TU-tagging: cell type–specific RNA isolation from intact complex tissues

Michael R Miller[1], Kristin J Robinson[1], Michael D Cleary[1,2] & Chris Q Doe[1]

We found that the combination of spatially restricted uracil phosphoribosyltransferase (UPRT) expression with 4-thiouracil delivery can be used to label and purify cell type–specific RNA from intact complex tissues in *Drosophila melanogaster*. This method is useful for isolating RNA from cell types that are difficult to isolate by dissection or dissociation methods and should work in many organisms, including mammals and other vertebrates.

Cell type–specific gene expression is a defining feature of multicellular organisms[1]. The analysis of cell type–specific transcriptomes can provide insight into the mechanisms used to generate cellular diversity[2] as well as help determine the underlying cause of disease[3]. Although a few methods are available for cell type–specific RNA isolation[4–7], each has constraints, and researchers are often limited by their ability to isolate RNA from cell types of interest[8]. Thus, developing new methods for cell type–specific RNA isolation is an important goal for genomic analysis of development and disease.

We previously had shown that the *Toxoplasma gondii* nucleotide salvage enzyme uracil phosphoribosyltransferase (UPRT) can be used to biosynthetically label newly synthesized RNA *in vivo*[9]. Under natural conditions, UPRT couples ribose-5-phosphate to the N1 nitrogen of uracil to yield uridine monophosphate (UMP), which is subsequently incorporated into RNA. When the modified uracil analog 4-thiouracil is provided to UPRT as a substrate, the resultant product is also incorporated into RNA, and this incorporation has little effect of cellular physiology[9]. Thio-substituted nucleotides are not a natural component of nucleic acids, and the resulting thio-labeled RNA can be readily tagged and purified using commercially available reagents. Owing to the ability to use this method to separate newly synthesized RNA from bulk cellular RNA, we and others have used it to measure RNA synthesis and decay rates[9,10].

Here we describe a different use for 4-thiouracil UPRT-based biosynthetic labeling that we call 'TU-tagging'. We reasoned that by spatially restricting UPRT expression in a multicellular organism, 4-thiouracil will be modified and subsequently incorporated into newly synthesized RNA only in cells expressing UPRT. Thus, even if RNA is isolated from the whole organism, RNA from the cells expressing UPRT can be recovered by purifying labeled RNA (**Fig. 1a**). This method would be particularly useful for isolating RNA from cell types that are difficult to isolate by dissection or dissociation methods, such as subsets of neurons or glia in the central nervous system.

To test the ability to biosynthetically label RNA in *Drosophila*, we delivered 4-thiouracil and monitored its incorporation into RNA by purifying total RNA, performing thio-biotin coupling and using streptavidin–horseradish peroxidase to detect labeled RNA (Online Methods). Wild-type larvae or adult flies, or larvae or adult flies containing only the GAL-4–inducible transgene UAS-UPRT (upstream activating sequence–UPRT) but no source of GAL4, had very small amounts of labeled RNA when fed 4-thiouracil (**Fig. 1b**). In contrast, 4-thiouracil–fed larvae or adult flies containing both *GAL4* and *UAS-UPRT* transgenes expressed UPRT in the cell types containing GAL4 (data not shown) and showed robust RNA labeling (**Fig. 1b**). Similarly, embryos soaked in 4-thiouracil–containing medium had robust RNA labeling only when both *UAS-UPRT* and *GAL4* were present (**Fig. 1c**). We conclude that the combination of UPRT and 4-thiouracil can be used to biosynthetically label RNA in *Drosophila* embryos, larvae and adults.

To determine the limits of sensitivity, we fed larvae 4-thiouracil and expressed UPRT in different subsets of the larval brain. When UPRT was expressed in about 2,000 neurons in the entire larva, we detected small amounts of labeled RNA (**Fig. 1b**). When we reduced the number of UPRT-expressing cells to about 250 neural progenitors per larva, we observed no detectable signal over background (**Fig. 1d**). However, a simple dissection of the intact larval brain before RNA purification yielded excellent signal and dramatically reduced background RNA labeling (**Fig. 1d**). We conclude that the amount of TU-tagged RNA correlated well with the number of cells expressing UPRT; that simple tissue isolation can reduce background and increase sensitivity; and that non–central nervous system tissue contributed to nearly all of the low-level background RNA labeling in larvae.

To confirm that thio-labeled RNA was from UPRT-expressing cells and demonstrate the utility of TU-tagging for cell type–specific RNA isolation, we purified TU-tagged and untagged RNA and compared them using microarrays. We isolated RNA from larval glia, which are present in low numbers, are highly dispersed and have a complex cell morphology (**Fig. 2a**), making them one of the most difficult cell types to isolate by dissection or dissociation methods. We used reversed polarity (*repo*)-*GAL4* to drive expression of *UAS-UPRT* specifically in glial cells of the larval brain (**Fig. 2a**). We purified TU-tagged and untagged RNA from 72–96-h larval brains and hybridized them to custom Agilent microarrays

[1]Institute of Neuroscience, Institute of Molecular Biology, Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon, USA. [2]Present address: School of Natural Sciences, University of California, Merced, California, USA. Correspondence should be addressed to C.Q.D. (cdoe@uoneuro.uoregon.edu).

RECEIVED 17 FEBRUARY; ACCEPTED 10 APRIL; PUBLISHED ONLINE 10 MAY 2009; DOI:10.1038/NMETH.1329

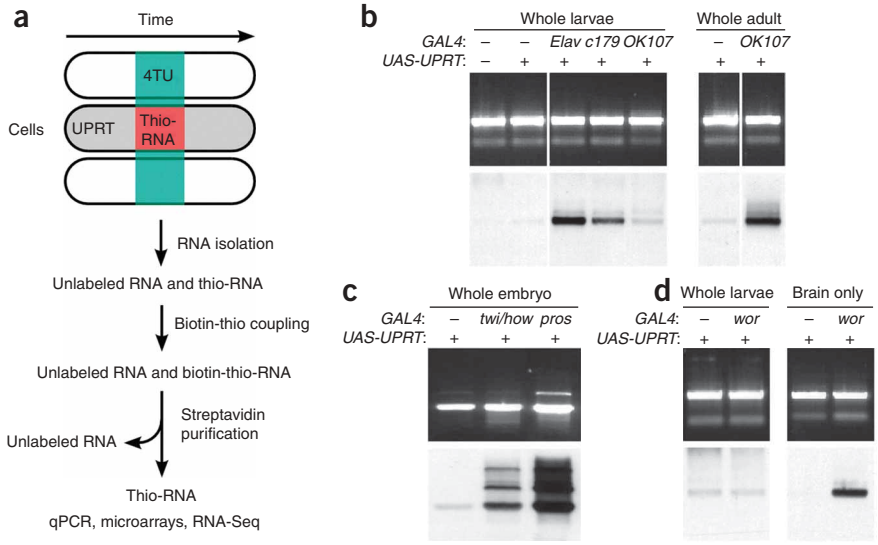

Figure 1 | TU-tagging: overview and cell type–specific labeling. (**a**) TU-tagging procedure. (**b**) TU-tagging in larvae and adult flies. RNA from larvae expressing GAL4 in none of the cells (–), all neurons (*elav*), muscle cells (*c179*) or the mushroom body (*OK107*) and either with or without *UAS-UPRT*, and from adults of the indicated genotypes were electrophoresed and stained with ethidium bromide to detect all RNA (top) and streptavidin-horseradish peroxidase to detect thio-RNA (bottom). (**c**) The 0–16-h embryos of the indicated genotypes were treated with 4-thiouracil for 2 h. RNA was analyzed as indicated above. (**d**) Comparison of RNA labeling before tissue isolation (whole larvae) and after tissue isolation (brain only). *twi/how*, mesoderm/muscle; *pros*, neural; and *wor*, neuroblasts.

(Online Methods). We detected signal for 7,354 of the 14,141 genes present on the microarray. If TU-tagged RNA was primarily from UPRT-expressing cells, known larval glia–specific genes should be among the most enriched in the TU-tagged relative to untagged RNA. There are four genes known to be expressed specifically in larval glia but not neurons or trachea of the brain. All four were enriched greater than twofold in the TU-tagged RNA (**Fig. 2b**) and three of the four were among the 5% most enriched genes *ana* (3.54-fold, top 0.3%, 19/7,354), *repo* (2.85-fold, top 1.6%, 115/7,354) and *moody* (2.75-fold, top 1.9%, 141/7,354), which are expressed in all larval glia[11,12], and *ced-6* (2.13-fold, top 7.6%, 561/7,354), which is expressed in a subset of larval glia[13]. Thus, known larval glia–specific genes were enriched in the TU-tagged RNA. We conclude that TU-tagging can effectively isolate glia-specific RNA from whole brain tissue without prior cell dissociation.

We next tested whether the number of uracils in a transcript influenced the extent of enrichment we observed, because long transcripts containing many uracils are expected to be labeled at a higher frequency than short transcripts with few uracils. Thus we plotted the number of uracils in a transcript against the observed microarray ratio for each transcript (Online Methods). There was a notable positive correlation (**Fig. 2c**). Transcripts with many uracils were often among the most enriched transcripts overall, even when they were not enriched relative to other transcripts with a similar number of uracils (**Fig. 2c**), whereas transcripts with few uracils that were clearly enriched relative to other transcripts with a similar number of uracils were unlikely to be among the most enriched transcripts overall (**Fig. 2c**). We conclude that the number of uracils in a transcript is a source of bias in TU-tagging experiments.

To remove this bias on the microarray results, we used the regression equation to calculate the expected Tu-tagged to untagged ratio and subtracted it from the observed ratio for each transcript (Online Methods and **Supplementary Table 1** online). This normalization procedure successfully removed the bias in data that resulted from the number of uracils in a transcript (**Fig. 2d**) and had a large impact on which genes were present in the top 1% or 5% most enriched genes (**Fig. 2e**). We conclude that enrichment bias owing to the number of uracils in a transcript can be removed using a simple normalization procedure and that normalization has a large impact on which genes are considered most enriched.

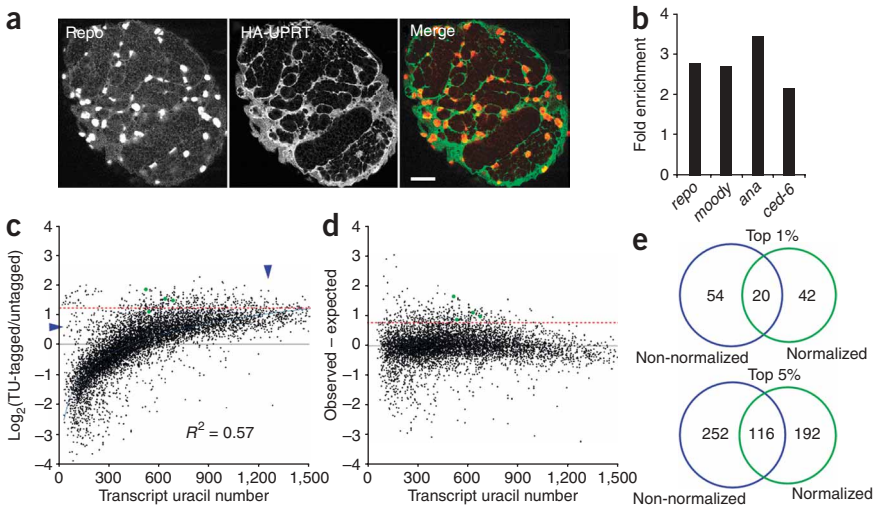

Figure 2 | Cell type–specific RNA isolation and analysis. (**a**) Single confocal section through a brain lobe from a 96-h after larval hatching (ALH) larvae expressing hemagglutinin-tagged UPRT (HA-UPRT) in all glia (*repo-GAL4 UAS-HA-UPRT*) stained for HA-UPRT (detected with an HA antibody) and glial nuclei (detected with Repo antibody). Scale bar, 20 μm (**b**) Fold enrichment of the indicated larval glia–specific genes. See **Supplementary Table 1** for a list of all enriched genes. (**c**) Average microarray ratios from two glia TU-tagging experiments plotted against the number of uracils in the transcript. Dashed red line indicates cutoff for top 5% enriched genes. Green dots, previously known larval glia–specific genes; vertical arrowhead, possible false positives; and horizontal arrowhead, possible false negatives. (**d**) TU-tagging microarray ratios after normalization and removal of transcripts with missing untranslated region annotations. (**e**) Comparison of top 1% and 5% enriched genes before and after normalization for uracil number.

We next determined how the normalization affected the enrichment of the known larval glia–specific genes. Without normalization three of four previously known larval glia–specific genes were within the top 5% of enriched genes (**Fig. 2c**). After normalization all four were within the top 3.2% of enriched genes: *ana*, top 0.2% (12/6,167); *repo*, top 1.6% (99/6,167); *moody*, top 2.4% (150/6,167); and *ced-6*, top 3.2% (235/6,167) (**Fig. 2d**). Although the characterization of new glia genes was beyond the scope of this work, these results suggest that the other highly enriched genes in this dataset are excellent candidates for regulating aspects of larval glia biology. We conclude that normalizing for transcript uracil number improves TU-tagging data analysis and that normalization should be useful for other 4-thiouracil/UPRT–based methods.

An important property of the TU-tagging method is that only newly synthesized RNAs are labeled. Thus, the percentage of labeled cellular RNA will depend on the duration of labeling. Although long labeling periods should work well for isolating the majority of RNA present in a particular cell type, short labeling periods could be used to detect changes in gene expression at successive time points in specific cell types, because newly synthesized RNA could be separated from bulk cellular RNA. This would be useful for studying rapid changes in gene expression after a particular developmental or physiological event.

The TU-tagging method is likely to work well in other systems including vertebrates. There is very little UPRT-independent 'background' incorporation of 4-thiouracil into RNA in mouse or human cell culture lines[9,14]. Spatial control of UPRT in vertebrates could be achieved by using transgenes, as in *Drosophila*, or by retroviral delivery, electroporation or mRNA injection. Thus, at least *Drosophila*, mice and humans appear suitable for TU-tagging experiments. It is likely that TU-tagging can be used for cell type–specific RNA isolation in many multicellular organisms and should be particularly useful for the study of development, neurobiology and disease.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank C. Canestro, T. Herman, Z. Lewis and S. O'Rourke for comments on the manuscript, and C. Cabernard for help with Imaris. M.R.M. was supported by a US National Science Foundation predoctoral fellowship, M.D.C. was supported by a National Institutes of Health National Research Service Award postdoctoral fellowship, and C.Q.D. was supported by National Institutes of Health HD27056 and the Howard Hughes Medical Institute.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Tomancak, P. *et al. Genome Biol.* **8**, R145 (2007).
2. Wang, Z., Gerstein, M. & Snyder, M. *Nat. Rev. Genet.* **10**, 57–63 (2009).
3. Schadt, E.E., Monks, S.A. & Friend, S.H. *Biochem. Soc. Trans.* **31**, 437–443 (2003).
4. Roy, P.J. *et al. Nature* **418**, 975–979 (2002).
5. Tanke, H.J. & van der Keur, M. *Trends. Biotechnol.* **11**, 55–62 (1993).
6. Doyle, J.P. *et al. Cell* **135**, 749–762 (2008).
7. Heiman, M. *et al Cel* **135**, 738–748 (2008).
8. Nelson, S.B., Hempel, C. & Sugino, K. *Curr. Opin. Neurobiol.* **16**, 571–576 (2006).
9. Cleary, M.D. *et al. Nat. Biotechnol.* **23**, 232–237 (2005).
10. Dolken, L. *et al. RNA* **14**, 1959–1972 (2008).
11. Bainton, R.J. *et al. Cell* **123**, 145–156 (2005).
12. Ebens, A.J. *et al. Cell* **74**, 15–27 (1993).
13. Awasaki, T. *et al. Neuron* **50**, 855–867 (2006).
14. Rinn, J.L. *et al. Genes Dev.* **22**, 303–307 (2008).

ONLINE METHODS

UPRT and 4-thiouracil toxicity tests. We crossed *UAS-UPRT* lines to various *GAL4* lines to express UPRT in several cell types (see below) and ubiquitously (using *tubulin-GAL4*). In no case did we observe lethality, developmental defects or delay in the time to pupariation and eclosion (data not shown). We next monitored development of larvae ubiquitously expressing UPRT when grown on food containing 4-thiouracil (4TU). We observed no effect on central nervous system development (developmental timing or neuroanatomy) in larvae grown on 4TU for 48 h or less; we then used an 8 h or less feeding interval for all experiments. Longer exposure to 4TU (with or without UPRT expression) produced slight developmental delays but did not affect viability.

Fly stocks. We used standard methods to clone the *T. gondii* UPRT coding sequence into pUAST to generate the *UAS-UPRT* and *UAS-HA-UPRT* plasmids, and obtained independent viable insertions on the X, II or III chromosome. Three *UAS-UPRT* lines consistently had low background expression and high GAL4-induced expression: *UAS-HA:UPRT2.1* (chromosome II), *UAS-UPRT3.1* (chromosome III) and *UAS-HA:UPRT3.2* (chromosome III). We crossed one of these three lines to the following previously described *GAL4* lines for all experiments (chromosome in parentheses) (i) *worniu-GAL4(II)* × *UAS-HA:UPRT3.2*; (ii) *OK107-GAL4(IV)* × *UAS-HA:UPRT3.2;* (iii) *c179-GAL4(II)* × *UAS-HA:UPRT3.2*; (iv) *elav-GAL4(III)* × *UAS-HA:UPRT3.2*; (v) *tubulin-GAL4(III)* × *UAS-HA-UPRT3.2*; (vi) *prospero-GAL4(III)* × *UAS-UPRT3.1*; (vii) *twist-GAL4(II);how24B-GAL4(III)* × *UAS-UPRT3.1*; and (viii) *repo-GAL4(III)* × *UAS-HA:UPRT3.2*.

4TU treatment and RNA extraction. To treat embryos with 4TU, the embryos were dechorionated in bleach, washed, rinsed with isopropanol, blotted dry and submerged in octane for 3 min (all in the basket). Embryos were blotted and air dried for ~3 min until soft to touch. Embryos were transferred to Schneider's media containing 1.0 mM 4TU (Sigma-Aldrich) for 2 h at 25 °C or 30 °C, blotted dry, moved using a paintbrush to an Eppendorf tube containing 1× PBS with 1% tween (PT) and centrifuged at 6,000*g* for 30 s. PT was removed and the embryos were homogenized in Trizol and stored at –80 °C until RNA purification. To treat larvae with 4TU, larvae of the desired stage were placed on mocha food caps (20 ml H_20, 0.4 g sucrose, 0.18 g agar, 1 g dextrose and 0.5 g brewers yeast) containing 0.5 mM 4TU for the indicated time at 30 °C, homogenized in Trizol and stored at –80 °C until RNA purification. To treat adult flies with 4TU, they were starved for ~16 h, placed on mocha food containing 1.0 mM 4TU for 6–8 h at 25 °C, homogenized in Trizol and stored at –80 °C until RNA purification. A more detailed protocol is available on request.

Total RNA was extracted from Trizol using standard methods with the following additional steps: an initial centrifugation at 12,000*g* for 10 min at 4 °C to remove insoluble material followed by a 5-min incubation at room temperature (20–25 °C) to ensure complete dissociation of RNA-bound proteins. Only RNA samples with absorbance 260/280 ratios of ≥ 2.0 were used for subsequent biotinylation and purification steps. In all cases, RNA samples were resuspended at a final concentration of ≥0.4 μg μl^{-1}.

Purification of TU-tagged RNA. Detailed methods for biotinylating and purifying thio-tagged RNAs have been published[9,15], and detailed protocols including the most recent improvements are available upon request. Relevant changes to previous protocols are summarized here. Biotinylation of RNA was performed using EZ-Link biotin-HPDP (*N*-(6-(biotinamido)hexyl)-3′-(2′-pyridyldithio)-propionamide; Pierce), as previously described. Biotinylation reactions contained 10 mM Tris (pH 7.4), 1 mM EDTA and 2 μl of a 1 mg ml^{-1} biotin-HPDP solution (in dimethylformamide) per 2 μg of RNA. The reaction volume was adjusted with water so that the concentration of biotin-HPDP was equal to 30% of the final reaction volume. Biotinylation reactions were incubated in the dark for 3 h at 25 °C before RNA precipitation. Biotinylated RNA was detected by blotting and probing with streptavidin-horseradish peroxidase as previously described. Purification of biotinylated TU-tagged RNA was performed as previously described with the following modifications: 2 μl of MPG streptavidin beads (Pure-Biotech) were used per microgram of input RNA. The input RNA was always at a concentration of 0.5 μg μl^{-1}. After blocking with yeast tRNA and washing, beads were resuspended in the input RNA sample plus a volume of MPG buffer (1 M NaCl, 10 mM EDTA and 100 mM Tris (pH 7.4) in RNase-free H_2O) equal to one-third the input RNA volume. Beads plus RNA were incubated at room temperature for 20 min before collecting the non-bound sample and washing with MPG buffer (one 5-min wash at room temperature, two 1-min washes at room temperature, one 1-min wash in 65 °C MPG buffer and one final 1-min wash at room temperature). After the removal of as much MPG buffer as possible, TU-tagged RNA was eluted by incubating the beads for 10 min in freshly prepared 5% 2-mercaptoethanol. RNA was precipitated using isopropanol and linear acrylamide. After resuspending RNA in water, samples were placed in the magnetic stand again to remove any remaining MPG beads. For the purification procedures, input amounts of biotinylated RNA were 14–20 μg.

Microarray analysis. We used the Agilent eArray platform to design a custom *D. melanogaster* oligonucleotide microarray representing 14,141 unique genes from the Flybase release 5.4 genome. Fifty nanograms of TU-tagged and 200 ng untagged RNA were fluorescently labeled and hybridized directly against each other to Agilent microarrays. The microarray experiments were performed according to Agilent's protocol (Version 5.5, February 2007) and scanned using an Axon GenePix 4000B scanner. Fluorescent ratios for each microarray element were recovered and normalized using GenePix Pro 6.0.

Normalizing for transcript uracil number. We downloaded the dmel-all-transcript-r5.4.fasta file from Flybase, which contained the sequences of every predicted transcript from genome release 5.4. We wrote a Perl script to count the number of uracils in a transcript (transcript uracil number) for each transcript. In cases where there was alternative splicing, we averaged the uracil number over the multiple isoforms. This was necessary because, for the most part, our microarray did not distinguish between different isoforms of the same gene. To normalize the data, these uracil counts were plotted against the observed microarray ratios using OpenOffice.org Spreadsheet and the regression equation was determined. This equation was used to calculate the expected ratio for each transcript based on the transcript's uracil number. For each transcript, the normalized ratio was calculated by subtracting the expected ratio from the observed ratio. After plotting the

doi:10.1038/nmeth.1329

initial normalized ratios, we noticed a group of transcripts, all with low uracil numbers, that had very high normalized ratios. Upon investigation we noticed that these transcripts were missing annotations for either or both untranslated regions. Thus, our transcript uracil counts for these transcripts were lower than their actual uracil number, leading to incorrect normalization. Therefore, transcripts with missing untranslated region annotations were excluded from normalization and subsequent analysis. We note that normalization errors owing to transcript misannotation will decrease as transcriptome annotations improve.

15. Zeiner, G.M. *et al. Methods Mol. Biol.* **419**, 135–146 (2008).