polysomes (G & H). (E) Clustal-omega alignment of human, rabbit IRFD2 protein sequence and *D.melanogaster* CG31694-PA protein sequence. (F) Comparison of the atomic models of IFRD2 and CG31694-PA. (I) Electron microscopy average of testis polysomes colour-coded according to its local resolution. Scale bars for A, C & G, 50 nm.

	Testis 80S	Ovary 80S	Testis Polysomes
Data collection			
Microscope	FEI Titan KRIOS	FEI Titan KRIOS	FEI Titan KRIOS
Voltage (keV)	300	300	300
Detector	FEI Falcon III	FEI Falcon III	FEI Falcon III
Magnification	x75,000	x75,000	x75,000
Defocus range	-2 to -4	-2 to -4	-2 to -4
Pixel size (Å)	1.065	1.065	1.065
Electron dose ($e^-/Å^2$)	80	80	80
Electron dose per frame ($e^-/Å^2$)	1.35	1.35	1.35
Exposure (sec)	2	2	2
No. of frames	60	60	60
No. of micrographs	5,241	9,076	2,758
Data processing			
Symmetry Point Group	C1	C1	C1
Final particle number	46,878	185,913	10,392
Map average resolution (Å, 0.143 FSC threshold)	3.5	3.0	4.9
Map sharpening B-factor ($Å^2$)	-150	-118	-197
Multi-body refinement		N/A	N/A
Large subunit			
Map average resolution (Å, 0.143 FSC threshold)	3.5		
Map sharpening B-factor ($Å^2$)	-143		
Small subunit without head			
Map average resolution (Å, 0.143 FSC threshold)	3.7		
Map sharpening B-factor ($Å^2$)	-165		
Head of small subunit			
Map average resolution (Å, 0.143 FSC threshold)	4.8		
Map sharpening B-factor ($Å^2$)	-224		
Refinement			
Initial model (PDB code)	4v6w	Testis 80S model	Testis 80S model
Model Composition			
Non-hydrogen atoms	219,859	217,079	219,021
Amino acid residues	12,106	11,755	11,764
Nucleotides	5,916	5,916	6,003
R.M.S.D. from ideal geometry			
Bond lengths (Å)	0.008	0.011	0.017
Bond angles (º)	1.027	1.315	1.794
Validation			
Clashscore	10.57	7.8	15.56
Rotamer outliers (%)	0.82	1.23	1.90
Ramachandran plot statistics			
Favored (%)	86.50	83.81	82.22
Allowed (%)	13.00	15.31	17.02
Outliers (%)	0.50	0.88	0.76

Sup Table 1: Summary of data collection, image processing, model building, refinement and validation statistics

Figure 5

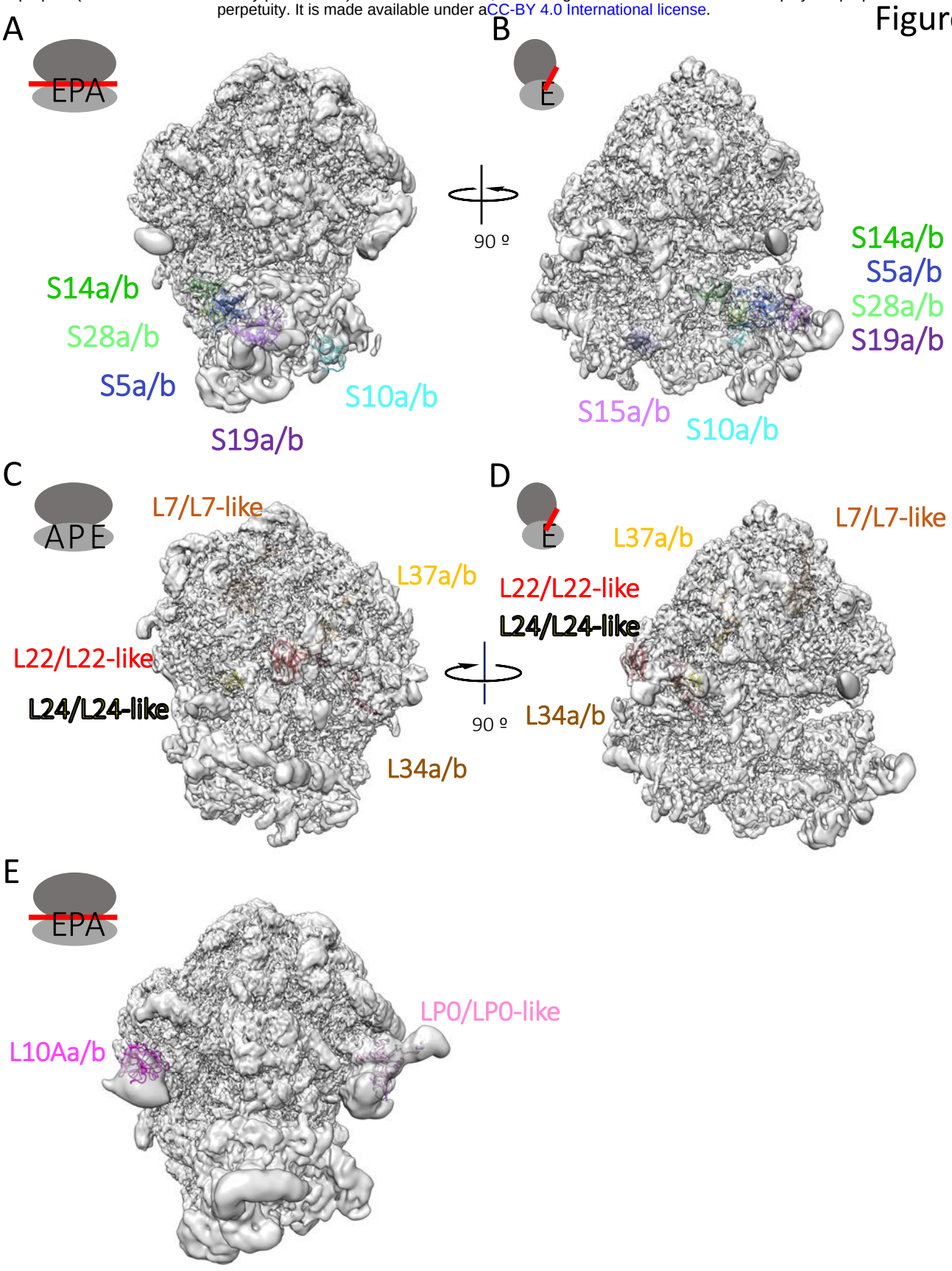


Figure 5: Location of *D. melanogaster* ribosomal paralogs

Ribosomal paralogs were mapped to the testis 80S EM average. (A & B) Small subunit paralogs: RpS14a/b (dark green), RpS28a/b (light green), RpS5a/b (dark blue), RpS19a/b (dark purple), RpS10a/b (cyan) and S15a/b (light purple). Paralogs are shown viewed from the front of ribosome (A) and from the side into mRNA channel (B). (C & D) Large subunit paralogs: L7/L7-like (light brown), L22/L22-like (red), L24/L24-like (yellow), L34a/b (dark brown) and L37a/b (orange). Paralogs are viewed from back of the ribosome (C) and from the side into mRNA channel (D). E) Paralogs that locate in ribosome stalks: L10Aa/b (dark pink) and LP0/LP0-like (light pink). Paralogs are shown viewed from the front of ribosome.

		Testis 80S			Ovary 80S			
		Paralog in Testis	# aac included in refinement	Resolution (Å)	Paralog in ovary	# aac included in refinement	% identity	RMSD across all atom pairs (Å)
Large subunit	L7/L7-like	L7	226	3.5	L7	226	100	0.481
	L37a/b	L37a	87	3	L37a	87	100	0.514
	L34a/b	L34b	103	3.5	L34b	103	100	0.530
	L24/L24-like	L24	58	3.5	L24	60	100	0.525
	L22/L22-like	**L22-like**	**96**	**4.5**	**L22**	**99**	**57**	**3.107**
Small subunit	S14a/b	S14b	127	5	S14b	127	100	0.797
	S5a/b	**S5a**	**189**	**5.5**	**S5b**	**189**	**90**	**1.805**
	S28a/b	S28b	62	5.5	S28b	62	100	1.990
	S19a/b	**S19b**	**132**	**6**	**S19a**	**126**	**69**	**2.512**

Table 1: Atomic model paralog comparison. Switched paralogs are highlighted in bold.

Figure 6

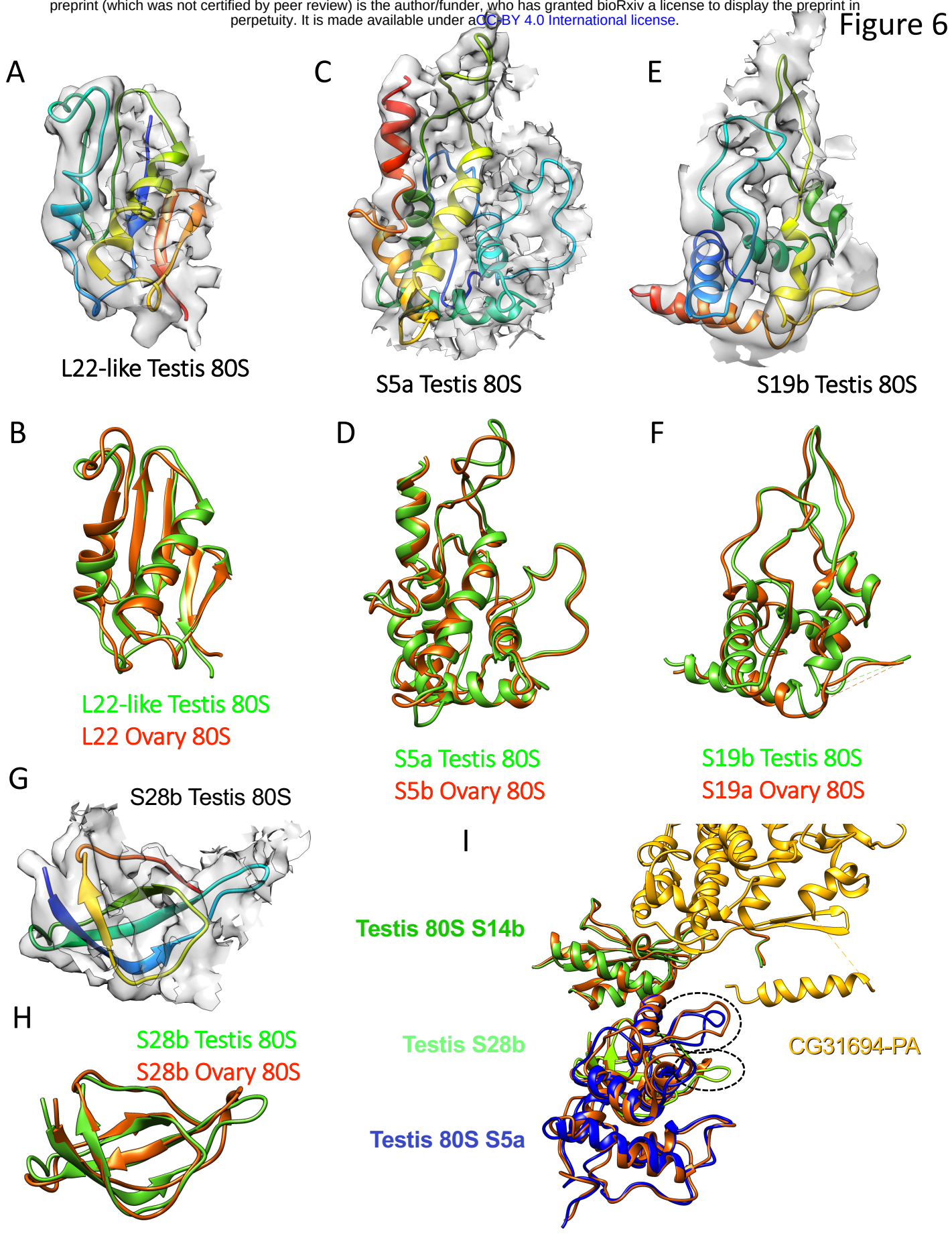
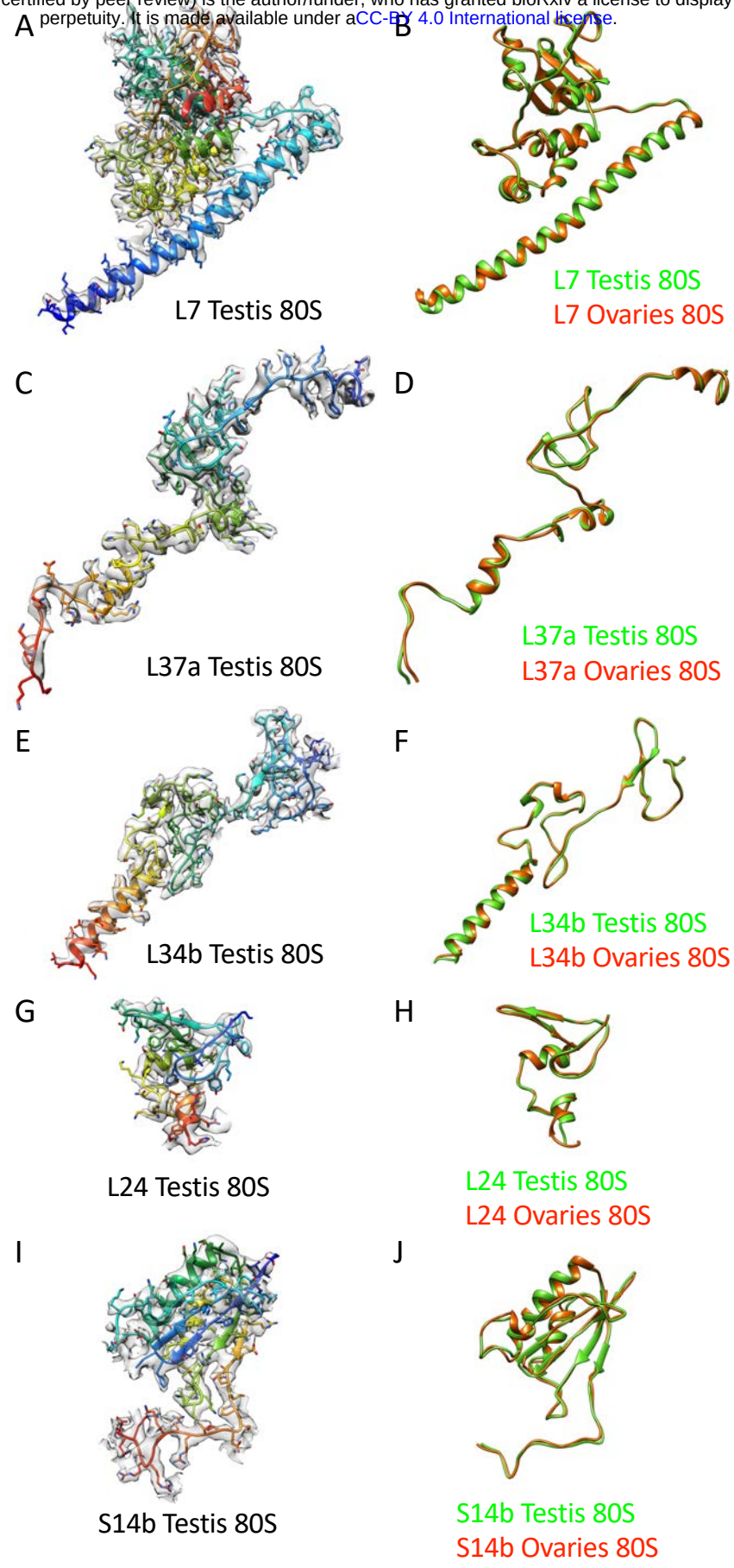


Figure 6: Structural implications of paralog switching events

Switched paralogs in testis 80S vs ovary 80S are shown. (A & B) L22-like (testis 80S) and L22 (ovary 80S). (C & D) S5a (testis 80S) and S5b (ovary 80S). (E & F) S19b (testis 80S) and S19a (ovary 80S). (G & H) S28b (testis 80S) and S28b (ovary 80S). (A, C, E & G) show the testis atomic model fitted into the EM density. The models are rainbow colored from n-terminus (blue) to c-terminus (red). (B, D, F & H) show the comparison between the testis 80S (green) and the ovary 80S (red) atomic models. (I) Area around mRNA channel, which in testis 80S is occupied by an alpha-helix from CG31694-PA. S14b (dark green), S28b (light green) and S5a (blue) from testis 80S are nearby. Ovary 80S paralogs (S14b, S5b and S28b) are superimposed, in red. The main differences between the pdb models, circled, are in regions close to CG31694-PA.



Sup 6: Structural implications of paralog switching events

Non-switched paralogs in Testis 80S vs Ovaries 80S are shown. (A & B) L7; (C & D) L37a; (E & F) L34b; (G & H), L24; (I & J) S14b. (A, C, E, G & I) show the testis atomic model fitted into the EM density. The models are rainbow colored from n-terminus (blue) to c-terminus (red). (B, D, F, H & J) show the comparison between the testis 80S (green) and the ovary 80S (red) atomic models.

Table 2

Paralog 1	Aa length	Paralog 2	Aa length	Amino acid Identity	Specialisation	Human ortholog	Human Disease
RpL22	299	RpL22-like	312	45%	Testis RpL22-like	RpL22 RpL22L1	Cancer/blood disease
RpL37a	93	RpL37b	89	75%	Testis RpL37b	RpL37	
RpS19a	160	RpS19b	159	66%	Testis RpL19b	RpS19	Diamond-Blackfan anaemia. Cancer.
RpS28a	64	RpS28b	65	82%	Testis	RpS28	Diamond-Blackfan anaemia.
RpS10a	163	RpS10b	160	61%	Testis	RpS10	Diamond-Blackfan anaemia. Aase syndrome
RpS5a	228	RpS5b	230	76%	Ovary>Embryo> Testis	RpS5	Diamond-Blackfan anaemia. Cancer.
RpL7	252	RpL7-like	257	28%	Ovary 80S	RpL7	Cancer
RpL24	155	RpL24-like	191	24%	Ovary 80S	RpL24	Tropoblast development

Table 2: Summary of paralog pair attributes
For each paralog pair, amino acid identity, ribosome of specialization, phenotype in *D. melanogaster* mutants, relationship to humans and human diseases.

Figure 7

A

```
CLUSTAL O(1.2.4) multiple sequence alignment

RpL22_dm          MAPTA-KTNKGDTKTAAAKPAEKKAAPAAAAAKGKVEKPKAEAAKP---------AAAAA   50
RpL22-like_dm     MSSQTQKKNASKAK-SKAGPQPSKKAPAKVT---PVETATPSAAPPLSSKKVAKAPAVAL   56
                  *:  : *.* ..:* : * *  .* *** .:    **. . .** *          *.*

RpL22_dm          KNVKKASEAAKDVKAAAAAAAKPAAAKPAAAKPAAASKDAGKKAPAAAAPKKDAKAAAAPA  110
RpL22-like_dm     KNLDLAMKESAK-KAAEAADQKLVS--KAKKGQNASRKNLMQVPEPVAPEVDNKKKSAP-  112
                  **:. * : : . *** ** :  .:   * *   **:.   :.*  .**: * *  :**

RpL22_dm          PAKAAPAKKAASTPAAAPPAKKAAPAKA-------AAPA-AAAPAPAAAAPAVAKPAPKP  162
RpL22-like_dm     -----PKRGNAEIPTVV-PAKKAALVQEKPKEQAAAAPKAENVEVPAAKAKSVKKS---L  163
                       * :  *. *:.. ****** .:         ***    . .*** * :* *

RpL22_dm          KAKAAPAPSKVVKKNVLRGKGQKKKKVSLRFTIDCTNIAEDSIMDVADFEKYIKARLKVN  222
RpL22-like_dm     ATPANPKPKPKRTKNVLRGKRLAKKKAWQRFVIDCACVAEDMILDLADFEQYLKTHIKIK  223
                   : * * *.   .*******   ***.  **.***: :*** *:*:****:*:*:::*::

RpL22_dm          GKVNNLGNNVTFERSK-LKLIVSSDVHFSKAYLKYLTKKYLKKNSLRDWIRVVANEKDSY  281
RpL22-like_dm     NKLNQLKDQVTFERTKNFSLIIHSGVHFSKRYFKYLTKRYLKKVSLRDWLRVVSTAKDTF  283
                  .*:*:* ::******:* :.**: *.***** *:*****:**** *****:***:. **::

RpL22_dm          ELRYFRISSNDDEDDDAE-----------   299
RpL22-like_dm     AMTYFKIQDKDFDDDDDDVADDNGGKTFA   312
                   : **:*..:* :*** :
```

B

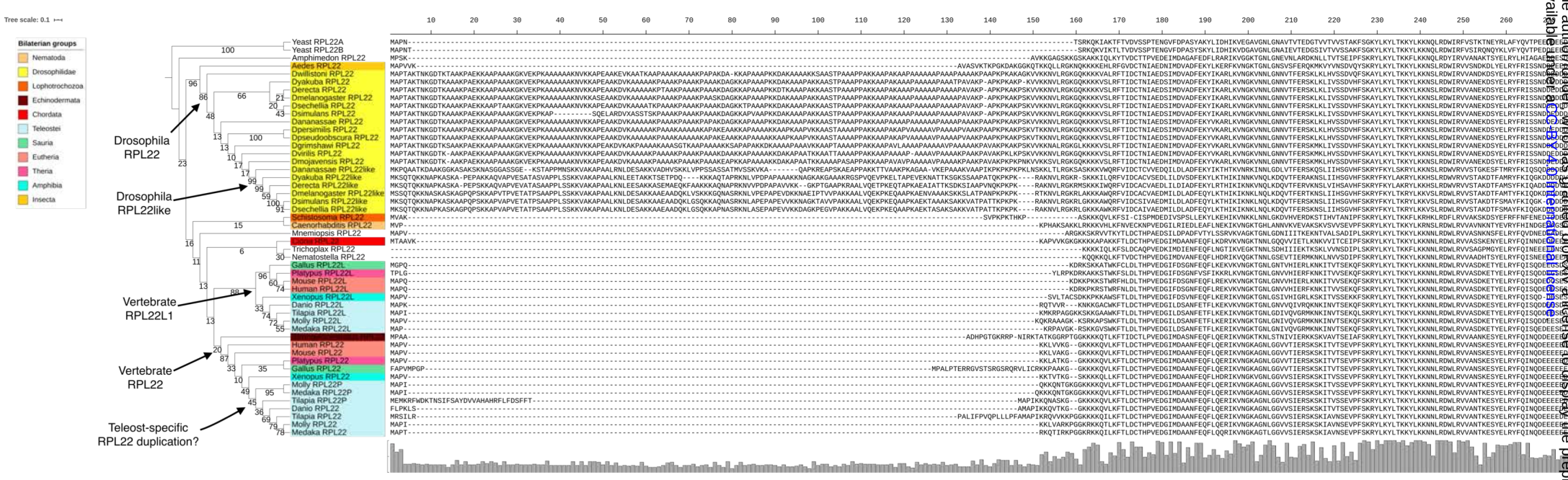
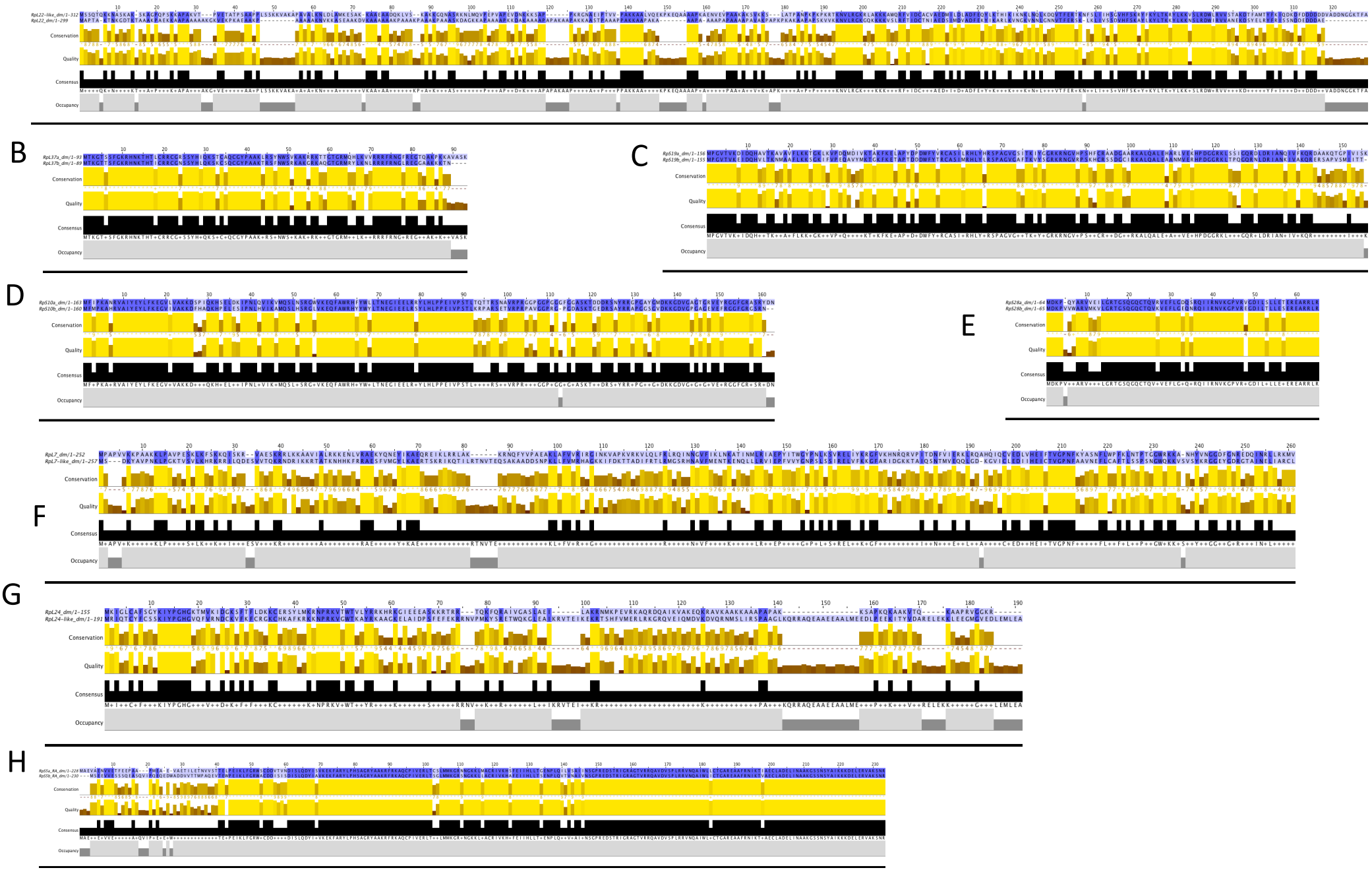


Figure 7: Potential changes to ribosome

A) Alignment of RpL22 and RpL22-like amino acid sequences for the two *D.melanogaster* proteins. In bold, amino acids included in the atomic models. B) Phylogenetic tree for RpL22 and RpL22-like paralogs across a range of animal genomes. The three duplication events detected are indicated.



Sup 7: Potential changes to ribosome

Clustal-omega alignment of paralog pairs' protein sequences; A) RpL22/22-like, B) RpL37a/b, C) RpS19a/b, D) RpS10a/b, E) RpS28a/b, F) RpL7/7-like , G) RpL24/24-like, H) RpS5a/b