Quantifying post-transcriptional regulation in the development of *Drosophila melanogaster*

Kolja Becker¹, Alina Bluhm¹, Nuria Casas-Vila¹, Nadja Dinges¹, Mario Dejung¹, Sergi Sayols¹, Clemens Kreutz², Jean-Yves Roignant¹, Falk Butter¹ & Stefan Legewie¹

Even though proteins are produced from mRNA, the correlation between mRNA levels and protein abundances is moderate in most studies, occasionally attributed to complex post-transcriptional regulation. To address this, we generate a paired transcriptome/proteome time course dataset with 14 time points during *Drosophila* embryogenesis. Despite a limited mRNA-protein correlation ($\rho = 0.54$), mathematical models describing protein translation and degradation explain 84% of protein time-courses based on the measured mRNA dynamics without assuming complex post transcriptional regulation, and allow for classification of most proteins into four distinct regulatory scenarios. By performing an in-depth characterization of the putatively post-transcriptionally regulated genes, we postulate that the RNA-binding protein Hrb98DE is involved in post-transcriptional control of sugar metabolism in early embryogenesis and partially validate this hypothesis using Hrb98DE knockdown. In summary, we present a systems biology framework for the identification of post-transcriptional gene regulation from large-scale, time-resolved transcriptome and proteome data.
According to the central dogma of molecular biology, protein is translated from mRNA, suggesting that mRNA levels can be predictive of protein concentrations. However, the relationship between mRNA and the concentration of its protein does not follow the simple monotonic correlation that higher mRNA levels always relate to concordantly more protein. This nontrivial mRNA–protein relation is a general phenomenon ranging from yeast to human (reviewed for example in refs. 1–3), that could arise from extensive post-transcriptional gene regulation in eukaryotic organisms. One important mechanism of post-transcriptional gene regulation is controlled protein translation. In line with widespread translational regulation, global studies in mammalian cells showed very different translation rates across mRNAs. Accordingly, ribosome occupancy of a transcript is a better predictor of protein expression when compared to its mRNA concentration. Furthermore, proteins are subject to active degradation via the Ubiquitin-proteasome system, controlling protein degradation rates independently of its transcript abundance.

Even in the absence of regulated protein translation or turnover, a mismatch between mRNA and protein levels can occur, due to the temporal delay that occurs when protein is translated from mRNA. Accordingly, it has been noted that while mRNA–protein correlations are generally low, the discrepancy is even more pronounced during dynamical cellular transitions. Theoretical considerations revealed that delayed protein dynamics relative to mRNA results in a nonlinear relationship between the two species, and is especially pronounced for stable proteins with long half-lives. Furthermore, during cellular transitions active regulation of protein translation and turnover is enforced. For instance, RNA binding proteins, micro-RNAs, or RNA modifications can affect processing, translation, and/or turnover of a transcript.

Drosophila embryogenesis poses an interesting case to study post-transcriptional regulation, since almost no transcription takes place during the first 3 h after oocyte fertilization. Instead, developmental dynamics rely on post-transcriptional regulation of maternally deposited mRNA and protein until the transcription of embryonic genes is switched on during maternal-to-zygotic transition (MZT). For example, it is well established that the maternally deposited positional information genes hunchback, caudal, nanos, and oskar are regulated post-transcriptionally. Global measurements of ribosome occupancy at various time points demonstrated that post-transcriptional regulation is a widespread phenomenon in the developing embryo. Given that RNA-binding proteins (RBPs) are major regulators of post-transcriptional gene regulation, it is not surprising that the mRNA-bound proteome is highly dynamic during MZT.

These complex dynamics of post-transcriptional regulation make it difficult to intuitively understand the relations between mRNA and protein expression changes. Mechanistic models of protein translation provide a possible mathematical framework for an improved understanding of gene expression regulation. For instance, using simple models of protein translation describing the spatio-temporal mRNA–protein relationship of three Drosophila gap genes, we concluded that protein abundance merely represents a time-delayed version of its corresponding mRNA not requiring post-transcriptional regulation. On a genome-wide level, Teo et al. and Cheng et al. developed a time-discrete model of protein turnover and directly inferred post-transcriptional regulation from characteristic features of protein and mRNA time courses. Applying kinetic, time-continuous, models of protein translation, Peshkin et al. concluded for Xenopus laevis development that the majority of protein expression changes can be explained by assuming proportional protein production from mRNA and first-order protein degradation. Likewise, in dendritic cells responding to external lipopolysaccharide treatment, Jovanovic et al. conclude that simple models of protein translation explain most of the variance in protein expression based on mRNA abundance changes.

In this work, we investigate the relation between mRNA and protein levels during Drosophila embryogenesis using highly time-resolved paired transcriptome/proteome measurements. Compared to the previous modeling studies, we specifically describe a framework for the systematic discovery of post-transcriptional regulation mechanisms controlling a biological process of interest: We combine model fitting and model rejection to explicitly name mRNA–protein pairs insufficiently described by four distinct model variants, thereby deriving lists of potentially post-transcriptionally regulated proteins. We show that post-transcriptionally regulated proteins are enriched for certain biological functions, including cell cycle and regulation of glucose metabolism. Furthermore, based on sequence motif analyses and knockdown experiments, we propose that the splicing factor Hrb98DE may be involved in post-transcriptional control of glucose metabolism.

Results

Paired mRNA/protein data reduces experimental variation. We previously followed proteome changes during Drosophila development with high temporal resolution. This dataset included measurement time points every 1 h within the first 6 h after egg-laying, and every 2 h hereafter until 20 h (Fig. 1a). When comparing our proteome dataset to a published developmental RNA-Seq time course, we observed limited correlation between RNA and protein levels. To further investigate post-translational gene regulation in more detail, and to exclude that the discrepancies between RNA and protein levels arose from experimental variation between laboratories, Drosophila strains, and measured time points, we generated a new RNA-Seq dataset using the same fly embryo samples as for our proteome measurement (Fig. 1a and Supplementary Figure 1a).

Our transcriptome matches the published RNA-Seq dataset at similar stages of development when calculating pairwise correlations of all common reads between samples (Supplementary Figure 1b). Time-course analysis however showed altered developmental speed between our data and the published dataset, indicating that laboratory conditions can affect the overall kinetics of gene expression even for a robust biological process as Drosophila embryonic development. Thus, pairing of RNA and proteome measurements from the same laboratory avoids the risk of systematic mis-estimations.

Proteome and transcriptome changes show limited correlations. We related transcriptome and proteome measurements to explore whether both gene expression layers exhibit a high degree of coordination during Drosophila embryogenesis. We based our analysis on median values across four replicates and focused on 3761 RNA–protein pairs detected reproducibly in at least ten time points (Supplementary Data 1).

We classified the temporal behavior of these 3761 mRNA–protein pairs into four groups with qualitatively distinct dynamics using a hierarchical clustering approach (see Fig. 1b). These four groups are: (1) mRNA and protein levels increase concordantly (525 genes), (2) mRNA and protein levels decrease concordantly (1063 genes), (3) mRNA increases, while protein levels decrease (372 genes) and (4) mRNA decreases, while protein levels increase (1801 genes). In total, 58% of genes showed inverse abundance changes between mRNA and protein,
indicating that the temporal expression dynamics of most mRNA transcripts and the corresponding proteins are not directly related.

To quantify the relationship between mRNA and protein abundance, we calculated Spearman correlation coefficients globally, relating mRNA and protein levels across all genes at the same time point. We included only proteins showing the largest fold-changes compared to 0 h \((n = 500)\) to avoid that random variations mask existing trends between mRNA and protein. In line with previous studies, the average correlation between mRNA and protein levels of the same time point is limited \((\rho = 0.41)\) (Fig. 2a, diagonal) with a maximum at 14 h \((\rho = 0.57)\) (Fig. 2b—green panel). Of note, this limited correlation is not due to experimental variation, as the mean correlation across four biological replicates at 14 h is \(\rho = 0.96\) for the transcriptome and \(\rho = 0.96\) for the proteome.

To account for time delays associated with mRNA processing and translation, we also assessed nonsynchronous correlations by globally relating mRNA and protein levels at different time points. In this scenario, the highest mRNA−protein correlation reaches \(\rho = 0.63\) between mRNA levels at 12 h and protein abundances at 16 h (Fig. 2b−orange panel). Despite this modest value, we observed a general trend of better transcriptome/proteome correlation when relating mRNA samples to later protein time points, most likely due to delays in protein synthesis. Furthermore, higher correlations are generally observed at later developmental time points, possibly because less pronounced dynamical changes of mRNA and protein occur after MZT.

To further test for the correspondence of mRNA and protein dynamics, we calculated the correlation between the mRNA and protein time courses for each gene individually. Only a subset of mRNA–protein pairs \((n = 429 (11.4\%))\) exhibits a significant positive Spearman correlation coefficient (Fig. 2c—bottom panel, Student’s \(t\) test, two-sided, \(p<0.05\)), and the median correlation coefficient over all genes is close to zero (Fig. 2c—top panel). To account for protein expression delays, we additionally calculated the same correlation coefficients after introducing a time shift between mRNA and protein measurements. These shifts only marginally improved the median correlation coefficient over all genes, with a maximum median correlation when the protein was assumed to lag behind the mRNA for 4−6 h (Fig. 2c—top panel). At a time shift of 4 h, we also observed the highest number \((n = 520 (13.8\%))\) of mRNA−protein pairs with significant positive correlation (Fig. 2c—lower panel, Student’s \(t\) test, two-
Kinetic models quantitatively relate mRNA/protein dynamics. We used a set of dynamical models describing protein expression based on ordinary differential equations (ODEs) and investigated whether the dynamics of each protein can be explained as a function of the corresponding mRNA time course. By fitting these models to the experimentally measured protein expression time courses, we assigned each mRNA—protein pair to a model representing one of four regulatory scenarios described below (schematically depicted in Fig. 3a), or excluded it from any of these.

The default model (production) follows the simple assumption that protein is synthesized from mRNA and additionally subject to degradation. The mRNA concentration serves as an input and proportionally affects the translation rate, implying the absence of
of the four different model variants was protein accumulation prior to the MZT (see Fig. 1b). Addition-
of the delayed-production model was motivated by very low model corresponds to the production model. The implementation translation sets in, from which point the delayed-production translation. In this model, the protein is initially degraded until delayed-production model, which assumes a temporal delay in complex post-transcriptional regulation. We also considered a delayed-production model, which assumes a temporal delay in translation. In this model, the protein is initially degraded until translation sets in, from which point the delayed-production model corresponds to the production model. The implementation of the delayed-production model was motivated by very low protein accumulation prior to the MZT (see Fig. 1b). Additionally, we assumed two mathematically less complex models: In the degradation model de novo translation can be neglected throughout embryonic development and therefore protein synthesis was omitted. In the stationary model, the protein is assumed to be stable during embryogenesis thereby eliminating protein production and degradation altogether.

The unknown model parameters such as protein synthesis and degradation rates, initial protein concentrations, and delay times were determined by fitting our models to the experimental data. We fitted each of the 3761 mRNA—protein pairs individually using the four model variants, and classified the genes into one of the regulatory scenarios (stationary, degradation, production and delayed-production) using a two-step strategy: First, we assessed whether a given model describes the measured mRNA—protein dynamics based on the difference between model fit and data using a \( \chi^2 \)-test (Benjamini–Hochberg (BH) corrected \( p \) value < 0.05). To exclude that deviations between model and data are correlated in time, we further applied a Durbin–Watson test (BH corrected \( p \) value < 0.05), and only considered a model variant feasible if both tests were passed. Second, for mRNA/protein pairs fitting multiple models, we performed a stepwise likelihood-ratio test, balancing the goodness-of-fit against the risk of overfitting, to select for the simplest model explaining the data. If the model selection assigned a protein to the degradation or stationary classes, but the mathematically more complex production model could not be rejected with a nonzero protein translation rate, we re-assigned the protein to the production class. Thus, we consider the production model as the simplest, unregulated protein expression scenario in biological terms, whereas the competing (mathematically simpler) models require the existence of an additional biological factor blocking translation or degradation. While Fig. 3a shows examples of classifications, see Supplementary Note 3 for a detailed description of the classification procedure.

Slightly more than half (54%) of mRNA—protein expression patterns follow the production model, which assumes continuous protein synthesis from mRNA, and are thus likely determined by
pure transcriptional control (Fig. 3b). Furthermore, 5% of proteins were stable during embryogenesis (stationary model), and another 3% of proteins were best fitted by the degradation model. Finally, the assumption of a delay time for translation (delayed-production model) explained another 21% of genes. Altogether, 84% of genes were explained by any of these four simple regulatory models, suggesting that simple ODE models quantitatively describe proteome dynamics by including transcriptome data. For each of the four models the Pearson correlation between fitted and measured protein expression values across all genes and time points is above 0.99 (Supplementary Figure 3).

For the remaining 16% of protein–mRNA pairs all four proposed models were rejected, suggesting that these proteins are under complex post-transcriptional control. Accordingly, we find a statistically significant overlap (1.85-fold enrichment, hypergeometric test, \( p = 2.3e-16 \), Supplementary Figure 4a) with a published set of translationally regulated developmental genes identified by ribosome profiling \(^{21} \). Potentially post-transcriptionally regulated proteins are found in all four dynamical groups obtained by hierarchical time-course clustering (Supplementary Figure 4b). This demonstrates the complementarity of our modeling analysis and suggests that post-transcriptional regulation occurs by various mechanisms rather than being controlled by a global factor. Interestingly, previous proteomic studies report similar fractions of 20–30% of potentially post-transcriptionally regulated proteins for other biological processes (see Discussion). The fact that we obtain slightly smaller numbers of post-transcriptionally regulated genes (16%) most likely reflects that our approach produces a conservative estimate as we show using in silico benchmarking (see Supplementary Note 1).

During the first 3 h until MZT, no transcription occurs in the fly embryo \( ^{18} \). Thus, any protein must be maternally deposited or translated from maternally deposited RNA. MZT is visible in our data as striking changes in mRNA as well as protein expression patterns at 3 h (Fig. 1b). Since MZT marks the advent of transcriptional activity, we expected a large fraction of post-transcriptional regulation specifically within the first 3 h of development. To test this hypothesis, we repeated the protein classification including only data from time points beyond MZT (3–20 h). In line with a strong decline of post-transcriptional control after MZT, we found that for 98% of proteins the post-MZT data can be explained by one of the four ODE model variants (Fig. 3b, right).

Among these 98%, only 3% of proteins (initially 21%) are still assigned to the delayed-production model, i.e. only few proteins show delayed translation post-MZT. This observation again supports that MZT is a major point of post-transcriptional control at which mRNA translation may be activated on demand \( ^{31} \). In line with this hypothesis, estimated delay times in the delayed-production model fitted to the full dataset (0–20 h) show a bimodal distribution with a group of proteins exhibiting delays of 0.5–2.5 h (381 proteins) and another group with translation delayed even longer for about 3–10 h (452 proteins) (Fig. 3c). This indicates two waves of translation: a first translation burst coinciding with MZT and a secondary delayed translation phase. The general ability of our classification method to distinguish between different mRNA/protein dynamics is shown in Supplementary Figure 4c.

**Kinetic models account for lack of mRNA–protein correlation.**

Given that 84% of mRNA–protein pairs are described without assuming complex post-transcriptional regulation, the low mRNA–protein correlation might be surprising. The proposed mathematical models provide mechanistic insights into the missing correlation: For the stationary, degradation, and delayed-production classes, the model predicts simple biological control mechanisms on protein translation, leading to an uncoupling of mRNA–protein dynamics.

However, even for the production model, where such control mechanisms are absent, mRNA and protein time courses are not necessarily correlated, as evidenced by a broad distribution of the corresponding correlation coefficients (Fig. 4a, top panel). The lacking correlation seems to arise, in part, because proteins with long half-life lag behind corresponding mRNA \( ^{32} \), hereby decreasing mRNA–protein correlation \( ^{33} \). Based on our fitting results, we obtained estimates for the kinetic parameters and indeed find that the lack of positive mRNA–protein correlation (\( \rho = 0.03 \)) is particularly evident when considering only stable proteins whose half-life confidence interval exceeds the median (8.4 h) of all protein half-lives (260 proteins; \( \rho = 0.55 \), Fig. 4a, second panel). In contrast, positive correlations are much more common in the inversely defined group of proteins with short half-lives (331 proteins, \( \rho = 0.24 \), Fig. 4a, third panel), supporting that protein degradation rates determine how closely protein dynamics follow the mRNA time course.

However, our analysis indicates that even proteins with short half-lives can show a weak mRNA–protein correlation (Fig. 4a, third panel, short half-life). Our model explains many of these mRNA–protein discrepancies with out-of-steady-state protein concentrations at the initial time-point (\( t = 0 \) h): Certain mRNAs deposited in the egg remain untranslated in the inactive egg and are only translated upon fertilization \( ^{34} \). Here, initial protein production at \( t = 0 \) exceeds protein degradation, leading to a net increase in protein level even at constant mRNA amounts. In an extreme case, this can result in an inverse protein–mRNA relationship during the developmental time course, where the protein level increases, while mRNA abundance decreases (Fig. 4b, left). Likewise, mRNA increase with a concomitant protein decay can occur if net protein degradation exceeds synthesis (Fig. 4b, right).

To test whether a large fraction of proteins is out of steady-state at egg deposition (\( t = 0 \) h), we compared the actual protein expression values with the theoretical steady-state predicted by the ratio of the fitted synthesis and degradation rates (see Fig. 4c). As expected, the correlation of measured protein levels with the theoretical steady-state is limited at fertilization, whereas they agree better later in development. Such a perturbation from steady-state indeed weakens the mRNA–protein correlation: If we consider only proteins with fast dynamics (short half-life), whose abundance is close to their estimated steady-state at 0 h (101 proteins), the median correlation between individual mRNA–protein pairs significantly increases to \( \rho = 0.45 \) (Fig. 4a, bottom panel) compared to \( \rho = 0.24 \) for all fast changing proteins (two-sided Kolmogorov–Smirnov test, \( p = 0.003 \)). In a dynamical context, a prerequisite for a high correlation between mRNA and protein is therefore not only fast protein turnover, but also an appropriate initial mRNA–protein ratio with balanced production and degradation rates.

**Regulatory classes show enrichment for biological functions.**

We performed Gene Ontology (GO) analyses to investigate whether certain biological functions are specifically represented in each class (Fig. 5). Proteins assigned to the production model showed enrichment for processes related to mRNA processing, splicing and RNA metabolism, suggesting that this class contains critical regulators of post-transcriptional gene expression. Moreover, proteins in the delay group are enriched for GO terms related to protein catabolic processes, possibly indicating
widespread expression of factors mediating the degradation of maternal protein at the onset of MZT. Among post-transcriptionally regulated proteins for which all of our four models needed to be rejected we found an enrichment of cell cycle-related genes. Indeed, early nuclear divisions in Drosophila embryos cannot be controlled transcriptionally, since transcription is virtually absent before MZT. Accordingly, widespread post-transcriptional regulation has been reported for maternal mRNAs involved in cell cycle regulation. In addition, GO-terms related to sugar metabolism are enriched in the group of potentially post-transcriptionally regulated proteins, which agrees with previous evidence showing post-transcriptional control of genes functioning in glucose metabolism.

The mode and time scale of post-transcriptional regulation, however, appears to be distinct for cell cycle and sugar metabolism: Protein levels of genes related to sugar metabolism predominantly upregulate 3 h after egg-laying and afterwards remain nearly stable, while their mRNAs sharply decrease already within 2 h after egg-laying (Supplementary Figure 5a). This suggests early post-transcriptional regulation of sugar metabolic processes before or at MZT. In contrast, a subset of proteins related to the cell cycle shows an abrupt downregulation of both mRNA as well as protein at 14 h, potentially indicating (post-)transcriptional regulation long after the completion of MZT (Supplementary Figure 5b).

**Identifying mechanisms of post-transcriptional regulation.** To uncover potential regulators of post-transcriptional gene regulation, we searched for enriched sequence motifs in the mRNA sequence of proteins in our protein classes. We considered Drosophila-specific RNA–binding protein (RBP) motifs corresponding to 51 different RBPs. We found no or little enrichment of RBP motifs in stationary, degradation, and delayed-production classes, whereas seven RBPs showed significant motif enrichment for the class of potentially post-transcriptionally regulated proteins (Fig. 6a).

Among the enriched motifs were the RBPs Pumilio and Bruno (also known as Aret), which are known post-transcriptional regulators of positional information genes in early Drosophila development. The strongest motif enrichment was observed for Hrb87F and Hrb98DE (also known as Hrp36 and Hrp38), two proteins recognizing highly similar RNA sequence motifs (Fig. 6a). In our time course, Hrb87F and Hrb98DE both show a ~5-fold upregulation in protein expression (Fig. 6b), indicating that they are developmentally regulated. Accordingly, a recent RNA interactome study demonstrated that the mRNA-bound fraction of Hrb87F and Hrb98DE changes during MZT in Drosophila. Interestingly, the known functions of Hrb87F and Hrb98DE match with results of our GO term analysis, as these proteins have been identified in large-scale screens for regulators of the cell cycle and sugar metabolism, respectively.

**Hrb98DE-dependent regulation of glucose metabolism.** For further analysis, we focused on Hrb98DE, because this protein (and its vertebrate homolog hnRNP A1) has been implicated in the regulation of protein translation. To assess the role of
Fig. 5 Protein classes are enriched for specific biological functions. Distinct GO terms are significantly (hypergeometric test, BH corrected \( p \) value < 0.05) overrepresented in the production and delay category as well as for the rejected class (indicated by color). GO terms were arranged according to their semantic similarity and a 2D projection was generated via multidimensional scaling. Each circle represents a single GO term and the size of each circle is proportional to the corrected \( p \) value (see legend).

Fig. 6 Post-transcriptionally regulated proteins are enriched for RBP binding motifs. a Seven RBP sequence motifs are enriched in the group of potentially post-transcriptionally regulated proteins (adjusted enrichment \( p \) value < 0.05). In total we scanned for enrichment of 67 Drosophila-specific motifs corresponding to 51 distinct RBPs. For each RBP, only the sequence logo with highest enrichment is shown. b Putative post-transcriptional regulators are regulated at the protein expression level during Drosophila development. Measured protein profiles (black solid line—normalized LFQ) of four RBPs, alongside their corresponding mRNA expression (dashed gray line) and the best model fit (colored line) are presented. Error bars represent standard deviation of protein according to a linear error model. The expression level of the three remaining RBPs (RBP1-LIKE, ARET, and CG17838) were below the proteomics detection limit.
Hrb98DE in post-transcriptional gene regulation during development, we performed in vivo knockdown experiments with two different dsRNAs (Supplementary Figure 6a) and measured protein expression 7:45 h (29°C) after egg deposition. The first dsRNA, which reduced Hrb98DE protein by 42% (Supplementary Figure 6a), led to differential expression of 78 proteins of 2600 genes consistently detected in the embryonic time course and the knockdown (3.0%, Supplementary Figure 1b right). In support for our model-based classification approach, there exists a significant overlap between genes predicted to be post-transcriptionally regulated and differentially expressed genes upon knockdown (24 genes observed vs. 13.26 expected, hypergeometric test, p = 2.6e-3; Fig. 7a). This significant overlap was preserved in the knockdown with the second dsRNA, even though knockdown efficiency was lower (Supplementary Figure 6a-c), suggesting that Hrb98DE indeed regulates the predicted targets during Drosophila development.

Since assessing the role of Hrb98DE during embryogenesis remains challenging due to limited knockdown efficiency, we turned to Drosophila S2R+ cell culture to obtain more robust Hrb98DE depletion. Based on RNA-Seq and mass spectrometry data, we observed a strong depletion of Hrb98DE mRNA and protein by about 84 and 76%, respectively (Supplementary Figure 6d). Off-target effects could be excluded, as two different dsRNA constructs induced highly overlapping transcriptome changes (Supplementary Figure 6e). In the knockdown showing

![Image](image-url)
greater Hrb98DE depletion, we detected an overlapping set of 4406 genes at both transcriptome and proteome, out of which 231 are differentially expressed at the mRNA level upon Hrb98DE knockdown (Wald test—DEseq2, BH corrected $p$ value < 0.05, absolute fold-change > 30%), whereas 40 change significantly at the protein level (independent $t$ test two-tailed, $p$ value < 0.01, absolute fold-change > 50%) (Fig. 7b).

Interestingly, a large fraction of genes responding at the protein level (26 out of 40, 65%) do not change at the mRNA level (Fig. 7c), which agrees with published data showing that Hrb98DE binding to mRNA leads only to few instances of changed transcript levels. Thus, we hypothesize that Hrb98DE binds these 26 RNAs to affect protein translation, but not mRNA turnover. In line with Hrb98DE controlling sugar metabolism at the post-transcriptional level, three out of the 26 candidate genes (14-3-3-zeta, dorsal, domino) are known to affect glucose metabolism.

Hrb98DE has been described as a regulator of alternative splicing. To identify additional targets of Hrb98DE-dependent post-transcriptional regulation, we analyzed splicing changes upon knockdown. We focused on 3542 transcripts consistently detected in the S2R+ cell line and the embryonic time course. We found 287 differentially spliced genes upon Hrb98DE knockdown (Fig. 7d), which significantly overlap with the set of direct Hrb98DE targets, which either contain a sequence motif or are published RIP-Seq targets (1.93-fold increase observed vs. expected, hypergeometric test, $p$ value = 6.0e-24) (Fig. 7e). We further tested for enrichment of model-predicted post-transcriptionally regulated genes in the set of differentially spliced genes, but did not find a significant overlap. The differentially spliced genes, however, are significantly enriched for the GO-term “regulation of glucose metabolism” (18 genes, 3.37-fold enrichment observed vs. expected, hypergeometric test, $p$ = 2.6e-6), and 11 genes in this set have previously been identified as Hrb98DE targets. Notably, Hrb98DE-dependent alternative splicing of domino, one of the genes involved in glucose metabolism, led to a detectable isoform switch at the protein level (Supplementary Figure 6f).

Our analysis thus suggests post-transcriptional regulation of glucose metabolism by the RBP Hrb98DE, the majority of effects being visible at the level of pre-mRNA splicing.

**Discussion**

Early *Drosophila* development occurs in the absence of de novo transcription, suggesting extensive post-transcriptional gene regulation. In this study, we set out to quantify how mRNA and protein levels are dynamically coordinated, and combined our *Drosophila* embryogenesis proteome of 14 time points with its embryonic development and classification of translation using model organisms. This creates an extensive systems-level dataset to study translational and post-translational gene regulation in a highly resolved temporal fashion.

We expressed mRNA and protein concentrations as relative units after normalization by the total amount of mRNA and protein at each time point, as was done in other dynamic transcriptome studies analyzing early *Drosophila* development. This implicitly assumes that total mRNA and protein mass do not change significantly during development. The conclusion of a stable total protein expression across developmental stages is supported by a recent single-embryo proteome survey. Furthermore, a recent study used spike-in controls for absolute quantification of 68 mRNAs in developing *Drosophila* embryos. We quantitatively compared our relative RNA expression values to their corresponding absolute transcript counts and found a good agreement in time-dependent transcriptome changes (Supplementary Figure 7).

There is a recent interest in explaining the moderate mRNA–protein correlation. While it is appealing to attribute the lack of correlation to a widespread post-transcriptional regulatory network, correlation measures—although pragmatic—do not reflect the complex temporal connection between mRNA and protein dynamics. This is especially true if the correlation analysis is limited to the global mRNA–protein relationship at a single time point. Using mathematical models, we show that a low correlation of mRNA and protein time courses does not necessarily imply post-transcriptional regulation, but could be explained based on protein turnover rates or deviations from steady-state at the onset of development. Thus, the analysis of time-resolved mRNA and protein expression data using mechanistic mathematical models of translation is superior when compared to simple correlation analysis.

Few studies have quantitatively modeled time-course data to investigate the transcriptome/proteome relationship. Most of these studies employ protein expression models comparable to ours, but do not explicitly identify post-transcriptionally regulated genes. Using our combined model rejection and model selection approach, we determined post-transcriptional regulation during *Drosophila* embryogenesis for about 16% of expressed proteins. We establish further confidence in these candidates using a combination of in silico benchmarking and published evidence for post-transcriptional regulation provided by ribosome profiling. We find that post-transcriptionally regulated proteins are enriched for certain biological functions and identify potential mechanisms by RBP motif analysis.

Previous studies also reported comparatively low numbers of post-transcriptionally controlled genes: Investigating the response of dendritic cells to LPS stimulation, Jovanovic et al. can explain 79% of the protein variance based on a simple model of protein translation. They further concluded that post-transcriptional regulation is relevant predominantly for the regulation of absolute protein levels rather than fold-changes upon stimulation. Another study characterized early *Xenopus* embryonic development and classified proteins into four simple models of translation using model fitting and selection. Only few proteins are not classified into one of the four groups, suggesting limited importance of post-transcriptional regulation during early embryonic development. Two proteomic studies on mammalian circadian rhythmicity and on yeast amino acid starvation report fractions of 20–30% of potentially post-transcriptionally regulated proteins. This indicates that the low degree of post-transcriptional regulation may be conserved across organisms.

For *Drosophila* development, our analysis shows that most post-transcriptional control occurs within the first few hours after fertilization, i.e. before or during MZT. In addition, based on our models, we evidenced wide-spread delay of protein production until the onset of MZT. Translation-on-demand mechanisms have been postulated in yeast and neurons but to our knowledge not yet for metazoan embryonic development. Interestingly, the proteins with delayed production are enriched for GO terms related to protein catabolic processes, suggesting that induced protein translation may in turn trigger the degradation of maternal proteins at the onset of MZT.

Based on sequence motif analyses, we hypothesized that the RBP Hrb98DE is a post-transcriptional regulator of sugar metabolism during early embryogenesis. Indeed, a global remodeling of energy production can be observed during *Drosophila* development. Collectively, our results and previous work suggest that this remodeling is largely coordinated at the post-transcriptional level.
Overall, we provide a framework for investigating gene regulation in large-scale paired RNA–protein datasets, by which we identified genes under strong post-transcriptional control.

**Methods**

**Collection of embryos for RNA-Seq.** Population cages of wild-type Oregon R flies containing only fertilized females were maintained at 25 °C. Embryos were collected on standard agar apple juice plates in 30 min laying time windows and processed immediately (0 h time point) or aged at 25 °C for the required period (1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, 20 h). After collection, embryos were dechorionated using 7.5% hypochlorite for 2 min and rinsed with water. At this point, approximately 30% of the embryos (20 μl embryo pellets) were transferred to PBS buffer for lysis and mass spectrometry measurement. To check for correct and homogeneity of stages, approximately 10% of each sample was fixed and staged. The remaining samples were snap-frozen in liquid nitrogen and stored at −80 °C. Proteome extraction was performed previously as described in Casas-Vila et al.27. Total RNA was extracted from approximately 20 μl embryo pellets with the RNeasy Mini Kit (Qiagen) and RNA integrity checked by Bioanalyzer.

**In vivo knockdown experiments.** Females carrying one of two Hrb98DE dsRNA transgenes (obtained from Bloomington Drosophila Stock Center: # 33103 and 32351) were crossed with males carrying the actin-GAL4 driver for 3 days to allow mating. On the fourth day, flies were transferred to conical flasks connected with apple agar plates and females were allowed to lay eggs for 45 min at 29 °C. Embryos were developed for another 7 h at 29 °C and subsequently transferred to two tubes containing 1× PBS. For control, actin-GAL4 males were crossed with WT females and experimental flies were harvested and treated in parallel to control flies.

**Cell culture.** Experimental knockdown of Hrb98DE in S2R+ cells was performed using two distinct dsRNAs. Drosophila S2R+ cells were cultured at 25 °C in Schneider’s Drosophila Medium (GIBCO, Cat-No 21720) supplemented with 10% FBS and 2% penicillin/streptomycin. For knockdown experiments, dsRNA was synthesized overnight at 37 °C using the Hi-Scribe T7 kit (NEB, Cat-No-E2040). dsRNA was transfected in S2R+ cells by serum starvation for 6 h. The treatment was repeated twice and cells were harvested 5 days after the first treatment. Primer sequences to amplify dsRNA templates are listed in Supplementary Table 1.

**qRT-PCR.** Three micrograms total RNA was transcribed into cDNA using MMLV reverse transcriptase (Promega). qRT-PCR analysis was performed using a ViiA7 real-time PCR system (Applied Biosystems). Measurements were done in triplicates and relative RNA levels were normalized to rpl15 levels. qRT-PCR primer sequences are listed in Supplementary Table 1.

**Mass spectrometry measurement and label-free analysis.** For mass spectrometry measurement of Hrb98DE in vivo knockdown, experiments were performed in biological quadruplicates for each sample in PBS with 1× LDS buffer complemented with 0.1 M DTT. Samples were boiled for 10 min at 80 °C and proteins were separated on a 4–12% NuPAGE Bis/Tris gel for 10 min at 180 V in MOPS buffer. For mass spectrometry measurement of Hrb98DE knockdown in S2R+ cells, experiments were performed in biological quadruplicates. Cell pellets were lyzed in RIPA buffer (150 mM NaCl, 50 mM Tris–HCl pH 7.5, 0.1% sodium deoxycholate, 1% iegal CA-630) supplemented with protease inhibitor mix (Roche) for 30 min on ice and vortexed in between. After centrifugation at max speed for 10 min at 4 °C, the supernatant was recovered and protein concentration determined using Bradford. Sixty micromgram protein lysate was fractionated by SDS PAGE (three slices per sample).

**Time-course clustering.** From the mass spectrometry data, we selected 3761 proteins for further analysis based on their number of missing values (maximum four LFQ values below the detection limit out of 14 time points). Both mRNA and protein data including imputed values (see Casas-Vila et al.27) was normalized by the respective value at 0 h, log10 transformed, and scaled to minimum and maximum values between −1 and 1. A hierarchical clustering approach was applied to the processed data with imputed values: For this, pairwise distances between all mRNA–protein time courses were calculated using the Euclidean distance. Clusters were merged using a complete-linkage criterion.

**Correlation analysis.** Global Spearman correlation coefficients between mRNA (RPKM) and protein (LFQ, IBAQ) samples were calculated using only proteins showing largest absolute fold-changes compared to the 0 h time point (top 500). Values for which no protein measurement value exists, were not considered in this as well as the following analysis, i.e. no imputed values were taken into account. A linear correlation analysis was conducted using the Spearman correlation coefficient. Significance was determined using the permutation test for the matched pairs (Statistical Analysis System, SAS, Inc.).
matching mRNA at time $t_n$ with protein at time $t_n$, with $i \in Z, -3 \leq i \leq 5$. Unmatched time points were left out, leading to a minimum number of six paired values for the assessment of correlation. mRNA/protein correlations with a $p$ value < 0.05 ($p$ value returned by scpy.stats.spearmanr—Student’s $t$ test, two-sided) were chosen as significant.

Model fitting and evaluation. Before fitting, protein data were rescaled by its global mean value in all proteins to avoid numerical issues due to overly large LFQ values. On each time interval between two subsequent measured time points the model variants (stationary, degradation, production, delayed-production) can be solved analytically if the input mRNA time course on each interval is described using a linear function $y(t) = mt + b$. The explicit solution for the delay model

$$\frac{dy(t, \theta, u)}{dt} = ah(t - r)(mt + b) - ly(t, \theta, u)$$  \hspace{1cm} (1)

with initial value

$$y(t = 0) = y_0$$  \hspace{1cm} (2)

then becomes

$$y(t) = ah(t - r)(b\lambda^{-1} - m\lambda^{-2} + m\lambda^{-1}) + ce^{-\lambda t}$$  \hspace{1cm} (3)

with

$$c = y_0 - ah(-r)(b\lambda^{-1} - m\lambda^{-2}).$$  \hspace{1cm} (4)

In the above equation $y$ denotes measured protein, $a$ and $\lambda$ the production and degradation rate, respectively. The Heaviside function $h$ is used to suppress the product term in the delay model with $r$ being the time delay.

Model parameters $\theta = [y_0, a, \lambda, \tau]$ were estimated by minimizing the weighted least-squares distance between modeled values (Eq. 3) and nonimputed protein data using a trust region reflective optimization scheme (scipy.optimize.least_squares in Python). Missing protein values (or their imputed counterparts) were not considered in the cost function.

$$\chi^2(\theta, u) = \sum_{i} \frac{y_i(y, u)_{\text{model}} - y_i(\text{data})}{\sigma^2_i}$$  \hspace{1cm} (5)

Here $n$ denotes the number of data-points, $y$ modeled or measured protein expression, $\theta = [y_0, a, \lambda, \tau]$ the parameter vector, and $\sigma$ the weight for each data point. By relating mean and standard deviation of the four biological replicates, a linear error model was chosen to express $\sigma$ dependent on mean protein expression (slope, 0.169, intercept, 0.0). Ranges for the parameters are chosen as follows: $y_0 \in [1e-1, 5e3]$, $a \in [2e1, 1e-1]$, $\lambda \in [1e-5, 1e-1]$, $\tau \in [1e-5, 1e1]$. In all cases, multi-start local optimization was carried out via Latin-hypercube sampling and parameters were sampled on a logarithmic scale. Depending on the model complexity, a different number of initial parameter samples was chosen (degradation model: 5, production model: 5, delayed-production model: 25).

Model rejection was carried out by applying a $\chi^2$-test to the weighted squared residuals between model and data. To further exclude strong systematic deviations between model and data, the residuals were subjected to a Durbin-Watson (DW) test. Normality of model residuals was checked comparing values as Student’s $t$-test different model classes or Hrb98DE targets was assessed using a hypergeometric test. Significance was assessed using an independent $t$ test ($p < 0.05$). The overlap of genes uniquely identified in the embryonic time course as well as the in vivo knockdown were considered (2600 genes). Differentially expressed proteins were determined using an independent $t$ test ($p < 0.01$, two-tailed) and absolute fold-change > 2.0. Significant overlap between differentially expressed genes and different model classes or Hrb98DE targets was assessed using a hypergeometric test.

Gene set enrichment analysis. GO term enrichment for protein groups was carried out using Gorilla. As background the list of 3761 reproducibly measured proteins was chosen. Results were filtered using a corrected $p$ value of < 0.05 (Benjamini and Hochberg correction). Plotting was done using Multi-Dimensional Scaling implemented in the sklearn package provided for python based on a distance metric obtained using semantic similarity between GO terms. Semantic similarities were retrieved from the bioconductor package “GOSemSim” using the Wang distance metric.

Seventeen proteins with the following GO terms have been selected for Supplementary Figure 5a: “monosaccharide metabolic process” (GO:0005996), “glucose metabolic process” (GO:0006006), “monosaccharide biosynthetic process” (GO:0048364), “carbohydrate biosynthetic process” (GO:0016051), “hexose metabolic process” (GO:0019318), ”gluconeogenesis” (GO:0006094), “hexose biosynthetic process” (GO:0019319). Proteins (34) assigned the following GO terms are plotted in Supplementary Figure 5b: “mitotic spindle organization” (GO:0007052), “mitotic cell cycle process” (GO:1903047), “cell cycle process” (GO:0022402), “microtubule cytoskeleton organization involved in mitosis” (GO:1902830), “mitotic centrosome separation” (GO:0007100).

Motif analysis. Motif analysis was carried out using AME. Input sequences were defined as one of the following: the common transcript sequence between all mapped transcripts for each protein, the longest transcript sequence, longest 3’ UTR sequence, longest 5’ UTR sequence, or longest coding sequence. The motif database was taken from Ray et al., which included 67 motifs mapping to 51 Ray motifs. All AME motif scoring method was based on the fold motif scoring method was based on using FIMO. Motif match threshold was $p < 0.0002$ (Fisher’s exact). Background was estimated from the sequences of the 3761 reproducibly measured proteins. As control all sequences not in the group were chosen. Statistical test for enrichment/association was Fisher’s exact. The threshold values for reported results was set to an adjusted $p$ value of 0.05 (Benjamini and Hochberg correction). Individual motif occurrences were identified using FIMO with a 0.0 value threshold of $p < 0.0002$ (Fisher’s exact) for reported motifs. Multiple motif copies in the motif database were grouped by calculating the union over all transcripts.

Statistical analysis of Hrb98DE knockdown in vivo. Only proteins quantified in all measured samples (actin KD1 and actin KD2, each in biological triplicates) were chosen for statistical analysis of the in vivo knockdown. Additionally only the overlap of genes uniquely identified in the embryonic time course as well as the in vivo knockdown were considered (2600 genes). Differentially expressed proteins were determined using an independent $t$ test ($p < 0.01$, two-tailed) and absolute fold-change > 2.0. Significant overlap between differentially expressed genes and different model classes or Hrb98DE targets was assessed using a hypergeometric test.

Genes are considered targeted by Hrb98DE if identified as such by the study of Ji and Tulin, containing an Hrb98DE binding motif (section Motif analysis). Target genes identified in Ji and Tulin were kindly provided by Alexis Tulin. Target genes according to Blanchette et al. were determined re-analyzing the raw data according to the methods described in ref. 57, and using a fold-change cutoff of log2 > 10 between immunoprecipitation and control samples.

Statistical analysis of Hrb98DE knockdown in S2R + cells. Protein with NaN values in any of the tested conditions were filtered out before differential expression was assessed using an independent $t$ test for Hrb98DE knockdown and lacZ control. As thresholds for differentially expressed protein a $p$ value < 0.01 ($t$ test, two-tailed, $n = 4$) and fold-change cutoff > 50% was chosen. In the RNAseq data, nonexpressed mRNAs were removed based on a mean RPKM over all samples > 1. We obtained mRNA statistics using DESeq version 1.14.1 using the default Wald test to calculate the significance and applying automatic independent filtering to avoid testing genes which were poor candidates of being differentially expressed (maximizes the number of adjusted $p$ values less than alpha = 0.1). Differentially expressed mRNA were identified using a threshold of $< 0.05$ for BH corrected $p$ values ($n = 3$) and a fold-change cutoff of >30% change.
without discarding singletons. DEXSeq is a specialization of the NB-GLM approach and internally uses DESeq2 to test for differential exon usage, and we modeled our data with the formula “~ sample + exon + condition:exon”, looking for changes in exon abundance that could be explained by the KD (hence the interaction term in the formula). Finally, we considered that genes could show a different splicing pattern if they included exon abundance changes of at least 50% between conditions at FDR 1%. Duspars, DESeq and DEXSeq are part of the Bioconductor project. Genes associated with the GO-term “regulation of glucose metabolic process” (GO:010906) were retrieved from FlyMine v45.1 2017.

**References**

1. de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global signatures of protein and mRNA expression levels. Mol. Biol. Systems. [https://doi.org/10.1039/b903831d](https://doi.org/10.1039/b903831d) (2009).

2. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in Drosophila melanogaster. *Nature* 473, 337–342 (2011).

3. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218–223 (2009).

4. McShane, E. et al. Kinetic analysis of protein stability reveals age-dependent degradation. *Cell* 165, 535–550 (2016).

5. Schwanhauser, B. et al. Global quantification of mammalian gene expression control. *Nature* 473, 337–342 (2011).

6. Lee, M. V. et al. A dynamic model of proteome changes reveals new roles for RNA binding proteins in vertebrate embryonic development. *Dev. Cell* 35, 383–394 (2015).

7. Jovanovic, M. et al. Immuno-sequencing. Dynamic profiling of the protein life cycle in response to pathogens. *Science* 347, 1259038 (2015).

8. Casas-Vila, N. et al. The developmental proteome of Drosophila melanogaster. *Genome Res.* 27, 1273–1285 (2017).

9. Graveley, B., Brooks, A. & Carlson, J. The developmental transcriptome of *Drosophila melanogaster*. Nature 471, 473–479 (2011).

10. Robles, M. S., Cox, J. & Mann, M. In vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet.* 10, e1004074 (2014).

11. Symmonds, P. et al. Accurate protein quantification by delayed normalisation and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* 13, 2513–2526 (2014).

12. Beyer, A., Hollunder, J., Nasheuer, H.-P. & Wilhelm, T. Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. *Mol. Cell. Proteom.* 3, 1083–1092 (2004).

13. Margrove, I. L. & Schmidt, F. H. The role of mRNA and protein stability in gene expression. FASEB J. 13, 2360–2370 (1999).

14. Wu, G., Nie, L. & Zhang, W. Integrative analyses of posttranscriptional regulation in the yeast Saccharomyces cerevisiae using transcriptomic and proteomic data. *Curr. Microbiol.* 57, 18–22 (2008).

15. Lasko, P. mRNA localization and translational control in Drosophila melanogaster. *Cold Spring Harbor Perspect. Biol.* 4, a012294–a012294 (2016).

16. De Renzis, S., Elemento, O., Tavazoie, S. & Wieschaus, E. F. Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biol.* 5, 1036–1051 (2007).

17. Grossman, I., Jung, M.-Y., Sarkissian, M., Cao, Q. & Richter, J. D. Translational control of the embryonic cell cycle. *Cell* 109, 473–483 (2002).

18. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–177 (2013).

19. Ducat, D., Shin-ichi, K., Hongbin, L., John, R. Y. & Zheng, Y. Regulation of microtubule assembly and organization in mitosis by the AAA ATPase pontin. *Mol. Biol. Cell* 19, 308–317 (2008).

20. Sysoev, V. O. et al. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *FASEB J.* 23, 2435–2447 (2009).

21. Fabre, B. et al. Analysis of Drosophila melanogaster proteome dynamics during embryonic development by a combination of label-free proteomics approaches. *Proteomics* 16, 2068–2080 (2016).

22. Zeileis, A. & Grothendieck, G. *R* S3 Class and Methods for Indexed Totally Ordered Observations (Institut für Statistik undMathematik, WU Vienna University of Economics and Business, Vienna, 2005).

23. Fabre, B. et al. Analysis of Drosophila melanogaster proteome dynamics during embryonic development by a combination of label-free proteomic approaches. *Proteomics* 16, 2068–2080 (2016).

24. Zeileis, A. & Grothendieck, G. *R* S3 Class and Methods for Indexed Totally Ordered Observations (Institut für Statistik undMathematik, WU Vienna University of Economics and Business, Vienna, 2005).

25. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional gene regulation. *Nature* 449, 323–330 (2007).

26. Rinn, J. L. & Sonenberg, N. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–745 (2009).

27. Qin, X., Ahn, S., Speed, T. P. & Rubin, G. M. Global analyses of mRNA translational control during early Drosophila embryogenesis. *Genome Biol.* 8, R6–R63 (2007).

28. Sysoev, V. O. et al. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nat. Commun.* 7, 12128 (2016).

29. Flint, S., Kozlov, C., Ghosh, S., Kim, S. & Choi, H. P. FCA: a novel statistical tool for deconvoluting time-dependent gene expression regulation. *Proteome Res.* https://doi.org/10.1002/pr008535 (2014).

30. Cheng, Z. et al. Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Mol. Syst. Biol.* https://doi.org/10.1523/mdb.20156423 (2016).

31. Pepinhon, L. et al. On the relationship of protein and mRNA dynamics in vertebrate embryonic development. *Dev. Cell* 35, 383–394 (2015).

32. Hargrove, J. L. & Schmidt, F. H. The role of mRNA and protein stability in gene expression. FASEB J. 13, 2360–2370 (1999).

33. Wu, G., Nie, L. & Zhang, W. Integrative analyses of posttranscriptional regulation in the yeast Saccharomyces cerevisiae using transcriptomic and proteomic data. *Curr. Microbiol.* 57, 18–22 (2008).

34. Lasko, P. mRNA localization and translational control in Drosophila melanogaster. *Cold Spring Harbor Perspect. Biol.* 4, a012294–a012294 (2016).

35. Grossman, I., Jung, M.-Y., Sarkissian, M., Cao, Q. & Richter, J. D. Translational control of the embryonic cell cycle. *Cell* 109, 473–483 (2002).

36. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–177 (2013).

37. Ducat, D., Shin-ichi, K., Hongbin, L., John, R. Y. & Zheng, Y. Regulation of microtubule assembly and organization in mitosis by the AAA ATPase pontin. *Mol. Biol. Cell* 19, 308–317 (2008).

38. Sysoev, V. O. et al. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *FASEB J.* 23, 2435–2447 (2009).

39. Zeileis, A. & Grothendieck, G. *R* S3 Class and Methods for Indexed Totally Ordered Observations (Institut für Statistik und Mathematik, WU Vienna University of Economics and Business, Vienna, 2005).

40. Fabre, B. et al. Analysis of Drosophila melanogaster proteome dynamics during embryonic development by a combination of label-free proteomic approaches. *Proteomics* 16, 2068–2080 (2016).

41. Zeileis, A. & Grothendieck, G. *R* S3 Class and Methods for Indexed Totally Ordered Observations (Institut für Statistik und Mathematik, WU Vienna University of Economics and Business, Vienna, 2005).

42. Fabre, B. et al. Analysis of Drosophila melanogaster proteome dynamics during embryonic development by a combination of label-free proteomic approaches. *Proteomics* 16, 2068–2080 (2016).

43. Zeileis, A. & Grothendieck, G. *R* S3 Class and Methods for Indexed Totally Ordered Observations (Institut für Statistik und Mathematik, WU Vienna University of Economics and Business, Vienna, 2005).

44. Fabre, B. et al. Analysis of Drosophila melanogaster proteome dynamics during embryonic development by a combination of label-free proteomic approaches. *Proteomics* 16, 2068–2080 (2016).
52. Andrews, S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/ fastqc/ or http://www.bioinformatics.babraham.ac.uk/projects/ (2010). citeulike-article-id:11583827

53. Sayols, S., Scherzinger, D. & Klein, H. dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. BMC Bioinformatics. 17, 428 (2016).

54. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 30, 923–930 (2014).

55. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 200–201 (2014).

56. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods 7, 1009–1015 (2010).

57. Cedersund, G. & Roll, J. Systems biology: model based evaluation and comparison of potential explanations for given biological data. FEBS J. 276, 903–922 (2009).

58. Raue, a, Kreutz, C., Maiwald, T., Klingmüller, U. & Timmer, J. Addressing parameter identifiability by model-based experimentation. IET Syst. Biol. 5, 120–130 (2011).

59. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 10, 48 (2009).

60. Yu, G. et al. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 26, 976–978 (2010).

61. McLeay, R. C. & Bailey, T. L. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics. 11, 165 (2010).

62. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018 (2011).

63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).

64. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. Genome Res. 22, 2008–2017 (2012).

65. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121 (2015).

Author contributions
A.B. and N.C.-V. carried out embryo collection and preparation of samples for proteome and RNA-seq measurement. Processing of raw protein data was carried out by M.D. Raw reads from RNA-seq were processed by S.S. S.S. and K.B. analyzed correlation between data from Gravelle et al. (2011) and our RNA-seq data. K.B. carried out time-course clustering, correlation analysis of transcriptome and proteome, model fitting and selection, model analysis, gene set enrichment analysis, and motif analysis. C.K. provided input concerning model fitting and selection procedures. N.D. prepared Htr98DE knockdown cell lines and prepared samples for RNA-sequencing. N.C.-V. and A.B. prepared Htr98DE knockdown cell-line samples for proteome measurement. Statistical analysis of Htr98DE knockdown results was done by K.B. while analysis of differential splicing was carried out by S.S. K.B., F.B., S.L., and J.-Y.R. designed the project. All authors discussed the results. K.B., F.B., and S.L. wrote the manuscript with input from all other authors.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07455-9.

Competing interests: S.S. is currently employed at Boehringer Ingelheim International GmbH. The remaining authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements
We acknowledge the help of Hannah Lukas and Maria Méndez-Lago (IMB Genomics Core Facility) for RNA sequencing. We thank Alexej Tulin for kindly providing a list of Htr98DE targets, Stefanie Ebersberger for valuable advice on motif analysis and Bertrand Fabre for discussion on protein normalization. S.L. and K.B. were supported by the ebio junior group program of the German Federal Ministry of Education and Research (BMBF, FKZ: 0316196) and by the German Research Foundation (grant LE 3473/2-1). Part of the work was supported by the Rhineland-Palatinate Forschungsschwerpunkt GeneRED.