RNA-binding profiles of *Drosophila* CPEB proteins Orb and Orb2

Barbara Krystyna Stepień^{a,1,2}, Cornelia Oppitz^{a,3}, Daniel Gerlach^{a,3,4}, Ugur Dag^b, Maria Novatchkova^a, Sebastian Krüttner^{a,5}, Alexander Stark^a, and Krystyna Keleman^{a,b,1}

^aThe Research Institute of Molecular Pathology, 1030 Vienna, Austria; and ^bHoward Hughes Medical Institute, Janelia Research Campus, Ashburn, VA 20147

Edited by Eric C. Lai, Sloan-Kettering Institute, New York, NY 10065, and accepted by Editorial Board Member Kathryn V. Anderson September 26, 2016 (received for review March 8, 2016)

Localized protein translation is critical in many biological contexts, particularly in highly polarized cells, such as neurons, to regulate gene expression in a spatiotemporal manner. The cytoplasmic polyadenylation element-binding (CPEB) family of RNA-binding proteins has emerged as a key regulator of mRNA transport and local translation required for early embryonic development, synaptic plasticity, and long-term memory (LTM). *Drosophila* Orb and Orb2 are single members of the CPEB1 and CPEB2 subfamilies of the CPEB proteins, respectively. At present, the identity of the mRNA targets they regulate is not fully known, and the binding specificity of the CPEB2 subfamily is a matter of debate. Using transcriptome-wide UV cross-linking and immunoprecipitation, we define the mRNA-binding sites and targets of *Drosophila* CPEBs. Both Orb and Orb2 bind linear cytoplasmic polyadenylation element-like sequences in the 3' UTRs of largely overlapping target mRNAs, with Orb2 potentially having a broader specificity. Both proteins use their RNA-recognition motifs but not the Zinc-finger region for RNA binding. A subset of Orb2 targets is translationally regulated in cultured S2 cells and fly head extracts. Moreover, pan-neuronal RNAi knock-down of these targets suggests that a number of these targets are involved in LTM. Our results provide a comprehensive list of mRNA targets of the two CPEB proteins in *Drosophila*, thus providing insights into local protein synthesis involved in various biological processes, including LTM.

CPEB | CLIP | Orb2 | translation | long-term memory

Local protein translation restricts protein synthesis to specific cellular domains. This mechanism of gene-expression regulation is essential for many biological processes, including the production of gametes (1), animal development (2, 3), and maintaining tissue polarity in adult organisms (2, 4). In the nervous system local protein synthesis is critical for the development of highly polarized cells such as neurons (5) and is necessary for dynamic changes in neural architecture in adulthood (6) including those that take place during long-term memory (LTM) formation (7). The molecular machinery controlling local protein synthesis has been best studied in vertebrate oocyte development during which members of the cytoplasmic polyadenylation element binding (CPEB) protein family, in concert with other RNA-binding proteins, control the temporal expression of maternal RNAs (8, 9). More recently, CPEB proteins also have been found to be expressed in adult nervous systems (10–13) where they are thought to be involved in synaptic plasticity and LTM (12, 14–17), suggesting evolutionary conservation of the mechanism of local translation across species, tissues, and developmental stages.

The *Drosophila melanogaster* genome encodes two CPEB proteins, Orb and Orb2, representing two distinct branches of this family: Orb belongs to the CPEB1 branch, and Orb2 is the single representative of the CPEB2 subfamily. Although both proteins play critical and nonredundant roles in germline formation (18, 19), Orb2 is also essential for nervous system development (13) and is acutely required for LTM (12, 20), as is Orb (15). Both proteins contain an unstructured N-terminal poly-glutamine (poly-Q) stretch and a well-conserved C-terminal RNA-binding domain (RBD) consisting of two RNA recognition motifs (RRMs) and a

zinc finger (Znf) region (12). Most CPEB proteins exist in multiple isoforms (11). Orb2 has two variants, Orb2A and Orb2B (12), which differ in the composition of the N terminus preceding the poly-Q and share a common RBD (12). The poly-Q is required exclusively for LTM, whereas the RBD is required for both development and LTM (20), and its mutations are lethal (12, 13). Moreover, the RBD of Orb2 can be functionally replaced by the RBD of mouse CPEB2 (mCPEB2) but not by that of Orb or mCPEB1, suggesting the conservation of target specificity within but not between the CPEB subfamilies, at least in regard to their developmental function (20).

The function of the CPEB proteins is critically dependent on their ability to bind specific mRNA targets. However, our knowledge about their binding specificity remains incomplete. Although the CPEB1/Orb subfamily has a well-established specificity toward the T-rich cytoplasmic polyadenylation element (CPE) sequence (21), reports about the Orb2 subfamily binding motif are conflicting. In a study involving SELEX (Systematic Evolution of Ligands by Exponential Enrichment), the rodent Orb2 orthologs CPEB3 and CPEB4 (22) were shown to bind stem-loop RNA structures rather

Significance

Local protein synthesis is a highly used mechanism to create functional asymmetries within cells. It underlies diverse biological processes, including the development and function of the nervous and reproductive systems. Cytoplasmic polyadenylation element-binding (CPEB) proteins regulate local translation in early development, synaptic plasticity, and long-term memory. However, their binding specificity is not fully resolved. We used a transcriptome-wide approach and established that *Drosophila* representatives of two CPEB subfamilies, Orb and Orb2, regulate largely overlapping target mRNAs by binding to CPE-like sequences in their 3' UTRs, potentially with a shift in specificity for motif variants. Moreover, our data suggest that a subset of these mRNAs is translationally regulated and involved in long-term memory.

Author contributions: K.K. conceived the project; B.K.S., S.K., and K.K. designed research; B.K.S., C.O., and U.D. performed research; B.K.S., D.G., M.N., and A.S. analyzed data; and B.K.S. and K.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. E.C.L. is a Guest Editor invited by the Editorial Board.

Data deposition: The sequence data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database (accession no. GSE59611).

¹To whom correspondence may be addressed. Email: stepien@mpi-cbg.de or kelemank@janelia.hmi.org.

²Present address: The Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany.

³C.O. and D.G. contributed equally to this work.

⁴Present address: Boehringer Ingelheim Austria, 1121 Vienna, Austria.

⁵Present address: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603715113/-DCSupplemental.

than the linear CPE sequence recognized by the CPEB1 paralogue. Other reports suggest that both CPEB1 and CPEB4 recognize and control at least partially overlapping sets of mRNA targets and bind the same motif (23–25). Although a small number of fly Orb2 mRNA targets was described (26), the authors could not identify the bound sequence unambiguously. Thus, the full spectrum of fly CPEB targets is largely unknown, with previous studies characterizing only a limited number of interacting mRNAs (26–28). Elucidation of the RNA-binding specificity of Orb2 and the identification of its mRNA targets is essential to understand the role of downstream effectors in local translation-dependent processes, including LTM formation.

In this study we aimed at the transcriptome-wide identification of Orb- and Orb2-binding sites and mRNA targets using cross-linking and immunoprecipitation (CLIP). Through an extensive bioinformatic analysis of the high-throughput sequencing data, we obtained a comprehensive list of mRNA targets regulated by *Drosophila* CPEB proteins and determined that Orb and Orb2 bind CPE-like sequences but potentially do so with shifted specificity for particular motif subtypes and length. We independently confirmed the requirement for this motif in Orb2-dependent translational repression and showed that a number of candidate Orb2 CLIP targets are translationally regulated in a cell-based assay and in fly head extracts. Finally, knockdown of a subset of the verified Orb2 targets impairs courtship LTM, suggesting the functional significance of these targets in this process in vivo. Our results suggest that both CPEB proteins might function as key regulators of local protein synthesis in a plethora of biological contexts, including LTM, by binding to distinct variants of the CPE motif and regulating a common set of target mRNAs.

Results

CPEB Proteins Bind 3' UTRs of mRNA Targets Through the RRM but not the Znf. To identify Orb and Orb2 RNA-binding sites and potential mRNA targets on a transcriptome-wide scale, we used UV CLIP (29). To generate CLIP libraries, we used the *D. melanogaster* S2 cell expression system (30) in which we expressed WT or mutant forms of fly CPEB proteins. Orb2 mRNA is expressed endoge-

nously in S2 cells [average reads per kilobase of transcript per million reads mapped (RPKM) = 8.5; $n = 2$], and more than 90% of their transcriptome is detectable (RPKM >1) in the neuronal and gonad tissues, which express CPEB proteins in vivo (*SI Analysis*) (31). This expression and the fact that S2 cells allow Orb2-mediated translational suppression (26) suggest that S2 cells constitute a suitable system for investigating the transcriptome-wide Orb2-binding profile. Importantly, in S2 cells the overexpressed Orb2A exists in the form of protein clusters, reflecting its endogenous propensity to aggregate (*Fig. S1A*). In contrast, the overexpressed Orb2B isoform forms smaller clusters that partially colocalize with the ribonucleoprotein (RNP) markers Trailer hitch (Tral) and Eukaryotic translation initiation factor 4E (eIF4E) (*Fig. S1A and C*), reminiscent of the endogenous Orb2 protein in *Drosophila* brain in vivo (20). CPEB aggregates are thought to be associated with their normal function, including LTM, and to be mediated by the poly-Q domain (20, 32–34). In comparison, Orb mRNA is not expressed in S2 cells (average RPKM = 0.1; $n = 2$), although the majority of its known interactors (88%) are present (RPKM >1) (35, 36). Also, overexpressed Orb forms particles that partially colocalize with RNP markers (*Fig. S1B and C*). Nevertheless, we have generated inducible S2 cell lines for both proteins, with Orb serving as a reference for our approach because its binding specificity is well defined (21, 37).

Several modifications of the original CLIP protocol (38) confer advantages in sample preparation and postexperimental data analysis. To increase UV cross-linking specificity and to facilitate robust bioinformatic analysis based on diagnostic mutations, we combined the PAR-CLIP (39) and iCLIP (40) protocols, allowing identification of cross-linking sites based on specific mutations and read termination events, respectively. Because efficient UV cross-linking is crucial for CLIP, we first assessed the ability of both Orb and Orb2 proteins to form UV-induced RNA cross-links. We expressed WT and mutated protein variants in S2 cells (30) in the presence of the photoactivatable nucleotide analog 4-thiouridine (4-SU). The proteins were tagged C-terminally with the myc epitope, which does not interfere with their in vivo function (12, 20).

Fig. 1. CPEB proteins bind to RNA using RRM but not Znf motifs. The RRM region of Orb2 (A) and Orb (B) RBD, but not the Znf or N-terminal regions, is necessary for UV-induced RNA cross-linking. Shown are autoradiography of P^{32} -labeled complexes of UV cross-linked RNA with different Orb2 variants and corresponding Western blot images of immunoprecipitates (IP). Control indicates a mock precipitation of Orb2B-GFP (A) or Orb-GFP (B) protein with α -myc antibody. The UV wavelength and power are indicated above the image. All Orb2 mutations are derived from the B isoform. The WT A isoform differs from B in its N-terminal exon. Schematics of Orb2 and Orb variants corresponding to tested constructs and RNA/protein signal ratios are shown below. The Q-rich region is shown in red; the RBD is outlined in black with RRM motifs in green and Znf in yellow. Alternative Orb2 exons are depicted with different shades of blue. Sequences of mutated regions are shown in boxes below the protein schematics.

First we investigated whether UV cross-linking is mediated by the RBD. We expressed Orb2A and Orb2B in S2 cells (12, 20). Both isoforms cross-linked to RNA robustly (Fig. 1A); however, we focused on Orb2B because it requires intact RBD for its *in vivo* function (20). We generated mutations in the regions outside and within the RBD. Mutations of conserved residues in either of the RRM motifs (Y492/F494 and R601/F604) (20) in the proximal part of RBD abolished binding to control levels. Surprisingly, mutating the conserved cysteines (C669/C672) in the Znf of the RBD, previously implicated in RNA binding *in vitro* (22), had only a minor effect. Similarly, deletion of the Orb2 poly-Q region ($\Delta 168$ –220), shown to be essential for LTM (12), had no impact on its RNA-binding ability. Analogous results were obtained for the corresponding mutations in the Orb protein, although the RRM1 mutation had a smaller effect (Fig. 1B). The lower UV cross-linking signal in Orb compared with Orb2 is likely caused by slightly lower Orb overexpression and possibly by intrinsic differences between the proteins, which share only about 40% of RBD homology. Detailed quantification of the RNA cross-linking can be found in Fig. S24. These data show that modifications of the regions outside the RBD (poly-Q deletion or the substitution of an alternative N-terminal exon) do not influence the robustness of RNA cross-linking, but the RRM sequence (although not the Znf in the RBD) is necessary for efficient RNA binding in both proteins.

To analyze the Orb- and Orb2-bound transcriptome, we performed PAR-CLIP experiments. We generated multiple replicate cDNA libraries from S2 cells expressing C-terminally myc-tagged Orb or Orb2 pretreated with 4-SU to enhance RNA cross-linking specificity (39) or with 6-thioguanosine (6-SG) to control for potential cross-linking bias. cDNA libraries were prepared with α -myc antibodies following an iCLIP protocol (40). We prepared six Orb2 and three Orb 4-SU cDNA libraries and a single 6-SG library for

each genotype. As a negative control, we prepared a library from cells expressing the loss of the RNA-binding mutant protein, Orb2B RRM1&2*-myc. The RNA-protein cross-linking was specific, because additional controls (no UV illumination, no photoactivatable nucleotide treatment, and mock precipitation with the α -GFP instead of α -myc antibodies) produced no signal on the autoradiography gel (Fig. S2B and C). Details on CLIP sample preparation, sequencing, and analysis can be found in *SI Materials and Methods*.

As expected from previous studies, the binding sites of *Drosophila* CPEB proteins (26, 27) are located preferentially in 3' UTR regions: 22.5% of all Orb2-bound and 37% of all Orb-bound reads are located in 3' UTR regions, although this genomic feature only covers 3.4% of the fly genome (6.6 $\times$ and 10.8 $\times$ enrichment, respectively) (Fig. 2A). Little binding was shown in intronic regions for either genotype. This finding is not surprising, because CPEB proteins are largely cytoplasmic under normal growth conditions, although they can shuttle to the nucleus (Fig. S1B).

To identify the potential target genes, we used three complementary methods. First, we used the PARalyzer method (41), which is based on the diagnostic substitutions (T > C or G > A) at cross-link sites stemming from the use of photoactivatable nucleotide analogs (39). Second, we created an in-house method, NovoCLIP, which incorporates information based both on substitutions and on read termination positions characteristic of iCLIP protocol (40). Finally, because deletions were reported to be indicative of cross-linking sites (42), potential targets were scored using T or G deletions as diagnostic mutations. The Orb dataset is highly biased for T > C substitutions (Fig. S3A), whereas Orb2 shows a higher bias for T deletions (Fig. S3B); both datasets were used for analysis. Potential target genes were identified based on cross-linking sites overlapping the 3' UTR regions of annotated

Fig. 2. CPEB proteins bind the 3' UTR of target mRNAs. (A) Fly CPEB-binding sites are enriched in 3' UTRs. Pie charts illustrate the relative contribution of reads in various gene regions in comparison with their genomic distribution shown at right. The percentages of reads mapping to 3' UTRs are indicated. Bar plots depict the enrichment of reads in indicated regions as the natural logarithm of the ratio between sample and genomic read fraction in each region. (B) Primary CLIP targets based on 3' UTR mapping clusters. Venn diagrams illustrate the number of CPEB targets identified with three complementary approaches. The numbers inside the circles represent the number of targets identified with the respective methods. Note that Orb2B RRM1&2* binds few targets.

genes. All methods led to largely overlapping results (82% and 79% of Orb and Orb2 targets were identified by both PARalyzer and NovoCLIP methods) (Fig. 2B), although deletion analysis yielded fewer target sites because of the relative rarity of such events.

Biological complexity (BC) is an established criterion used to filter CLIP datasets to minimize experimental noise (43). Candidate targets were filtered by their reproducibility in at least three independent libraries ($BC \geq 3$). We identified 3,128 targets for Orb and 2,154 for Orb2. Most targets were identified by both PARalyzer and NovoCLIP: 73% for Orb and 66% for Orb2 (Fig. S3C). In contrast, the single Orb2B RRM1&2* control library identified only 376 genes. The list of genes containing CPEB cross-linking sites in their 3' UTRs and their reproducibility are provided in Fig. S3D and Dataset S1.

Orb and Orb2 Bind to Linear CPE Motifs. To resolve a long-standing question about the RNA-binding specificity of Orb2, we turned to de novo motif discovery using BioProspector (44). The analysis of the Orb dataset returned a T-rich binding motif comprising four T nucleotides followed by A in the fifth position. This motif closely resembles the linear CPE sequence bound by the vertebrate Orb ortholog CPEB1 (21, 37). In contrast, analysis of the Orb2 dataset revealed a T₄ motif that could not be resolved toward the 3' end (Fig. 3A). For comparison, nontarget 3' UTR sequences are not enriched in CPE motifs (Fig. S4A).

Because the 3' variability in the CPE motif composition and length (45–48) can obscure the de novo analysis, we searched both datasets for CPE variants containing either A or G in fifth position (Fig. 3B) and calculated enrichment over background 3' UTR control. The analyzed groups of CPE motifs were indeed enriched in CLIP sequences for both proteins. A-containing motifs are consistently enriched in the Orb dataset, as expected from de novo analysis. In contrast, the Orb2 dataset showed overall enrichment for G-containing motifs. A more detailed comparison of individual motif types shows that the Orb dataset is enriched in CPE variants containing A in the fifth position [TTTT(A)_{1–3}T], whereas the TTTTGT and TTTTAAAT motifs are the most enriched in the Orb2 dataset (Fig. 3C); this difference could explain the difficulty in resolving the fifth nucleotide in de novo analysis. Modification of this analysis using only nontarget 3' UTRs as the background model supports the strong preference of Orb protein for A-containing CPE variants (Fig. S4B and C). Interestingly, in this case the Orb2 dataset enrichment of TTTTAAAT is much lower, whereas the TTTTGT sequence remains a predominant motif (Fig. S4B and C). In contrast, no clear de novo motif or enrichment of the CPE sequences was found for the Orb2B RRM1&2* control dataset (Fig. S4D). A CPE-like motif also can be found in clusters derived from intronic binding (Fig. S4E), especially for Orb, although introns were not enriched in the CLIP dataset (Fig. 2A). Additionally, because CPEB proteins are known to influence splicing in proximity to poly(A) signal sequences (49), we analyzed the average distance between CPEB-bound motifs and poly(A) signal sequences (Fig. S4F). There is a small but significant difference in the average distance compared with the nontarget 3' UTR background (13.8 for Orb and 14.9 for Orb2 compared with 17.3 and 16.8 for the control sets, Student *t* test, $P < 0.001$).

Taken together our results show that, although both CPEB proteins recognize CPE motifs, Orb has a clear preference for canonical TTTT(A)_{1–3}T sequences. Orb2 appears to have a broader specificity that may include but not be restricted to motifs containing G in the fifth position. However, it must be mentioned that the results of the analysis of the Orb2 motif variant are influenced by the method used, and further biochemical studies will be needed to confirm the specificity shift between the paralogues and its functional significance.

To confirm that the CPE motif is indeed required for the Orb2-mediated translational control, we turned to a dual luciferase reporter assay. Orb2 was previously shown to repress the translation of a small number of targets, including its own 3' UTR (26). Indeed,

Fig. 3. Fly CPEB proteins bind the T-rich CPE sequence motif necessary for Orb2-mediated translational repression. (A) Sequence logos for Orb and Orb2 obtained using the PARalyzer and NovoCLIP methods separately. The canonical CPE motif is shown below. (B) CPE motifs are enriched in the CPEB CLIP dataset. Orb-binding sites are enriched in typical A-containing CPE motifs (light green), whereas Orb2 shows more preference for G-containing versions (dark green). The CPE variations used are listed on the right. (C) Enrichment scores for distinct versions of the CPE motif containing A or G in the fifth position. Orb and Orb2 are enriched for different CPE versions, whereas the Orb2 RRM1&2* mutant protein does not show enrichment for any analyzed sequences. Dashed lines indicate the 1.5-fold enrichment threshold.

expression of the firefly luciferase tethered to a long version of 3' UTR of Orb2 isoform H (RH) or short of Orb2 isoform A (RA), was repressed by Orb2 (RH: ratio 0.51 ± 0.01 SEM, RA: ratio 0.77 ± 0.03 SEM, Student *t* test, $P < 0.001$) (Fig. 4A). RRM mutations (Orb2B RRM1*, 2*, and 1&2*) abolish the repression of target translation (Student *t* test, $P = 0.16$), but poly-Q or ZnF alterations have no effect (Student *t* test, $P < 0.001$) (Fig. 4B), as is consistent with the RNA UV cross-linking pattern (Fig. 1A). We did not assess the effect of Orb mutations in this assay, because WT Orb caused a widespread, unspecific repression of all tested 3' UTRs, perhaps caused by pleiotropic effects or by a requirement for additional components lacking in S2 cells.

(Fig. 4C). These results show that the presence of CPE motifs in target 3' UTRs is necessary for Orb2-mediated translational control. Moreover, mutating the single CPE motif in Ald-RE from TTTTAT to TTTTGT leads to equally potent translational suppression mediated by Orb2 (Fig. 4D), indicating that Orb2 can bind both A- and G-containing CPE variants, at least under conditions of overexpression.

Orb2 Regulates the Translation of Its Targets in Vitro and in Vivo. Having established that Orb2 regulates translation through a CPE motif, we narrowed the list of Orb2 targets to the most reproducible ones occurring in at least three independent libraries ($BC \geq 3$) for which individual CLIP clusters overlap at least one CPE motif (CPE⁺) (Dataset S1). The final list contained 1,637 3' UTRs (76% of $BC \geq 3$ primary targets and 9% of *Drosophila* genes). Similarly, 2,691 Orb targets overlapped a CPE element (86% of $BC \geq 3$ primary targets and 15% of *Drosophila* genes). More than 94% of identified Orb2 targets and 98% of Orb targets are also expressed in fly head extracts, suggesting that these mRNAs could be targets in vivo. In comparison only 94 3' UTRs of all the Orb2B RRM1&2* targets have a cluster overlapping a CPE motif, as is consistent with the lack of enrichment for the motif in the mutant library.

To test functionally whether Orb2 regulates the translation of the identified target genes, we assessed the translational repression of the luciferase reporter induced by a set of target 3' UTRs in the S2 cell-based assay as described above. We tested a set of 222 3' UTRs representing 152 Orb2 target genes from the stringent dataset ($BC \geq 3$, CPE⁺). Additionally, we tested 22 3' UTRs of 21 genes classified as potential false positives: that is, genes with reproducibility of $BC \geq 3$ but no CLIP cluster overlapping a CPE motif ($BC \geq 3$, CPE⁻), and seven negative controls ($BC < 3$), whose RNA is expressed in the untransfected S2 cells with RPKM > 10 (Dataset S2). Of all the Orb2 target 3' UTRs tested, 92 3' UTRs (41.4%) of 69 genes (45.4%) were regulated (all but two were repressed), compared with only two 3' UTRs in the predicted false-positive set and a single 3' UTR in the negative control set (Fig. 4E). In addition, we investigated the Orb2B regulated 3'UTRs for other 3'UTR features that might determine the regulation by Orb2. Although the 3' UTRs regulated by Orb2 are longer on average and have higher CPE density (SI Analysis and Dataset S3), we failed to find any additional features that would predict whether a given CPE is active in Orb2-mediated translational repression (Fig. S5).

To test whether mRNA targets regulated by Orb2 in the S2 cell-based assay are also regulated in vivo, we investigated the expression of four proteins encoded by Orb2 targets [aPKC, Calcium/calmodulin-dependent protein kinase II (CaMKII), Fragile X mental retardation (Fmr1), and Upstream of N-ras (Unr)] in fly head extracts. We compared the protein levels in WT and Orb2-null fly heads using quantitative Western blotting (Fig. 4F). In all cases target protein levels are higher in heads of mutants lacking Orb2 suggesting that Orb2 acts as a translational repressor of these genes. The changes seen at the protein level are unlikely to be caused by Orb2's effect on RNA transcription or stability, because the global S2 cell mRNA levels do not change drastically upon Orb2 overexpression except for a small number (169) of genes. Only 13% of the genes that change mRNA levels in the presence of Orb2 (22 genes) are Orb2 CLIP targets (Fig. S6A and Dataset S4). Moreover, quantitative PCR (qPCR) for chosen target and nontarget 3' UTRs does not show a significant change in mRNA levels that would be consistent with translational regulation (Fig. S6B).

Orb2 Targets Regulate Cell and Tissue Polarity Genes and Play a Role in Courtship LTM. Orb2 has been best studied for its role in LTM formation and nervous system development, and some of its previously identified targets were suggested to contribute to

these processes (12, 13, 19, 20, 26). To determine what other biological processes may be regulated by Orb2 targets, we turned to Gene Ontology (GO) analysis (50) performed with the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (51, 52). Neurogenesis is one of the top-ranking GO terms associated with Orb2 CLIP targets, supporting Orb2's crucial role in this process (Fig. 5A) (12, 13, 20). Moreover, Orb2-bound transcripts are generally associated with the development of epithelial tissues and germline, suggesting that Orb2 may regulate a wide array of tissue-patterning and polarization processes (Dataset S5). Biological processes associated with Orb targets are reminiscent of Orb2, a finding that is unsurprising given the high overlap of Orb and Orb2 mRNA targets (Fig. 5A). The molecular function GO terms for the targets of both CPEB proteins shows a high involvement of protein kinases and small GTPases, suggesting regulation of signaling pathway components. The full list of significant terms for both proteins is provided in Dataset S5.

Because Orb2 is specifically required for LTM (12, 20, 32), we were interested in testing whether any of its translationally regulated targets have a role in this process. To this end, we knocked down candidate genes in all postmitotic neurons using the UAS-RNAi/Gal4 system and tested these flies for LTM induced by courtship conditioning (12, 53, 54). This form of memory is protein synthesis dependent (55) and requires an intact Orb2 RBD (20); thus the knockdown of relevant Orb2 RNA targets should impact LTM specifically. Of the 69 genes regulated by Orb2 in the luciferase reporter assay, seven are known to be involved in LTM [*CaMKII*, *Fmr1*, *murashka* (*mura*), *pumilio* (*pum*), *spoonbill* (*spoon*), *signal-transducer and activator of transcription protein at 92E* (*Stat92E*), and *staufen* (*stau*)]. We knocked down the expression of the remaining 61 candidates for which RNAi lines were available (there was no line for *nct*) in all neurons. Fourteen targets were lethal upon knockdown, possibly because of their involvement in the nervous system development, and were not analyzed further. The remaining 47 targets were tested using the LTM courtship conditioning assay. The knockdown of 10 (21.3%) of these genes was associated with the courtship LTM defect (Fig. 5B). This effect is specific to LTM and is not simply a developmental defect, because the function of these genes is required acutely for courtship LTM memory as assessed by conditional RNAi knockdown using the Gal80ts system (Fig. 5C). For all genes, knockdown under restrictive temperature resulted in impaired courtship LTM, although residual memory [learning index (LI) > 10] was seen with the knockdown of two genes, *polypeptide GalNAc transferase 5* (*pgant5*) and *metal response element-binding transcription factor-1* (*MTF-1*); this result requires further validation. Moreover, the impairment of courtship LTM is not caused by a failure of flies to learn, given that the short-term courtship memory (STM) remains unaffected (Fig. 5D). Detailed statistical analysis of the behavioral assays can be found in Dataset S6.

Discussion

In this study we characterized transcriptome-wide RNA-binding profiles of the *Drosophila* CPEB proteins Orb and Orb2. We determined that both proteins use RRM motifs but not ZnF or N-terminal regions to bind to the 3' UTR of target mRNA. Consistent with previous reports, we established that the Orb protein binds to a canonical T-rich CPE sequence previously described as a binding motif of its rodent ortholog, CPEB1 (21), thus validating our approach. Interestingly, Orb2 also appears to bind CPE sequences, but with potentially broader binding specificity, including G-containing CPE sequences in addition to A-harboring motifs. Consequently, both Orb and Orb2 bind a highly overlapping (50%) but not identical set of mRNA targets that have products associated with a wide range of processes involving cell and tissue polarization. Finally, we determined that

The annotation of Orb and Orb2 CLIP binding sites led to the identification of a high number of potential target genes reminiscent of the CLIP binding profiles of other RNA-binding proteins (29, 38). However, these targets likely contain a significant number of false positives, and identifying true targets from CLIP datasets with high confidence remains a challenge (43). Here, because of the combination of high biological reproducibility and CLIP cluster overlap to the identified motif, we were able to narrow the target list to ~10% of the *Drosophila* transcriptome. Additionally, we compared the smaller set of Orb ovary targets investigated using RNA immunoprecipitation (RIP) (62) with our data and discovered 243 high-confidence genes that appear in both sets (*SI Analysis* and *Dataset S1*). It would be interesting to validate and characterize the downstream functions of these targets further.

Additional analysis of more than 200 candidate 3' UTRs in a functional assay verified this approach. Including the measure of reproducibility and motif recognition substantially improved target identification, as shown using an in vitro assay for Orb2-mediated translational control. The actual positive rate in our assay was 45.4%, consistent with previous reports (26). The overall positive rate might be higher, because our assay is based on the overexpression of both Orb2 and the reporter 3' UTR in S2 cells; therefore, the degree of regulation obtained for some mRNAs might be low because of the requirement for additional components lacking in S2 cells. Importantly, the great majority of targets identified in S2 cells are also expressed in head extracts, suggesting that they are available for binding in vivo.

Orb2 plays a crucial role in the development and function of the nervous system (12, 13). Consistently, neurogenesis is one of the top GO terms associated with Orb2's newly identified targets. A number of targets regulated by Orb2 (e.g., *brat*, *pum*, *spoon*, *aPKC*, *CaMKII*) are known to function in these processes (59). Furthermore, known LTM genes are among the translationally regulated Orb2 targets, and a number of other genes

were implicated in courtship LTM by RNAi knockdown. Further analysis will be required to confirm and mechanistically explain their function in LTM. Our work identifies a comprehensive list of potential mRNA targets regulated by Orb2, thus providing a framework for further study of the regulatory network linking translation and nervous system development and function.

Materials and Methods

CLIP experiments were performed on *D. melanogaster* S2 cells (Invitrogen) transfected with Orb or Orb2 constructs. Orb2 isoform A and B ORFs were cloned from cDNA libraries prepared from fly heads using the Gateway system (Life Technologies), and mutations were introduced by site-specific mutagenesis. The 3' UTR sequences were amplified from genomic DNA. Incubation with photoactive nucleotide analogs, and UV cross-linking was done essentially as described previously (39). Monoclonal α -myc Abs (9E10) were used for immunoprecipitation. Immunoprecipitation and library preparation were done essentially as in ref. 40. UV cross-linking assessment of Orb2 variants was done following the PAR/CLIP protocol. For CLIP data analysis, three methods were used: PARalyzer (41), a method developed in-house (NovoCLIP), and scoring of deletions. For assessment of basal RNA expression levels, the two replicates were prepared from both Orb2B WT and nontransfected S2 cells and were analyzed as described (63). Fly head RNA-Seq samples were prepared as described in ref. 64. RNA-Seq and CLIP samples were sequenced using Illumina technology at the Campus Science Support Facilities (CSF) Next-Generation Sequencing (NGS) Unit (www.vbcf.ac.at). All sequencing data were submitted to the Gene Expression Omnibus (GEO) database under accession no. GSE59611. The luciferase reporter assay was performed as described (26). Behavioral analysis was done as in ref. 54. Primer sequences and detailed protocols can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Jernej Ule for help with optimizing the CLIP protocol and Maciej Pasternacki for help with bioinformatic analysis. pAC-Fluc-6xBS and pAC-Rluc plasmids were a kind gift from Chunghun Lim. The pMK33/pMThy plasmid was obtained from Silke Dörner, and the Gateway Vector Collection plasmids were provided by Jürgen Knoblich. This work was supported by Wiener Wissenschafts-, Forschungs-, und Technologiefonds Grant P21854.

- Richter JD, Lasko P (2011) Translational control in oocyte development. *Cold Spring Harb Perspect Biol* 3(9):a002758.
- Mohr E, Richter D (2001) Messenger RNA on the move: Implications for cell polarity. *Int J Biochem Cell Biol* 33(7):669–679.
- Piccioni F, Zappavigna V, Verrotti AC (2005) A cup full of functions. *RNA Biol* 2(4):125–128.
- Bramham CR, Wells DG (2007) Dendritic mRNA: Transport, translation and function. *Nat Rev Neurosci* 8(10):776–789.
- Jung H, Holt CE (2011) Local translation of mRNAs in neural development. *Wiley Interdiscip Rev RNA* 2(1):153–165.
- Jung H, Yoon BC, Holt CE (2012) Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat Rev Neurosci* 13(5):308–324.
- Wells DG, Fallon JR (2000) In search of molecular memory: Experience-driven protein synthesis. *Cell Mol Life Sci* 57(10):1335–1339.
- Charlesworth A, Meijer HA, de Moor CH (2013) Specificity factors in cytoplasmic polyadenylation. *Wiley Interdiscip Rev RNA* 4(4):437–461.
- Mendez R, Richter JD (2001) Translational control by CPEB: A means to the end. *Nat Rev Mol Cell Biol* 2(7):521–529.
- Kurihara Y, et al. (2003) CPEB2, a novel putative translational regulator in mouse haploid germ cells. *Biol Reprod* 69(1):261–268.
- Theis M, Si K, Kandel ER (2003) Two previously undescribed members of the mouse CPEB family of genes and their inducible expression in the principal cell layers of the hippocampus. *Proc Natl Acad Sci USA* 100(16):9602–9607.
- Keleman K, Krüttner S, Alenius M, Dickson BJ (2007) Function of the *Drosophila* CPEB protein Orb2 in long-term courtship memory. *Nat Neurosci* 10(12):1587–1593.
- Hafer N, Xu S, Bhat KM, Schedl P (2011) The *Drosophila* CPEB protein Orb2 has a novel expression pattern and is important for asymmetric cell division and nervous system function. *Genetics* 189(3):907–921.
- Si K, Choi YB, White-Grindley E, Majumdar A, Kandel ER (2010) Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. *Cell* 140(3):421–435.
- Pai TP, et al. (2013) *Drosophila* ORB protein in two mushroom body output neurons is necessary for long-term memory formation. *Proc Natl Acad Sci USA* 110(19):7898–7903.
- Chao HW, et al. (2013) Deletion of CPEB3 enhances hippocampus-dependent memory via increasing expressions of PSD95 and NMDA receptors. *J Neurosci* 33(43):17008–17022.
- Huang WH, Chao HW, Tsai LY, Chung MH, Huang YS (2014) Elevated activation of CaMKII α in the CPEB3-knockout hippocampus impairs a specific form of NMDAR-dependent synaptic depotentiation. *Front Cell Neurosci* 8:367.
- Lantz V, Chang JS, Horabin JL, Bopp D, Schedl P (1994) The *Drosophila* orb RNA-binding protein is required for the formation of the egg chamber and establishment of polarity. *Genes Dev* 8(5):598–613.
- Xu S, Hafer N, Agunwamba B, Schedl P (2012) The CPEB protein Orb2 has multiple functions during spermatogenesis in *Drosophila melanogaster*. *PLoS Genet* 8(11):e1003079.
- Krüttner S, et al. (2012) *Drosophila* CPEB Orb2A mediates memory independent of its RNA-binding domain. *Neuron* 76(2):383–395.
- Hake LE, Richter JD (1994) CPEB is a specificity factor that mediates cytoplasmic polyadenylation during *Xenopus* oocyte maturation. *Cell* 79(4):617–627.
- Huang YS, Kan MC, Lin CL, Richter JD (2006) CPEB3 and CPEB4 in neurons: Analysis of RNA-binding specificity and translational control of AMPA receptor GluR2 mRNA. *EMBO J* 25(20):4865–4876.
- Igea A, Méndez R (2010) Meiosis requires a translational positive loop where CPEB1 ensues its replacement by CPEB4. *EMBO J* 29(13):2182–2193.
- Novoa I, Gallego J, Ferreira PG, Méndez R (2010) Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nat Cell Biol* 12(5):447–456.
- Afroz T, et al. (2014) A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev* 28(13):1498–1514.
- Mastushita-Sakai T, White-Grindley E, Samuelson J, Seidel C, Si K (2010) *Drosophila* Orb2 targets genes involved in neuronal growth, synapse formation, and protein turnover. *Proc Natl Acad Sci USA* 107(26):11987–11992.
- Chang JS, Tan L, Schedl P (1999) The *Drosophila* CPEB homolog, orb, is required for oskar protein expression in oocytes. *Dev Biol* 215(1):91–106.
- Chang JS, Tan L, Wolf MR, Schedl P (2001) Functioning of the *Drosophila* orb gene in gurken mRNA localization and translation. *Development* 128(16):3169–3177.
- Ule J, et al. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302(5648):1212–1215.
- Schneider I (1972) Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embryol Exp Morphol* 27(2):353–365.
- Celniker SE, et al.; modENCODE Consortium (2009) Unlocking the secrets of the genome. *Nature* 459(7249):927–930.
- Majumdar A, et al. (2012) Critical role of amyloid-like oligomers of *Drosophila* Orb2 in the persistence of memory. *Cell* 148(3):515–529.

33. White-Grindley E, et al. (2014) Contribution of Orb2A stability in regulated amyloid-like oligomerization of *Drosophila* Orb2. *PLoS Biol* 12(2):e1001786.
34. Khan MR, et al. (2015) Amyloidogenic oligomerization transforms *Drosophila* Orb2 from a translation repressor to an activator. *Cell* 163(6):1468–1483.
35. Costa A, et al. (2013) Rasputin functions as a positive regulator of orb in *Drosophila* oogenesis. *PLoS One* 8(9):e72864.
36. Wong LC, Schedl P (2011) Cup blocks the precocious activation of the orb autoregulatory loop. *PLoS One* 6(12):e28261.
37. Stebbins-Boaz B, Hake LE, Richter JD (1996) CPEB controls the cytoplasmic polyadenylation of cyclin, Cdk2 and c-mos mRNAs and is necessary for oocyte maturation in *Xenopus*. *EMBO J* 15(10):2582–2592.
38. König J, Zarnack K, Luscombe NM, Ule J (2012) Protein-RNA interactions: New genomic technologies and perspectives. *Nat Rev Genet* 13(2):77–83.
39. Hafner M, et al. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141(1):129–141.
40. König J, et al. (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* 17(7):909–915.
41. Corcoran DL, et al. (2011) PARalyzer: Definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* 12(8):R79.
42. Zhang C, Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 29(7):607–614.
43. Darnell RB (2010) HITS-CLIP: Panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* 1(2):266–286.
44. Liu X, Brutlag DL, Liu JS (2001) BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* 2001: 127–138.
45. McGrew LL, Richter JD (1990) Translational control by cytoplasmic polyadenylation during *Xenopus* oocyte maturation: Characterization of *cis* and *trans* elements and regulation by cyclin/MPF. *EMBO J* 9(11):3743–3751.
46. Fox CA, Sheets MD, Wickens MP (1989) Poly(A) addition during maturation of frog oocytes: Distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUU. *Genes Dev* 3(12B):2151–2162.
47. McGrew LL, Dworkin-Rastl E, Dworkin MB, Richter JD (1989) Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev* 3(6):803–815.
48. Barkoff AF, Dickson KS, Gray NK, Wickens M (2000) Translational control of cyclin B1 mRNA during meiotic maturation: Coordinated repression and cytoplasmic polyadenylation. *Dev Biol* 220(1):97–109.
49. Piqué M, López JM, Foissac S, Guigó R, Méndez R (2008) A combinatorial code for CPE-mediated translational control. *Cell* 132(3):434–448.
50. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
51. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.
52. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1–13.
53. Siwicki KK, Ladewski L (2003) Associative learning and memory in *Drosophila*: Beyond olfactory conditioning. *Behav Processes* 64(2):225–238.
54. Reuter MS, et al. (2014) NDST1 missense mutations in autosomal recessive intellectual disability. *Am J Med Genet A* 164A(11):2753–2763.
55. Krüttner S, et al. (2015) Synaptic Orb2A bridges memory acquisition and late memory consolidation in *Drosophila*. *Cell Reports* 11(12):1953–1965.
56. Hake LE, Mendez R, Richter JD (1998) Specificity of RNA binding by CPEB: Requirement for RNA recognition motifs and a novel zinc finger. *Mol Cell Biol* 18(2): 685–693.
57. Merkel DJ, et al. (2013) The C-terminal region of cytoplasmic polyadenylation element binding protein is a ZZ domain with potential for protein-protein interactions. *J Mol Biol* 425(11):2015–2026.
58. Singh LN, Hannehalli S (2008) Functional diversification of paralogous transcription factors via divergence in DNA binding site motif and in expression. *PLoS One* 3(6): e2345.
59. McQuilton P, St Pierre SE, Thurmond J; FlyBase Consortium (2012) FlyBase 101—The basics of navigating FlyBase. *Nucleic Acids Res* 40(Database issue):D706–D714.
60. Vinuesa Veloz MF, et al. (2012) The effect of an mGluR5 inhibitor on procedural memory and avoidance discrimination impairments in Fmr1 KO mice. *Genes Brain Behav* 11(3):325–331.
61. Dubnau J, et al. (2003) The staufen/pumilio pathway is involved in *Drosophila* long-term memory. *Curr Biol* 13(4):286–296.
62. Rojas-Rios P, et al. (2015) Translational control of autophagy by Orb in the *Drosophila* germline. *Dev Cell* 35(5):622–631.
63. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628.
64. Han TW, et al. (2012) Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell* 149(4):768–779.
65. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162(1):156–159.
66. Koelle MR, et al. (1991) The *Drosophila* EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. *Cell* 67(1):59–77.
67. Evan GI, Lewis GK, Ramsay G, Bishop JM (1985) Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol* 5(12):3610–3616.
68. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
69. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
70. R Development Core Team (2008) The R Project for Statistical Computing (R Foundation for Statistical Computing, Vienna). Available at www.R-project.org. Accessed August 9, 2016.
71. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16):9440–9445.
72. Nakamura A, Sato K, Hanyu-Nakamura K (2004) *Drosophila* cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. *Dev Cell* 6(1):69–78.
73. Boag PR, Nakamura A, Blackwell TK (2005) A conserved RNA-protein complex component involved in physiological germline apoptosis regulation in *C. elegans*. *Development* 132(22):4975–4986.
74. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3(6):1101–1108.
75. Kamyshev NG, Iliadi KG, Bragina JV (1999) *Drosophila* conditioned courtship: Two ways of testing memory. *Learn Mem* 6(1):1–20.
76. Schindelin J, et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9(7):676–682.