Applying thiouracil tagging to mouse transcriptome analysis

Leslie Gay[1–3,5], Kate V Karfilis[1,5], Michael R Miller[1,2,4], Chris Q Doe[1–3] & Kryn Stankunas[1]

[1]Institute of Molecular Biology, University of Oregon, Eugene, Oregon, USA. [2]Institute of Neuroscience, University of Oregon, Eugene, Oregon, USA. [3]Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon, USA. [4]Present address: Department of Animal Science, University of California, Davis, Davis, California, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to C.Q.D. (cdoe@uoneuro.uoregon.edu) or K.S. (kryn@uoregon.edu).

Transcriptional profiling is a powerful approach for studying mouse development, physiology and disease models. Here we describe a protocol for mouse thiouracil tagging (TU tagging), a transcriptome analysis technology that includes *in vivo* covalent labeling, purification and analysis of cell type–specific RNA. TU tagging enables the isolation of RNA from a given cell population of a complex tissue, avoiding transcriptional changes induced by cell isolation trauma, as well as the identification of actively transcribed RNAs and not preexisting transcripts. Therefore, in contrast to other cell-specific transcriptional profiling methods based on the purification of tagged ribosomes or nuclei, TU tagging provides a direct examination of transcriptional regulation. We describe how to (i) deliver 4-thiouracil to transgenic mice to thio-label cell lineage–specific transcripts, (ii) purify TU-tagged RNA and prepare libraries for Illumina sequencing and (iii) follow a straightforward bioinformatics workflow to identify cell type–enriched or differentially expressed genes. Tissue containing TU-tagged RNA can be obtained in 1 d, RNA-seq libraries can be generated within 2 d and, after sequencing, an initial bioinformatics analysis can be completed in 1 additional day.

INTRODUCTION

Development of the protocol

The emergence of microarrays and next-generation sequencing has revolutionized the use of gene expression profiling to understand complex biological processes. However, existing cell-specific transcriptome analysis methods in mice have limitations that restrict the accurate and unbiased characterization of actively transcribed genes. Therefore, we developed a genetic and chemical intersectional technology, known as mouse TU tagging, that provides the *in vivo* labeling of RNA in specific cell types and during defined periods[1]. TU tagging uses the uracil analog 4-thiouracil (4TU) to label RNA *in vivo*. 4TU is only converted into 4-thiouridine and subsequently incorporated into newly transcribed RNA in cells engineered to express *Toxoplasma gondii* uracil phosphoribosyltransferase (UPRT). The thio-RNA is biotinylated and streptavidin beads are used to purify it from total RNA prepared from a complex tissue. The TU-tagged RNA is then analyzed by reverse-transcription (RT)-qPCR, microarrays or next-generation sequencing. TU tagging has been effectively used in cell culture models and *Drosophila* studies for several years[2–5]. Recently, we adapted TU tagging for transgenic mouse studies. We engineered a modular system based around a mouse line that provides Cre recombinase–dependent, spatially restricted expression of a transgene-encoding UPRT (*CA > GFPstop > HA-UPRT*). The timing of 4TU delivery provides temporal control of RNA labeling. The resulting 'TU-tagged' RNA is then transcriptionally profiled by RNA-seq via Illumina sequencing. In our recent report introducing mouse TU tagging, we defined endothelial lineage transcriptomes of the developing brain and heart and characterized the global transcriptome response of splenocytes to acute lipopolysaccharide (LPS) exposure[1].

Applications of the method

Mouse TU tagging is applicable to any project requiring gene expression analyses of specific cell types or for defining immediate transcriptome responses. Spatial control of RNA labeling enables the characterization of cell type–specific expression programs. Temporal control allows the study of transcriptome changes upon, for example, inducing a physiologic response or perturbing the system genetically or chemically. TU tagging can also be used to characterize the differences between a given cell type that is present in a variety of organs, such as our comparison of brain and heart endothelial transcriptomes. One major advantage of TU tagging is that it only covalently labels nascent transcripts, and therefore it is ideally suited for detecting dynamic changes in gene expression. This attractive feature has been well documented in TU-tagging experiments on cell lines[3,4] and in mouse tissue[1]. Controlled labeling also may allow pulse-chase experiments to study, for example, RNA stability and RNA splicing kinetics. TU tagging should also effectively integrate with photoactivatable ribonucleoside–enhanced cross-linking and immunoprecipitation (PAR-CLIP)[6] to define tissue-specific interactions between RNA-binding proteins and RNA molecules. Mouse TU tagging can also be undertaken without a Cre line. We have produced a constitutively expressed UPRT mouse line that is useful for studies in which only temporal labeling is desired or for generating UPRT-chimeric mice by cell transplantations (e.g., bone marrow transplants)[1].

Comparison with other methods

Current technologies used to isolate cells for transcriptome analysis, including FACS, immunopanning and manual or laser capture dissection, are effective but have functional limitations. FACS exerts damaging physical forces on cells, it takes time and it can result in the loss of RNA from processes of morphologically complex cell types such as neurons and glia. Immunopanning depends on the availability of antibodies against surface antigens and requires placing cells in unnatural environments for extended periods with likely effects on gene expression. Manual dissections are not always feasible; they are prone to operator variability, and they rarely allow high-purity separations.

Figure 1 | Flowchart of the mouse TU-tagging protocol. This chart outlines the entire procedure from the delivery of 4TU to the analysis of Illumina sequencing data by using DESeq (Type I experiment). The black-bordered boxes indicate the three major stages of the protocol. The gray-shaded boxes highlight distinct substages with step numbers corresponding to those in the PROCEDURE. Blue wavy lines represent RNA produced from two mixed cell populations, one of which is UPRT expressing (lighter blue color). The orange circles indicate the incorporation of thiolated uracils, which are then biotinylated (B) *in vitro*.
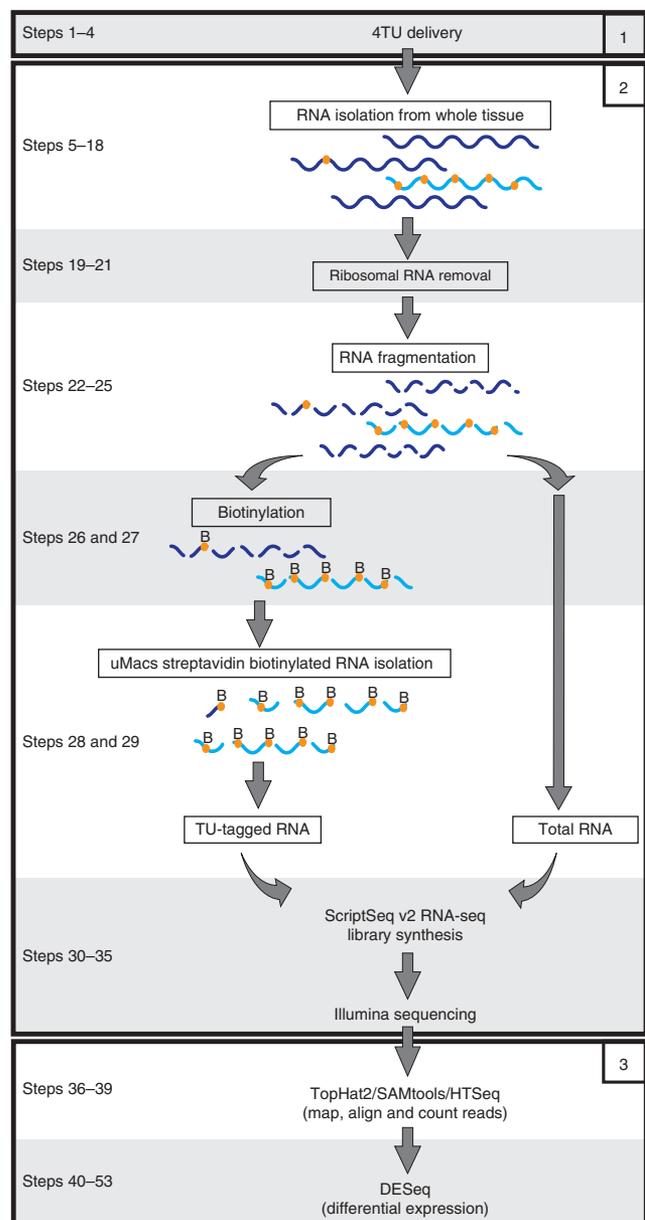

Laser capture microdissection typically uses fixed, sectioned tissue, limiting both the quality and quantity of recovered RNA. All of the above approaches lead to isolation of the bulk pool of RNA, irrespective of when the RNA was transcribed. TU tagging avoids these limitations by covalently tagging cell type–specific RNA *in vivo* to enable its subsequent purification from total RNA prepared from a complex tissue.

Alternative genetic-based methods that use *in vivo* labeling for cell type–specific transcriptome studies include INTACT, TRAP and Ribo-tag[7–10]. INTACT provides for the purification of cell type–specific nuclei; therefore, isolated RNA exclusively represents the accumulated nuclear pool. However, INTACT has the advantage of allowing simultaneous chromatin landscape or transcription factor association studies (by ChIP-seq), and therefore it provides extra information on gene regulatory networks. TRAP and Ribo-tag examine ribosomal protein-bound mRNA and are therefore especially useful as surrogate approaches to study global protein translation. These methods may not detect noncoding RNAs (ncRNA) including miRNAs. In contrast, TU tagging labels all newly transcribed RNA and thus enables simultaneous studies of mRNA and ncRNA. Mouse TU tagging is also designed as a modular system, being readily combined with the many existing Cre lines and usually eliminating the need to develop new transgenic lines before experiments can be pursued. As each of these alternative methods, including TU tagging, have unique advantages, a researcher should carefully consider which approach would most efficiently answer their questions of interest before pursuing a new study.

Experimental design
There are three stages to the presented mouse TU-tagging protocol (**Fig. 1**). The first stage describes the delivery of 4TU to initiate *in vivo* RNA labeling. The second stage describes, in detail, the molecular biology protocol for isolating TU-tagged RNA and preparing libraries for Illumina sequencing. The third stage outlines one straightforward bioinformatics approach to quickly process raw Illumina data into a table of enriched or differentially expressed TU-tagged transcripts. Successful completion of each stage will require that the research team have at least some experience in each of mouse genetics, molecular biology and statistical analysis of large data sets.

TU tagging takes advantage of the many existing cell-specific and tissue-specific Cre lines and our recently developed *CA > GFPstop > HA-UPRT* transgene to direct spatially restricted UPRT expression in the desired Cre-positive cell lineage[1]. This transgene incorporates a broadly expressed, constitutively active (CA) chicken β-actin/CMV promoter driving a *loxP-GFP-3stop-loxP* cassette, followed by a hemagglutinin (HA) epitope-tagged *UPRT* cDNA. The GFP-3stop cassette includes three SV40 polyadenylation sequences to prevent transcription of HA-UPRT until the cassette is excised by Cre activity. In mice carrying both transgenes, UPRT becomes permanently expressed in the Cre-expressing cell lineage. By GFP immunostaining, we demonstrated that the *CA > GFPstop > HA-UPRT* has widespread promoter activity in embryonic and postnatal tissues[1]. Nevertheless, Cre-induced expression of HA-UPRT should be confirmed before undertaking TU tagging in a new cell type. The *tissue-specific:Cre* line should be chosen carefully, as UPRT will be permanently expressed in any cell lineage that expressed Cre at any point in its development. Where available, the use of a tamoxifen-inducible Cre-ER line may facilitate tighter control of cell type specificity[11]. Tissues are sectioned and immunostained with HA antibodies (to detect UPRT) and a cell-specific antibody that labels the cells in which UPRT induction is expected[1]. For example, in our studies characterizing endothelial transcriptomes, we used Pecam1-specific antibodies to label all endothelial cells. Only cells in the desired Cre-expressing lineage should stain with HA antibodies, with considerable or complete overlap with the chosen cell-type marker.

For reliable and reproducible TU-tagging experiments, both the *UPRT* and *Cre* transgene copy numbers should be consistent between experimental repeats. Therefore, for most experiments, it is most convenient to use double-heterozygous *UPRT;Cre* animals. If possible, interbreed homozygous *CA > GFPstop > HA-UPRT*, which are viable and fertile, with homozygous *Cre* animals. All resulting progeny will be heterozygous for both the *CA > GFPstop > HA-UPRT* and *Cre* transgenes, requiring no genotyping and, as is especially useful for embryonic studies, allowing immediate pooling of samples (except for sex-specific studies). Mice are then injected with 4TU at a desired postnatal age or at a desired stage of pregnancy to provide temporal control of TU tagging.

There are two distinct TU-tagging experimental designs. In 'Type I' experiments, transcript levels are compared between the TU-tagged RNA ('pure') and the total RNA ('total') from which the TU-tagged RNA was purified. This approach reports on how enriched each transcript is within the UPRT-expressing cells compared with the total tissue from which the RNA was prepared. Therefore, this experimental design is ideal for observational studies when the goal is to characterize a given cell type's unique active transcriptome during a defined period of 4TU exposure. The maximum fold enrichment of a transcript is the inverse of the cell type's fractional representation within the starting material. Perfect enrichment, however, is never achieved, as there is always limited background labeling in non-UPRT-expressing cells[1]. The choice of starting material is of paramount importance. A careful organ or tissue dissection that enriches for the cell type of interest will avoid failure to detect cell-specific transcripts expressed at low levels. However, if the cell type being studied comprises more than ~33% of the starting material, the maximum fold enrichment will be low and therefore statistical significance will be more difficult to achieve. Note that short-lived transcripts will be over-represented in the pure compared with the total RNA, which could generate a bias when defining genes with cell type–enriched expression.

An alternative control for nascent RNA bias is to expose control mice to 4-thiouridine (4TUd), in parallel with the littermates exposed to 4TU. Whereas 4TU requires UPRT to be incorporated into nascent RNA, 4TUd will be added to newly synthesized RNA in all cells[12]. Thus, a comparison can be made between pure RNA from experimental mice with cell type–specific 4TU and mice with all nascent RNA thio-labeled 4TUd. For example, *tissue-specific:Cre; CA > GFPstop > HA-UPRT* littermate mice are injected with either 4TU or 4TUd, and then the organ of interest is collected after a fixed period. TU-tagged RNA is purified and sequenced in each case. Genes with relatively more reads in the 4TU-exposed mice are more abundantly expressed in the cell type in which Cre has been active. The workflow for a Type I experiment using this approach would be similar to the 'Type II' experiment described below (only pure TU-tagged RNA is sequenced in each case). Regardless, the output of a Type I experiment is a table of statistically significant enriched and depleted transcripts in the UPRT-expressing cells.

In Type II experiments, a transcriptome comparison is performed between two or more purified, TU-tagged RNA samples isolated from animals that were differentially treated. As examples, the differences could be mutant versus control mice, before and after a drug delivery or induced physiologic response

(e.g., the aforementioned spleen LPS-response experiment), or a comparison between two or more treatments/manipulations. Type II experiments are also useful when comparing transcriptomes of the same cell type in different organs of the same animal (e.g., our organ-specific endothelial experiments). The output in Type II experiments is a table of differentially expressed genes. For both Type I and Type II experiments, the molecular biology protocol is nearly identical, with the major difference being the samples prepared for sequencing (Type I = pure versus total, Type II = pure versus pure). The bioinformatics analysis outlined here conveniently uses the identical workflow for both experimental types, but the output is interpreted as either fold enrichment (Type I) or fold change (Type II). The bioinformatics analysis example presented is from a Type I experiment characterizing postnatal day 6 (P6) brain endothelial and macrophage/microglia transcriptomes[1].

Regardless of design, it is preferable to predefine a set of tissue-specific positive control transcripts. This will empower the analysis and allow an immediate assessment of the experiment's data quality. For example, in our endothelial transcriptome experiments, we used 13 well-known positive control endothelial genes: *Cdh5 (VEcad), Cd34, Egfl7, Emcn, Esam, Ets1, Flt1, Kdr (Flk1/VEGFR2), Nos3, Pecam1, Tek (Tie2), Tie1* and *Thsd1*. The positive control transcripts should cluster among the most enriched in a Type I experiment. For *de novo* transcriptome definition experiments, a gene ontology term analysis should be performed (using, e.g., Princeton Gene Ontology Tools (http://go.princeton.edu/)) on the full set of enriched transcripts to verify that expected GO terms for the studied cell type are over-represented. The statistically enriched transcripts can also be examined by using a gene expression database such as Eurexpress[13] (http://www.eurexpress.org/) to confirm the preponderance of enriched transcripts expressed in the Cre-labeled cell lineage. For all experiments, at least two biological replicates of each sample (e.g., experimental and control) should be sequenced.

The workflow for a typical TU-tagging experiment begins with 4TU delivery (Stage 1; **Fig. 1**). Typically, the *tissue-specific:Cre* and *CA > GFPstop > HA-UPRT* double transgenic adults or postnatal pups are injected with 4TU and collected 4–6 h later. We have observed a drop in labeling if the tissue is collected ≥12 h after injection, enabling pulse-chase experiments to study transcriptome dynamics. For embryonic studies, pregnant mice carrying double transgenic embryos are exposed to 4TU for at least 6 h. Embryonic studies use a longer 4TU exposure period because of an anticipated time lag for the compound to cross the placental barrier into embryonic or fetal circulation. The desired tissue is collected and total RNA is isolated in ~3 h, depending on the number of tissue samples collected. Either two biological replicates (Type I experiments) or two experimental and two matched control RNA samples (Type II experiments) can be comfortably processed at one time by the subsequent molecular biology steps to Illumina sequencing (Stage 2).

Stage 2 (**Fig. 1**) requires at least 10 μg of RNA from each sample. Less RNA may be used, but it may yield inconsistent outcomes. First, we deplete the RNA from rRNA (rRNA) by using Ribo-Zero kits, as described in the PROCEDURE. Alternative approaches include an RNase H method[14] that is likely to be less expensive over the long run and has worked well in pilot experiments. Both the Ribo-Zero and RNase H methods can recover pre-mRNAs, mRNAs and ncRNAs, and they can therefore provide a broad

report of transcriptomes. However, instead of rRNA depletion, poly(A) mRNA can be purified or oligo(dT) can be used to prime first-strand cDNA synthesis from non-rRNA-depleted RNA samples. These two methods fail to recover pre-mRNA and ncRNA, but they produce a higher fraction of exon reads, which are scored by conventional bioinformatics approaches including the method outlined in this protocol.

The RNA is then fragmented (optional for Type II experiments). RNA fragmentation increases the relative number of thio-labeled RNA fragments from UPRT-expressing cells compared with spuriously but infrequently labeled transcripts from non-UPRT-expressing cells. Fragmenting the RNA before purification not only reduces background but will also remove the capture bias for large (uridine-rich) RNA molecules that was observed in earlier experiments[2]. RNA fragmentation is not required for most Type II experiments because usually all the pure RNA samples will have similar background and large-transcript bias that is negated by the bioinformatics analysis. TU-tagged RNA is then biotinylated and affinity-purified by using streptavidin-conjugated magnetic beads. Fragmentation, biotinylation and purification together take ~6 h. Note that TU-tagged RNA concentrations are usually too low to be quantified, and thus the entire amount of TU-tagged RNA of each sample is used for cDNA synthesis, amplification and the addition of distinct Illumina index adaptors. For Type I experiments, an aliquot of total fragmented RNA is also prepared for sequencing. The indexed samples are pooled at equimolar concentrations and stored at −20 °C until Illumina sequencing.

Upon the return of the raw RNA-sequencing data, a bioinformatics analysis is performed (Stage 3). Many different analysis options are available; users should consider their experimental needs when designing a workflow, and they should consult with bioinformatics experts as necessary. Here we describe one straightforward workflow that uses popular software packages to characterize mRNAs. This approach efficiently provides a basic analysis while requiring only minimal knowledge of Unix commands and the R statistical programming language. Illumina sequencing provides raw reads data as a series of fastq-formatted files. These files are concatenated and processed on a Unix server with multicore processing. By using the alignment and mapping tools TopHat2 (ref. 15), SAMtools[16] and HTSeq[17], raw reads mapping to exons are assigned to each gene within a mouse reference genome, producing a tab-delimited text file that contains the total number of counts (reads) for each gene. Each file is downloaded to a personal computer and opened in Excel (or other spreadsheet software). The counts data for each data set are combined into a single file and resaved as a tab-delimited file for import into R. We suggest using the freely available RStudio package (http://www.rstudio.com/) to provide a convenient R workspace. We suggest the DESeq package[18] as one tool for identifying enriched (Type I) or differentially expressed (Type II) genes. Briefly, DESeq uses count-based data and a negative binomial distribution to statistically determine genes with differential expression levels between data sets, including replicates. DESeq returns a table with adjusted *P* values (Benjamini-Hochberg correction method) indicating whether transcripts from each corresponding gene are differentially present between data sets. A table showing only statistically different genes is generated, which can be further analyzed by R or exported back to Excel or other software. As mentioned, some experience with Illumina data, Unix commands and R will facilitate following this protocol. Available TopHat2/HTSeq/DESeq manuals and vignettes should be consulted to help less-experienced researchers to become familiar with the programs and their parameters.

Limitations

Applications of mouse TU tagging by using the *CA > GFPstop > HA-UPRT* mice are limited by the availability of an appropriate Cre-recombinase-expressing transgenic line. Certain experiments may be difficult or impossible if there are insufficient numbers of UPRT-expressing cells in the tissue sample used to prepare the total RNA. We have not determined the lower limit, but our chimeric mouse experiments suggest that TU-tagged RNA can be recovered if even 1% of the cells express UPRT[1]. Although initial experiences using Cre/*lox*-driven mouse TU tagging agree with this result, researchers should proceed with caution when performing TU tagging in rare cell types. Finally, transcripts with unusually low turnover may be infrequently labeled or not labeled during short 4TU exposure periods, and therefore they are excluded from tissue-specific expression analyses.

MATERIALS
REAGENTS
- Transgenic, tissue-specific Cre recombinase mice, e.g., *Tie2:Cre* (*B6.Cg-Tg(Tek-cre)1Ywa/J* (The Jackson Laboratory, stock no. 008863)
- Transgenic *CA > GFPstop > HA-UPRT* mice (The Jackson Laboratory, stock no. 021469) **❗ CAUTION** All mouse procedures must be authorized by your institutional animal care and use committee in compliance with institutional and governmental animal care guidelines.
- DMSO (Sigma, cat. no. D1435)
- Corn oil (Sigma, cat. no. C8267)
- 4TU (Sigma, cat. no. 440736)
- TRIzol reagent (Invitrogen, cat. no. 15596-026) **❗ CAUTION** TRIzol is toxic and should be used in a fume hood; wear gloves and a lab coat.
- Chloroform (Macron Fine Chemicals, cat. no. 444004)
  ❗ CAUTION Chloroform is toxic and should be used in a fume hood; wear gloves and a lab coat.
- Isopropyl alcohol (Macron Fine Chemicals, cat. no. MK303202)
- Ethanol (Pharmco Aaper, cat. no. 11100020G)
- RNase-free water (H$_2$O; Ambion, cat. no. AM9938)

- Tris, 1 M, pH 8.0, RNase free (Ambion, cat. no. AM9855G)
- EDTA, 0.5 M, pH 8.0, RNase free (Ambion, cat. no. AM9260G)
- TURBO DNase (Invitrogen, cat. no. AM2238)
- RNeasy mini kit (Qiagen, cat. no. 74101)
- Qubit RNA BR assay kit (Invitrogen, cat. no. Q10210)
- Qubit dsDNA HS assay kit (Invitrogen, cat. no. Q32854)
- Qubit assay tubes (Invitrogen, cat. no. Q32856)
- NEBNext magnesium RNA fragmentation module (NEB, cat. no. E6150S)
- Ribo-Zero magnetic kit (Epicentre, cat. no. MRZH11124)
- *N,N*-Dimethylformamide (Sigma, cat. no. D4551)
- EZ-Link biotin-HDPD (Thermo Scientific, cat. no. 21341)
- µMacs streptavidin kit (Miltenyi Biotec, cat. no. 130-074-101)
- 2-Mercaptoethanol (Sigma, cat. no. M3148) **❗ CAUTION** 2-Mercaptoethanol is toxic and should only be used in a fume hood. Gloves and a lab coat should be worn when you are handling it. Avoid inhalation.
- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63880)
- RNeasy MinElute (Qiagen, cat. no. 74204)
- MinElute reaction cleanup kit (Qiagen, cat. no. 28204)
- Tween-20 (Sigma, cat. no. P7949)
- ScriptSeq v2 RNA-seq library preparation kit (Epicentre, cat. no. SSV21106)

PROTOCOL

- FailSafe PCR enzyme mix (Epicentre, cat. no. FSE51100)
- ScriptSeq index PCR primers (Epicentre, cat. no. RSBC10948 (set 1))

EQUIPMENT
- Low-adhesion RNase-free microcentrifuge tubes, 1.5 ml (USA Scientific, cat. no. 1415-2600)
- Centrifuge tubes, 15 ml (Corning, cat. no. 430791)
- Thin-walled PCR tubes, 0.2 ml (USA Scientific, cat. no. 1402-4300)
- Tuberculin syringe, 1 ml, with 26 G × 3/8 inch detachable needle (VWR, cat. no. BD309625)
- Nalgene dewar, 1 liter (Sigma, cat. no. F9401)
- Refrigerated centrifuge (Eppendorf, cat. no. 5417R)
- Kontes pellet pestle grinder (VWR, cat. no. KT749540-0000)
- Pestle, RNase free (VWR, cat. no. KT749521-1590)
- Magnetic separation rack, 6 tubes (Cell Signaling, cat. no. 7017)
- MACS MultiStand (Miltenyi Biotec, cat. no. 017383)
- 6S Super Magnet plate (Alpaqua Engineering, cat. no. A001322)
- Qubit 2.0 fluorometer (Invitrogen, cat. no. Q32871)
- Agilent bioanalyzer 2100 (Agilent)
- Illumina HiSeq 2000 (Illumina)
- **Unix server, installed software**
- FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- TopHat2 (http://tophat.cbcb.umd.edu/)
- Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml)
- SAMtools (http://samtools.sourceforge.net/)
- HTSeq (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html)

- **Personal computer, installed software**
- R (http://www.r-project.org/)
- R Studio (http://www.rstudio.com/)
- Microsoft Excel (http://office.microsoft.com/en-us/excel/)

REAGENT SETUP
4TU Make a stock solution by dissolving 200 mg of 4TU in 1 ml of DMSO, which can be stored for at least 2 months at −20 °C. On the day of injection, dilute the stock at a 1:4 ratio in corn oil to a concentration of 50 mg ml^{-1}, and vortex it vigorously. Note that phase separation will occur if the solution is left standing for an extended period. Therefore, revortex the 4TU immediately before injection.
Biotin Dissolve 10 mg of EZ-Link Biotin-HPDP in 10 ml of *N,N*-dimethylformamide. Prepare 250-μl aliquots and store them in a covered container at −20 °C for up to 6 months.
Tween-20, 10% (vol/vol) Pipet 9 ml of RNase-free water into a 15-ml tube. Add 1 ml of Tween-20 to a total volume of 10 ml. Mix the solution well and store it in the dark.
2-Mercaptoethanol, 100 mM Freshly prepare 2-mercaptoethanol immediately before use. Mix 3.5 μl of 2-mercaptoethanol with 500 μl of RNase-free dH$_2$O. Preheat the solution to 80 °C.
Tris-EDTA (TE) Make a 10× TE stock solution in a 15-ml tube by mixing 1 ml of RNase-free 1M Tris-HCl (pH 8.0), 200 μl of RNase-free 0.5 M EDTA (pH 8.0) and 8.8 ml of RNase-free dH$_2$O. Dilute 10× TE to a 1:10 ratio with RNase-free dH$_2$O to make a 1× TE working solution.

PROCEDURE
4TU delivery ● **TIMING 5 h for adults or P6 pups, and 7 h for embryos**
1| Weigh each mouse and calculate the volume of the prepared 4TU solution required to deliver 400 mg kg^{-1} (e.g., 200 μl of a 50 mg ml^{-1} 4TU solution for a 25-g mouse). Lower 4TU doses generate less labeling; higher doses are not practical and are potentially toxic.

2| Inject 4TU intraperitoneally by using a tuberculin syringe and a 27-G needle.

3| Allow at least 4 h of 4TU exposure before tissue collection for adults and pups and 6 h for pregnant females.
▲ **CRITICAL STEP** Exposure time can be varied depending on the experiment being performed. Labeling can be detected in as little as 2 h after 4TU exposure to postnatal mice, but with considerably reduced yield. We have not defined the minimum exposure period for embryonic studies. We observe decreased labeling when waiting longer than 12 h after 4TU injection.

4| Dissect out the organ or tissue of interest, transfer it immediately into a 1.5-ml microcentrifuge tube and flash-freeze it in liquid nitrogen. Continue with RNA preparation or store it at −80 °C.
■ **PAUSE POINT** Frozen tissue can be stored at −80 °C indefinitely.

RNA purification ● **TIMING 2–3 h**
▲ **CRITICAL** From this point on, care should be taken to maintain RNase-free conditions. Benchtop surfaces should be thoroughly cleaned, RNase-free tubes and tips should be used and gloves should be worn at all times.

5| Precool the centrifuge to 4 °C.

6| Homogenize the tissue in a 1.5-ml tube by using a Kontes pestle and 500 μl of TRIzol per 100 mg of tissue until it is completely solubilized.
! **CAUTION** TRIzol is toxic and should be handled in a fume hood. Gloves and a lab coat should be worn.
? **TROUBLESHOOTING**

7| Add 500 μl of additional TRIzol and vortex the tissue mixture. Incubate the tube for 5 min at room temperature (~22 °C).

8| Add 200 μl of chloroform and vortex it for 15 s. Incubate the mixture for 2–3 min at room temperature.
! **CAUTION** Wear gloves and use a fume hood when you are working with chloroform.
▲ **CRITICAL STEP** The extended chloroform vortex is important for high RNA yield and quality.

9| Centrifuge the tube at 12,000g for 15 min at 4 °C.

10| Transfer the upper aqueous phase to a new tube and add 500 µl of isopropanol. Incubate the tube for 10 min at room temperature.

11| Centrifuge the tube at 20,000g for 10 min at 4 °C.

12| Decant the liquid and wash with 1 ml of 75% (vol/vol) ethanol.

13| Centrifuge the tube at 20,000g for 5 min at 4 °C. Decant the liquid.

14| Centrifuge the tube at 20,000g for 2 min to remove excess ethanol. Carefully pipet the remaining liquid away from the pellet and place the tube upside down on a laboratory tissue for 2–3 min to dry.
▲ **CRITICAL STEP** Do not overdry the RNA because this will make it difficult to resolubilize.

15| Resuspend the pellet in 50–100 µl of RNase-free H_2O. Dilute the RNA to a concentration of ~200 ng µl^{-1}. The RNA concentration should be determined by spectrophotometry (e.g., using a NanoDrop) or with a Qubit fluorometer (see Step 17).

16| Treat the RNA with DNase to remove residual genomic DNA contamination. We recommend using TURBO DNase according to the manufacturer's protocol.

17| Dilute 2 µl of the total RNA in 98 µl of RNase-free H_2O. Quantify the RNA concentration with a Qubit RNA BR assay kit and a Qubit system according to the manufacturer's directions.

18| Determine the integrity of the RNA samples by using an Agilent Bioanalyzer. The RNA integrity number should be >8.0.
▲ **CRITICAL STEP** RNA should be aliquotted to avoid repeated freezing and thawing.
? TROUBLESHOOTING
■ **PAUSE POINT** (end of day 1) RNA should be frozen and stored at −80 °C overnight or until you are ready to proceed.

Removal of rRNA ● TIMING 2 h
▲ **CRITICAL** At this point, there are several alternatives for the removal of rRNA. Our standard approach involves Ribo-Zero kits and is described below. There are two alternative approaches to consider (see Experimental design): one is to use RNase H to remove rRNA[14]; the other is to purify poly(A) mRNA or use oligo(dT) to prime first-strand cDNA synthesis. Of note, the bioinformatics workflow described in this protocol only analyzes exon-mapped reads, and thus it is well-suited for the poly(A) purification or oligo(dT) priming approaches.

19| Remove rRNA from 10–100 µg of total RNA, using the Ribo-Zero Magnetic kit and a 6-tube magnetic separation rack, according to the manufacturer's instructions.
▲ **CRITICAL STEP** The use of a maximum of 100 µg of RNA will increase the yield of TU-tagged RNA, which is especially important if the UPRT-expressing cells are rare, but at the expense of requiring the use of multiple Ribo-Zero rRNA removal reactions per sample.

20| Clean the rRNA-depleted total RNA by using an RNeasy mini column according to the manufacturer's protocol. When starting with more than 10 µg of RNA, the combined volume of rRNA-depleted RNA will exceed the capacity of the columns. Therefore, divide each RNA sample into 100-µl aliquots. To the first aliquot, add 350 µl of RLT buffer, followed by 250 µl of 100% RNase-free ethanol. Run this mixture through the RNeasy column to bind the RNA. Repeat this with the other aliquots until all of the RNA from a given sample has been bound to the same column. Proceed with column washes according to the manufacturer's protocol.

21| Elute the RNA in a final volume of 50 µl of RNase-free H_2O. Quantify the RNA by using a Qubit and determine the percentage yield after rRNA depletion (compared with the starting amount of RNA determined in Step 17). The rRNA-depleted RNA yield typically ranges from 1.0 to 6.0%.

(Optional) RNA fragmentation ● TIMING 1 h
▲ **CRITICAL** This section is important for Type I experiments (see Experimental design) when you are comparing transcript representation in pure TU-tagged RNA with the starting total RNA. Omitting Steps 22–24 for Type II experiments simplifies the protocol, and it may improve the yield and Illumina library synthesis.

22| Fragment the entire 50 µl of the rRNA-depleted RNA in a 100-µl final volume according to the NEBNext RNA fragmentation kit protocol. Use thin-walled PCR tubes and a thermocycler to incubate at 85 °C for 4 min with cooling to 10 °C before proceeding.

23| Purify the fragmented RNA by using an RNeasy mini kit according to the manufacturer's instructions. Elute RNA in 28 µl of RNase-free water.

24| (Optional) For Type I experiments, reserve 1 µl of fragmented RNA for cDNA synthesis of the total RNA (Step 30). Dilute this aliquot with RNase-free H_2O to a concentration no greater than 50 ng µl^{-1}.

25| Determine the RNA fragmentation efficacy by using an Agilent Bioanalyzer. Dilute 1 µl of the fragmented RNA with RNase-free dH_2O to a concentration of 2 ng µl^{-1}. Run the Bioanalyzer according to the manufacturer's instructions or submit the samples to a core facility. Fragmentation should produce fragmented RNA with an average size of ~200–500 nt.
? TROUBLESHOOTING

Biotinylation ● TIMING 4 h
26| To all the remaining RNA (from Step 21 if unfragmented, from Step 23 if fragmented), add 10 µl of 10× TE, H_2O to 75 µl, and 25 µl of biotin-HDPD for a final volume of 100 µl. Incubate the samples in the dark for 3 h at room temperature.

27| Purify biotinylated RNA by using the RNeasy mini kit according to the manufacturer's instructions, eluting it in 50 µl of RNase-free H_2O.

Isolation of biotinylated RNA ● TIMING 1 h
28| Isolate biotinylated RNA from nonlabeled RNA according to the µMacs streptavidin kit's directions, with the exception of using 100 mM 2-mercaptoethanol instead of DTT for the elution step. Elute it twice with 100 µl of preheated 100 mM 2-mercaptoethanol and pool the eluates.
! CAUTION 2-mercaptoethanol is toxic and should be used in a fume hood while you are wearing gloves and a lab coat. Avoid inhalation.

29| Purify the eluted biotinylated TU-tagged RNA by using RNeasy MinElute columns, eluting in 12 µl of RNase-free H_2O. Attempt to measure the RNA concentration via a Qubit and/or Bioanalyzer.
▲ CRITICAL STEP The yield of TU-tagged purified RNA is usually below detection limits. Essentially all the RNA, which is recovered in ~10 µl, is used for the subsequent cDNA synthesis step.
■ PAUSE POINT (end of day 2) RNA should be frozen and stored at −80 °C indefinitely. Optionally, proceed with library synthesis.

Illumina library preparation ● TIMING 5 h 30 min
30| Prepare cDNA from 9.5 µl of the purified RNA from Step 29 and, for Type I experiments, 50 ng of the diluted total RNA reserved at Step 24, by using the ScriptSeq v2 RNA-Seq library preparation kit. Follow the manufacturer's protocol (ScriptSeq v2 manual rev. A., 12/2012) beginning with appendix sections 4.1.A and 4.1.B and finishing with section 3.C.

31| Purify cDNA of primers by using Agencourt AMPure XP beads according to the directions in section 3.D of the ScriptSeq v2 manual. For best results, use the 6S Super Magnet plate for immobilizing the beads.

32| Amplify the cDNA samples with 12–15 PCR cycles by following the directions in section 3.E of the ScriptSeq v2 manual.
▲ CRITICAL STEP Although we typically require 15 cycles to generate sufficient DNA, limiting PCR cycles when possible will increase library diversity and improve downstream analyses. Use different index reverse PCR primers from the ScriptSeq index PCR primers set for each RNA sample.

33| Purify the RNA-seq libraries with AMPure XP beads according to the directions in section 3.F of the ScriptSeq v2 manual. Elute samples in 20 µl of RNase-free water.

34| Qubit-quantify each library. Start with 1 µl and incrementally increase up to 5 µl until a Qubit reading can be attained.
? TROUBLESHOOTING

35| Mix equimolar amounts of each library (typically four, but up to six, all synthesized with their distinct index primers) to a total concentration and volume requested by your sequencing facility, and then submit the pooled libraries for Illumina sequencing. If necessary, use a MinElute reaction cleanup column to concentrate the pooled RNA-seq libraries. We recommend adding Tween-20 to a final concentration of 1% (vol/vol) to prevent the DNA from adhering to tube walls. We typically perform single-read sequencing on a single lane of an Illumina HiSeq 2000. If paired-end reads are desired, the bioinformatics protocol described below should be modified accordingly.

Processing Illumina data ● TIMING 6–12 h
36| Download fastq files onto a Unix server. If multiple fastq files are produced for a given sample, use the Unix command 'cat' to concatenate them into a single file. You will produce one fastq file per library sequenced, which we name by sample run and bar code (e.g. 1234_ATCGAC.fastq). Assess the quality of your reads (e.g., with FastQC) and trim for quality and adaptor sequences if necessary. Note the number of total raw reads for each fastq file.

37| Align reads to the reference genome by using the TopHat2 splice aligner package. If possible, use parallel processing to expedite read aligning. The example TopHat command listed below uses 48 cores, which is a parameter that should be adjusted to suit your server's capabilities. The command will generate a folder called 1234_ATCGAC containing the alignment file, acceptedhits.bam. Note the number of mapped reads for each fastq file after running the SAMtools view command.

```
$ tophat -p 48 --library-type fr-secondstrand -o 1234_ATCGAC -G [path to and
name of reference gtf file] [path to bowtie2 indices] 1234_ATCGAC.fastq

$ samtools view -c -F 4 acceptedhits.bam
```

38| Use SAMtools to generate a sam file from the acceptedhits.bam file.

```
$ samtools view acceptedhits.bam > 1234_ATCGAC_acceptedhits.sam
```

39| Map the reads to exon features within the reference genome with the HTSeq package. Only reads mapping to the correct strand will be scored.

```
$ htseq-count -m union -s yes 1234_ATCGAC_acceptedhits.sam [path to and name
of reference gtf file] > 1234_ATCGAC.htseq
```

Determining enriched, depleted or differentially expressed genes ● TIMING 1–2 h
40| Download each counts table, designated by the file extension '.htseq', and open files in Microsoft Excel.

41| Create a combined counts table in Excel for all the samples to be compared. The instructions that follow are for a Type I experiment performing a comparison between two sets of total RNA and TU-tagged pure RNA. Start with one of the total sets. Copy the entire counts column from the corresponding pure.htseq file into the adjacent third column. Do likewise with the second total and pure data sets. This spreadsheet will now have five columns, the first being gene symbols and the adjacent four containing the corresponding counts data from each sample. Name the columns 'Symbol', 'T1', 'P1', 'T2' and 'P2'. Save as a tab-delimited text file (tutag_counts.txt). An example counts table is included as **Supplementary Table 1**.

42| Import the tab-delimited counts table into R studio. Note that everything following the '>'s below are the R commands. The head command will show the first ten rows of imported table, which will confirm that the formatting is correct.

```
> countsTable <- read.delim ( "tutag_counts.txt", header=TRUE, row.names=1)

> head( countsTable )
```

43| Describe the organization of your counts table (i.e., which columns correspond to which samples). For the Type I experiment example outlined here, the condition table indicates the column position of each pure sample with a 'P' and each total sample position with a 'T'.

```
> condition = factor(c( "T", "P", "T", "P"))
```

PROTOCOL

44| Install the DESeq package into R studio. Then, load the DESeq library.

```
> source("http://www.bioconductor.org/biocLite.R")
> biocLite("DESeq")
> library (DESeq)
```

45| For manipulation with DESeq package, generate a CountDataSet.

```
> cds <- newCountDataSet( countsTable, condition )
```

46| Filter out low-count genes that have no chance of showing statistical significance (removing the lowest 40% quantile by summed overall counts).

```
> rs = rowSums ( counts ( cds ))
> theta = 0.4
> use = (rs > quantile(rs, probs=theta))
> table(use)
> cdsFilt = cds[ use, ]
```

▲ **CRITICAL STEP** Failure to filter low-count genes may prevent DESeq from functioning.

47| Estimate the effective library size (normalization).

```
> cdsFilt <- estimateSizeFactors( cdsFilt )
```

48| Estimate the dispersions of the data set.

```
> cdsFilt <- estimateDispersions( cdsFilt )
```

49| Determine differentially expressed genes. The head command will display the DESeq analysis of the first ten rows of your counts table, including the Benjamini-Hochberg adjusted *P* value ('padj') for each gene ('id').

```
> res = nbinomTest( cdsFilt, "T", "P")
> head(res)
```

50| Generate a scatter plot of the data by plotting the log fold change against the mean normalized counts. A graph will be produced that shows each transcript plotted as factors of the mean normalized counts and the $\log_2$ fold change. Differentially expressed transcripts (with a false discovery rate (FDR) of 10%) are highlighted in magenta. An example plot is shown in **Figure 2** (green dots additionally indicate positive control genes).

```
> plotMA(res)
```

51| Filter the data for differentially expressed genes, according to a chosen FDR threshold. Here the adjusted *P* value is set at <0.05, according to convention.

```
> resSig = res[ res$padj < 0.05, ]
```

52| Save the output to an exportable file.

```
> write.table (resSig, file= "DESeq_results.txt", sep= "\t ",
row.names=TRUE, col.names=TRUE)
```

53| Open the DESeq_results.txt file in Excel for sorting and analysis. (Optional) Filter the data to only display genes that have at least the same number of total counts as that of the least-expressed positive control gene. This will highlight genes

Figure 2 | DESeq Output plot for visualizing differential expression. Shown is a scatter plot of the entire data set from a Type I experiment (endothelial and microglia/macrophages of P6 mouse brain) analyzed by DESeq, where the $\log_2$ fold change of each gene is plotted against the total number of counts recorded for that gene. Differentially expressed genes ($P < 0.10$) are highlighted in magenta. The 13 positive control endothelial transcripts are marked in green. Triangles represent data points outside the graph area.
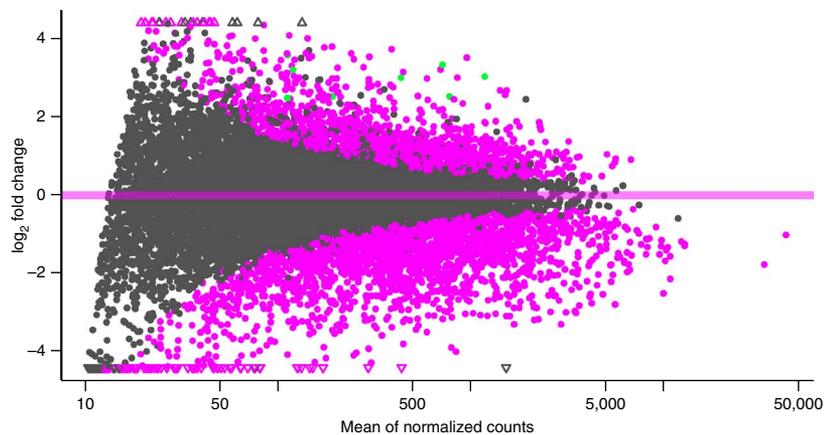
with sufficient expression likely to be biologically relevant. **Supplementary Table 2** is an example of the filtered output showing the 20 most-enriched transcripts in the Type I experiment described in ANTICIPATED RESULTS. **Supplementary Table 3** shows the results for all transcripts. A Type II experiment will produce the same table, but the fold change should be interpreted as such rather than the fold enrichment.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
6	Homogenate looks foamy and white	Tissue:TRIzol ratio is too high	Divide the tissue sample into two tubes and dounce it in at least twice the TRIzol
	Residual tissue fragments in the TRIzol after douncing	Insufficient mechanical action was used to disrupt the tissue	Use a syringe and needle to further break up the tissue; use a tissue homogenizer; spin down cell debris before adding chloroform
18	RNA is degraded or contaminated (RNA integrity number (RIN) < 8.0)	Degradation during sample preparation, excessive amount of tissue used for RNA extraction, contamination of the aqueous phase with the interface phase during RNA extraction with TRIzol	For extremely valuable samples, RNA can be further purified using RNeasy columns (up to 100 µg per column) and then re-analyzed. To avoid future low RIN values of RNA extractions from tissue, flash-freeze tissues in liquid nitrogen immediately after harvesting; increase the TRIzol:tissue ratio; leave ~20% of the aqueous phase behind after chloroform extraction during the TRIzol procedure
25	Fragmentation is either excessive or insufficient	Fragmentation conditions are not optimized for the sample RNA	Adjust the fragmentation reaction time and temperature following instructions in the NEBNext kit protocol to generate the desired fragmentation profile
34	Low or undetectable DNA concentration	Low or insufficient starting material	First, try quantifying the Illumina library DNA concentration by quantitative PCR. Otherwise, re-purify TU-tagged RNA starting with a larger amount of total RNA (up to 100 µg, pooling samples together if necessary)

● TIMING

Day 1 (7–10 h)
Steps 1–4, 4TU delivery: 5–7 h
Steps 5–18, RNA purification: 2–3 h
Day 2 (8 h)
Steps 19–21, removal of rRNA: 2 h

Steps 22–25, RNA fragmentation: 1 h (optional)
Steps 26 and 27, biotinylation: 4 h
Steps 28 and 29, isolation of biotinylated RNA: 1 h
Day 3 (5.5 h)
Steps 30–35, Illumina library preparation: 5.5 h
Day 4 (time varies)
Steps 36–39, processing Illumina data: 6–12 h
Steps 40–53, determining enriched, depleted or differentially expressed genes: 1–2 h

ANTICIPATED RESULTS

This protocol includes example data from a successful experiment describing the P6 brain *Tie2:Cre*-lineage transcriptome (endothelial cells and macrophages/microglia). P6 *Tie2:Cre;CA > GFPstop > HA-UPRT* double-transgenic mice were given a 400 mg kg^{-1} s.c. 4TU injection and killed after 6 h. The whole brain was removed for RNA preparation. We used 50 μg of starting RNA from each of two littermate pups. After rRNA removal and RNA fragmentation, we recovered 1.1 and 1.3 μg of rRNA-depleted total 1 and total 2 RNA samples, respectively. TU-tagged RNA was purified from both samples. The two total and two pure RNA preparations were used to synthesize bar-coded Illumina libraries that were pooled and sequenced on one lane of a HiSeq 2000. We obtained 39 million total reads for the total 1 sample, 19 million total reads for the pure 1 sample, 34 million total reads for the total 2 sample and 33 million total reads for the pure 2 sample. Of those reads, 27 million, 14 million, 24 million and 24 million reads aligned to the mouse genome, respectively. After HTSeq read mapping, DESeq analysis and expression-level filtering, we identified 913 differentially expressed genes that were more than twofold enriched in *Tie2:Cre*-lineage cells. The HTSeq-generated counts tables for this experiment are included as **Supplementary Table 1**. A DESeq/R-generated scatter plot highlighting differentially expressed genes and positive control genes is shown in **Figure 2**. **Supplementary Table 2** shows the top 20 most-enriched transcripts and **Supplementary Table 3** shows the statistical analysis of the entire transcriptome.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS L.G., K.V.K., M.R.M., C.Q.D. and K.S. developed the protocol. L.G., K.V.K., C.Q.D. and K.S. prepared and wrote the manuscript.

1. Gay, L. *et al.* Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. *Genes Dev.* **27**, 98–115 (2013).
2. Miller, M., Robinson, K., Cleary, M. & Doe, C. TU-tagging: cell type–specific RNA isolation from intact complex tissues. *Nat. Methods* **6**, 439–441 (2009).
3. Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R. & Boothroyd, J.C. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat. Biotechnol.* **23**, 232–237 (2005).
4. Friedel, C.C. & Dölken, L. Metabolic tagging and purification of nascent RNA: implications for transcriptomics. *Mol. Biosystems* **5**, 1271–1278 (2009).
5. Lai, S.L., Miller, M.R., Robinson, K.J. & Doe, C.Q. The Snail family member Worniu is continuously required in neuroblasts to prevent Elav-induced premature differentiation. *Dev. Cell* **23**, 849–857 (2012).
6. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
7. Doyle, J.P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
8. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
9. Deal, R.B. & Henikoff, S. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat. Protoc.* **6**, 56–68 (2011).
10. Sanz, E. *et al.* Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. USA* **106**, 13939–13944 (2009).
11. Kenzelmann, M. *et al.* Microarray analysis of newly synthesized RNA in cells and animals. *Proc. Natl. Acad. Sci. USA* **104**, 6164–6169 (2007).
12. Feil, R. *et al.* Ligand-activated site-specific recombination in mice. *Proc. Natl. Acad. Sci. USA* **93**, 10887–10890 (1996).
13. Diez-Roux, G. *et al.* A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582 (2011).
14. Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **10**, 623–629 (2013).
15. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
16. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079 (2009).
17. Anders, S. HTSeq: Analyzing high-throughput sequencing data with Python http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html (2011).
18. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).