UC Davis

UC Davis Previously Published Works

Title

Applying thiouracil tagging to mouse transcriptome analysis

Permalink

https://escholarship.org/uc/item/855809bx

Journal

Nature Protocols, 9(2)

ISSN

1754-2189

Authors

Gay, Leslie
Karfilis, Kate V
Miller, Michael R
et al.

Publication Date

2014-02-01

DOI

10.1038/nprot.2014.023

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

NIH Public Access
Author Manuscript
Nat Protoc. Author manuscript; available in PMC 2014 August 01.

Published in final edited form as:
Nat Protoc. 2014 February ; 9(2): 410–420. doi:10.1038/nprot.2014.023.

Applying thiouracil (TU)-tagging for mouse transcriptome analysis

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

[1]Institute of Molecular Biology, University of Oregon, Eugene, OR 97403

[2]Institute of Neuroscience, University of Oregon, Eugene, OR 97403

[3]Howard Hughes Medical Institute, University of Oregon, Eugene, OR 97403

Abstract

Transcriptional profiling is a powerful approach to study 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 1) the isolation of RNA from a given cell population of a complex tissue, avoiding transcriptional changes induced by cell isolation trauma, and 2) the identification of actively transcribed RNAs and not pre-existing transcripts. Therefore, in contrast to other cell-specific transcriptional profiling methods based on purification of tagged ribosomes or nuclei, TU-tagging provides a direct examination of transcriptional regulation. We describe how to: 1) deliver 4-thiouracil to transgenic mice to thio-label cell lineage-specific transcripts, 2) purify TU-tagged RNA and prepare libraries for Illumina sequencing, and 3) follow a straight-forward bioinformatics workflow to identify cell type-enriched or differentially expressed genes. Tissue containing TU-tagged RNA can be obtained in one day, RNA-Seq libraries generated within two days, and, following sequencing, an initial bioinformatics analysis completed in one additional day.

Keywords

TU-tagging; 4-thiouracil; uracil phosphoribosyltransferase; transcriptomics; transcriptional profiling; differential gene expression; mouse genetics; nascent RNA; intersectional genetics; chemical genetics

[5]Corresponding authors: Chris Doe: cdoe@uoneuro.uoregon.edu, Kryn Stankunas: kryn@uoregon.edu.
[*]these authors contributed equally
[4]Current address: Department of Animal Science, University of California, Davis

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.

Competing Financial Interestes The authors declare that they have no competing financial interests.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Introduction

Development of the protocol

The emergence of microarrays and next generation sequencing have 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 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. Resulting “TU-tagged” RNA is then transcriptionally profiled by RNA-Seq using 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 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 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 Crosslinking 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 when only temporal labeling is desired or for generating UPRT-chimeric mice by cell transplantations (e.g. bone marrow transplants)[1].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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, takes time, and can result in 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, are prone to operator-variability, and rarely allow high purity separations. 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), providing extra information on gene regulatory networks. TRAP and Ribo-tag examine ribosomal protein-bound mRNA, and therefore are especially useful as surrogate approaches to study global protein translation. These methods may not detect non-coding RNAs (ncRNA) including miRNAs. In contrast, TU-tagging labels all newly transcribed RNA and therefore 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 (Figure 1). The first 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 the research team have at least some experience in each of mouse genetics, molecular biology, and statistical analysis of large datasets.

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-*

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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 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, 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 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 number 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, 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 to 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 enriched” 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 more difficult to achieve. Note that short-lived transcripts will be overrepresented in the “pure” compared to 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 with 4-thiouridine (4TUd) in parallel with littermates exposed to 4TU. Whereas 4TU requires UPRT to be incorporated into nascent RNA, 4TUd will be added to newly synthesized RNA in all

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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 harvested 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 vs. 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 vs. total, Type II = pure vs. 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 brain endothelial and macrophage/microglia transcriptomes[1].

Regardless of design, it is preferable to pre-define 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 using a gene expression database such as Eurexpress[13] (http://www.eurexpress.org/) to confirm the preponderance of enriched transcripts are 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; Figure 1). Typically, the *tissue-specific:Cre* and *CA>GFPstop>HA-UPRT* double transgenic adults or postnatal pups are injected with 4TU and harvested 4-6 hours later. We have observed a drop in labeling if tissue is harvest >= 12 hours after injection, enabling pulse-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

chase experiments to study transcriptome dynamics. For embryonic studies, pregnant mice carrying double transgenic embryos are exposed to 4TU for at least six hours. Embryonic studies use a longer 4TU exposure period because of an anticipated time lag for the compound to cross the placental barrier into embryonic/fetal circulation. The desired tissue is harvested and total RNA is isolated in approximately three hours, 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 (Figure 1) requires at least 10 μg of RNA from each sample. Less RNA may be used, but with inconsistent outcomes. First, we deplete the RNA from ribosomal RNA (rRNA) using Ribo-Zero kits, as described in the Procedure. Alternative approaches include an RNase H method[14] that is likely 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 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 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 to spuriously but infrequently labeled transcripts from non-UPRT expressing cells. Fragmenting the RNA prior to 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 using streptavidin-conjugated magnetic beads. Fragmentation, biotinylation, and purification take together about six hours. 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 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 consult with bioinformatics experts as necessary. Here, we describe one straightforward workflow that uses popular software packages for characterizing 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 multi-core processing. Using the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

alignment and mapping tools TopHat2[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 re-saved as a tab-delimited file for import into R. We suggest using the freely available RStudio package (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 for less-experienced researchers to become familiar with the programs and their parameters.

Limitations

Applications of mouse TU-tagging 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 an insufficient number 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 TU-tagged RNA can be recovered if even 1% of the cells express UPRT[1]. While 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 or not labeled during short 4TU exposure periods and therefore 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 national animal care guidelines

DMSO (Sigma, cat. no. D1435)

Corn oil (Sigma, cat. no. C8267.)

4-thiouracil (4TU) (Sigma, cat. no. 440736)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TRIzol (Invitrogen, cat. no. 15596-026)

Chloroform (Macron Fine Chemicals, cat. no. 444004)

!CAUTION TRIzol and chloroform are toxic and should be used in a fume hood while wearing gloves and a lab coat.

Isopropyl alcohol (Macron Fine Chemicals, cat. no. MK303202)

Ethanol (EtOH) (Pharmco Aaper, cat. no. 11100020G)

RNase-free water (H_2O) (Ambion cat. no. AM9938)

1M Tris pH 8.0 RNase-free (Ambion cat. no. AM9855G)

0.5M EDTA 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. Glove and lab coat should be worn when handling. 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)

FailSafe PCR Enzyme Mix (Epicentre, cat. no. FSE51100)

ScriptSeq Index PCR Primers (Epicentre, cat. no. RSBC10948 (set 1))

Equipment

1.5 mL low adhesion RNase-free microcentrifuge tubes (USA Scientific, cat. no. 1415-2600)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

15 mL centrifuge tubes (Corning, cat. no. 430791)

0.2 mL thin-walled PCR tubes (USA Scientific, cat. no. 1402-4300)

1 mL Tuberculin syringe with 26 G × 3/8 in. detachable needle (VWR, cat. no. BD309625)

1 liter Nalgene Dewar (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)

6-tube Magnetic Separation Rack (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 4TU in 1 mL DMSO, which can be stored at least 2 months at -20°C. On the day of injection, dilute the stock 1:4 in corn oil to a concentration of 50 mg/mL and vortex vigorously. Note, phase separation will occur if the solution is left standing for an extended period. Therefore, re-vortex the 4TU immediately prior to injection.

Biotin—Dissolve 10 mg EZ-Link Biotin-HPDP in 10 mL N,N-Dimethylformamide. Store 250 µL aliquots in a covered container at -20°C for up to six months.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

10% (vol/vol) Tween-20—Pipette 9 mL RNase-free water into a 15 mL tube. Add 1 mL of Tween-20 to a total volume of 10 mL. Mix well. Store in the dark.

100 mM 2-mercaptoethanol—Prepare fresh, immediately before use. Mix 3.5 μL of 2-mercaptoethanol with 500 μL of RNase-free dH_20. Preheat to 80°C.

TE—Make a 10× TE stock solution in a 15 mL tube by mixing 1 mL RNase-free 1M Tris-HCL pH 8.0, 200 μL RNase-free 0.5M EDTA pH 8.0, and 8.8 mL of RNase-free dH_20. Dilute 10× TE 1:10 with RNase-free dH_20 to make a 1× TE working solution.

Procedure

4TU Delivery

TIMING 5 hours for adults or P6 pups, and 7 hours for embryos

1| Weigh each mouse and calculate volume of the prepared 4TU solution required to deliver 400 mg/kg (e.g. 200 μL of a 50 mg/mL 4TU solution for a 25 g mouse). Lower 4TU doses generate less labeling; higher doses are not practical and potentially toxic.

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

3| Allow at least 4 hours of 4TU exposure prior to tissue harvesting for adults and pups and 6 hours 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 hours following 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 hours after 4TU injection.

4| Dissect out the organ or tissue of interest, transfer immediately into a 1.5 mL microcentrifuge tube, and flash freeze in liquid nitrogen. Continue with RNA preparation or store at -80°C.

PAUSE POINT Frozen tissue can be stored at -80°C indefinitely.

RNA Purification

TIMING 2-3 hours

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 used, and gloves worn at all times.

5| Precool the centrifuge to 4°C.

6| Homogenize tissue in a 1.5 mL tube using a Kontes pestle and 500 μl TRIzol per 100 mg tissue, until completely solubilized.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

!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. Incubate 5 minutes at room temperature (approximately 22°C).

8| Add 200 µl chloroform and vortex 15 seconds. Incubate for 2-3 minutes at room temperature.

!CAUTION Wear gloves and use a fume hood when working with chloroform.

CRITICAL STEP The extended chloroform vortex is important for high RNA yield and quality.

9| Centrifuge at 12,000 *g* for 15 minutes at 4°C.

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

11| Centrifuge at 20,000 *g* for 10 minutes at 4°C.

12| Decant liquid and wash with 1 mL 75% (vol/vol) EtOH.

13| Centrifuge at 20,000 *g* for 5 minutes at 4°C. Decant liquid.

14| Centrifuge at 20,000 *g* for 2 minutes to remove excess EtOH. Carefully pipette remaining liquid away from pellet and place tube upside down on a laboratory tissue for 2-3 minutes to dry.

CRITICAL STEP Do not over dry the RNA because this will make it difficult to re-solubilize.

15| Resuspend the pellet in 50-100 µL of RNase-free H_2O. Dilute the RNA to a concentration of approximately 200 ng/µL. The RNA concentration should be determined by spectophotometry (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 following the manufacturer's protocol.

17| Dilute 2 µL of the total RNA in 98 µL of RNase-free H_2O. Quantify the RNA concentration using a Qubit RNA BR Assay kit and a Qubit system following the manufacturer's directions.

18| Determine the integrity of the RNA samples using an Agilent Bioanalyzer. The RNA integrity number (RIN) should be > 8.0.

CRITICAL STEP RNA should be aliquoted to avoid repeated freezing and thawing.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

PAUSE POINT (End Day One) RNA should be frozen and stored at -80°C overnight or until ready to proceed.

? TROUBLESHOOTING

Removal of Ribosomal RNA

TIMING 2 hours

CRITICAL At this point, there are several alternatives for the removal of rRNA. Our standard approach uses 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 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, following the manufacturer's instructions.

CRITICAL STEP Using the maximum 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 using an RNeasy Mini column following 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 RLT buffer followed by 250 μL 100% RNase Free EtOH. Run this mixture through the RNeasy column to bind the RNA. Repeat this with the other aliquots until all the RNA from a given sample has been bound no the same column. Proceed with column washes following the manufacturer's protocol.

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

(Optional) RNA Fragmentation

TIMING 1 hour

CRITICAL This step is important for Type I experiments (see **Experimental design**), when comparing transcript representation in pure TU-tagged RNA, to the starting total RNA. Omitting steps 22-24 for Type II experiments simplifies the protocol and may improve yield and Illumina library synthesis.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

22| Fragment all 50 µL of the rRNA-depleted RNA in a 100 µL final volume following the NEBNext RNA fragmentation kit protocol. Use thin-walled PCR tubes and a thermocycler to incubate at 95°C for 4 minutes with cooling to 10°C before proceeding.

23| Purify fragmented RNA using an RNeasy Mini kit following the manufacturer's instructions. Elute RNA in 28 µL 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.

25| Determine RNA fragmentation efficacy using an Agilent Bioanalyzer. Dilute 1 µL of the fragmented RNA with RNase-free dH_20 to a concentration of 2 ng/µL. Run the Bioanalyzer following the manufacturer's instructions or submit the samples to a core facility. Fragmentation should produce fragmented RNA with an average size of approximately 200-500 nucleotides.

? TROUBLESHOOTING

Biotinylation

TIMING 4 hours

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

27| Purify biotinylated RNA using the RNeasy Mini kit according the manufacturer's instructions, eluting in 50 µL RNase-free H_2O.

Isolation of biotinylated RNA

TIMING 1 hour

28| Isolate biotinylated RNA from non-labeled RNA according to the µMacs Streptavidin kit's directions with the exception of using 100 mM 2-mercaptoethanol instead of dithiothreitol for the elution step. Elute twice with 100 µL of pre-heated 100 mM 2-mercaptoethanol and pool the eluates.

!CAUTION 2-mercaptoethanol is toxic and should be used in a fume hood while wearing gloves and a lab coat. Avoid inhalation.

29| Purify the eluted biotinylated TU-tagged RNA using RNeasy MinElute columns, eluting in 12 µL RNase-free H_2O. Attempt to measure the RNA concentration using 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 approximately 10 µl, is used for the subsequent cDNA synthesis step.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

PAUSE POINT (End Day Two) RNA should be frozen and stored at -80°C indefinitely. Optionally, proceed with library synthesis.

Illumina Library Preparation

TIMING 5 hours 30 minutes

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, using the ScriptSeq v2 RNA-Seq Library Preparation Kit. Follow the manufacturer's protocol (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 using Agencourt AMPure XP beads following 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 using 12-15 PCR cycles following directions 3.E of the ScriptSeq v2 manual.

CRITICAL STEP While we typically require 15 cycles to generate sufficient DNA, limiting PCR cycles where 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 following directions 3.F of the ScriptSeq v2 manual. Elute samples in 20 μL 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 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 hours

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 barcode (e.g. 1234_ATCGAC.fastq). Assess the quality of your reads (e.g. with FastQC) and trim for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

quality and adapter sequences if necessary. Note the number of total raw reads for each fastq file.

37| Align reads to the reference genome using the TopHat2 splice aligner package. If possible, use parallel processing to expedite read aligning. The example Tophat command listed below uses 48 cores, 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 using 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
```

Determine enriched/depleted or differentially expressed genes

TIMING 1-2 hours

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" datasets. 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 the formatting is correct.

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

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

```
> 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”))
```

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 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 dataset.

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

49| Determine differentially expressed genes. The head command will display the DESeq analysis of the first ten rows of your countsTable, including the Benjamini-Hochberg adjusted p-value (“padj”) for each gene (“id”).

```
> res = nbinomTest(cdsFilt, “T”, “P”)

> head(res)
```

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

50| Generate a scatter plot of the data by plotting the log fold change against the mean normalized counts. A graph will be produced showing each transcript plotted as factors of the mean normalized counts and the log2foldchange. Differentially expressed transcripts (with false discovery rate (FDR) of 10%) are highlighted in red. An example plot is shown in Figure 2 (blue 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, following 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 with sufficient expression likely to be biologically relevant. Supplementary Table 2 is an example of the filtered output showing the twenty 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 “fold enrichment”.

TIMING

Day 1 (7-10 hours)

Step 1-4, 4TU delivery: 5-7 hours

Step 5-18, RNA purification: 2-3 hours

Day 2 (8 hours)

Step 19-21, Removal of Ribosomal RNA: 2 hours

Step 22-24, RNA fragmentation: 1 hour

Step 25-27, Biotinylation: 4 hours

Step 28-29, Isolation of biotinylated RNA: 1 hour

Day 3 (5.5 hours)

Step 30-35, Illumina library preparation: 5.5 hours

Day 4 (time varies)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Step 36-39, Processing Illumina data: 6-12 hours

Step 40-53, Determine enriched/depleted or differentially expressed genes: 1-2 hours

TROUBLESHOOTING

Step	Problem	Possible reason	Solution
6	Homogenate looks foamy and white	Tissue to TRIzol ratio is too high	Divide the tissue sample into two tubes and dounce it in at least twice the TRIzol
6	Residual tissue fragments in the TRIzol after douncing	Insufficient mechanical action was used to disrupt tissue	Use a syringe and needle to further break up the tissue; use a tissue homogenizer; spin down cell debris prior to adding chloroform
18	RNA is degraded or contaminated (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/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 TRIzol:tissue ratio; leave approximately 20% of aqueous phase behind following 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 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)

Anticipated Results

This protocol includes example data from a successful experiment describing the postnatal day 6 (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 subcutaneous 4TU injection and sacrificed after six hours. The whole brain was removed for RNA preparation. We used 50 μg of starting RNA from each of two littermate pups. Following ribosomal RNA removal and RNA fragmentation, we recovered 1.1 μg 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 barcoded 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. Following HTSeq read mapping, DESeq analysis, and expression level filtering, we identified 913 differentially expressed genes more than two fold 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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the University of Oregon Genomics Core Facility for Illumina sequencing; V. Devasthali, U. Hostick, and the UO Transgenic Mouse Facility for support generating the *CA>GFPstop>HA-UPRT* mice; P. Batzel for critical input on the bioinformatics workflow; and B. Simek for beta-testing the bioinformatics protocol. Use of the UO's ACISS server was supported by a Major Research Instrumentation grant from the National Science Foundation (OCI-0960354). Funding was provided by NIH 5R00HL087598 (K.S.), NIH 3R01DE013085 (V. Kaartinen, K.S.) and the Howard Hughes Medical Institute (C.Q.D.).

References

1. Gay L, et al. Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes & development. 2013; 27:98–115. [PubMed: 23307870]
2. Miller M, Robinson K, Cleary M, Doe C. TU-tagging: cell type–specific RNA isolation from intact complex tissues. Nature methods. 2009; 6
3. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nature biotechnology. 2005; 23:232–7.
4. Friedel CC, Dölken L. Metabolic tagging and purification of nascent RNA: implications for transcriptomics. Molecular bioSystems. 2009; 5:1271–8. [PubMed: 19823741]
5. Lai SL, Miller MR, Robinson KJ, Doe CQ. The Snail family member Worniu is continuously required in neuroblasts to prevent Elav-induced premature differentiation. Developmental cell. 2012; 23:849–57. [PubMed: 23079601]
6. Hafner M, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell. 2010; 141:129–141. [PubMed: 20371350]
7. Doyle JP, et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell. 2008; 135:749–62. [PubMed: 19013282]
8. Heiman M, et al. A translational profiling approach for the molecular characterization of CNS cell types. Cell. 2008; 135:738–48. [PubMed: 19013281]
9. Deal RB, Henikoff S. The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nature protocols. 2011; 6:56–68.
10. Sanz E, et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:13939–44. [PubMed: 19666516]
11. Kenzelmann M, et al. Microarray analysis of newly synthesized RNA in cells and animals. Proc natl acad sci U S A. 2007; 104:6164–9. [PubMed: 17405863]
12. Feil R, et al. Ligand-activated site-specific recombination in mice. Proc natl acad sci U S A. 1996; 93:10887–90. [PubMed: 8855277]
13. Diez-Roux G, et al. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS biology. 2011; 9:e1000582. [PubMed: 21267068]
14. Adiconis X, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nature methods. 2013; 10:623–9. [PubMed: 23685885]
15. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013; 14:R36. [PubMed: 23618408]
16. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009; 25:2078–9.
17. Anders, S. HTSeq: Analyzing high-throughput sequencing data with Python. 2011. http://www.huber.embl.de/users/anders/HTSeq/
18. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biology. 2010; 11:R106. [PubMed: 20979621]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

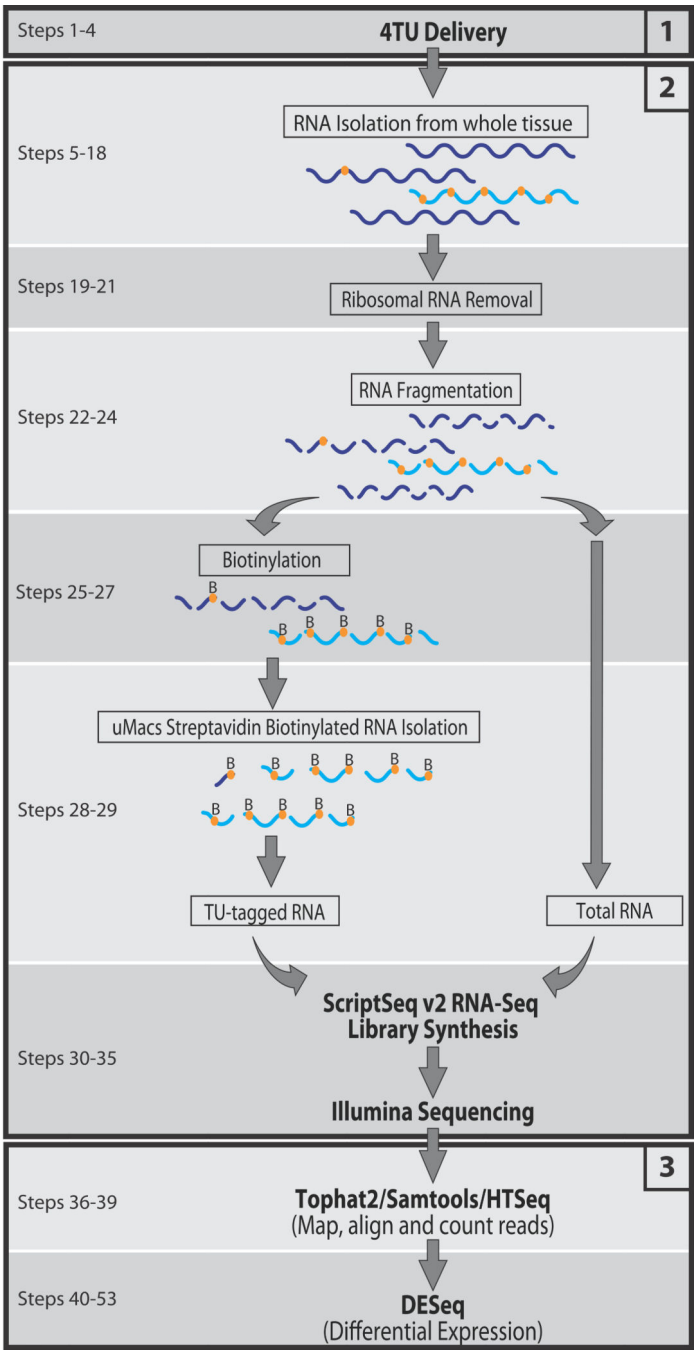

Figure 1.

Flowchart of the mouse TU Tagging protocol. This chart outlines the entire procedure from the delivery of 4-thiouracil to the analysis of Illumina sequencing data using DESeq (Type I experiment, see text). The black bordered boxes indicate the three major stages of the protocol. The gray shaded boxes highlight distinct sub-stages with step numbers corresponding with those in the Procedure. Blue wavy lines represent RNA produced from two mixed cell populations, one which is UPRT expressing (lighter color). The orange

circles indicate the incorporation of a thiolated uracil, which are then biotinylated (B) in vitro.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

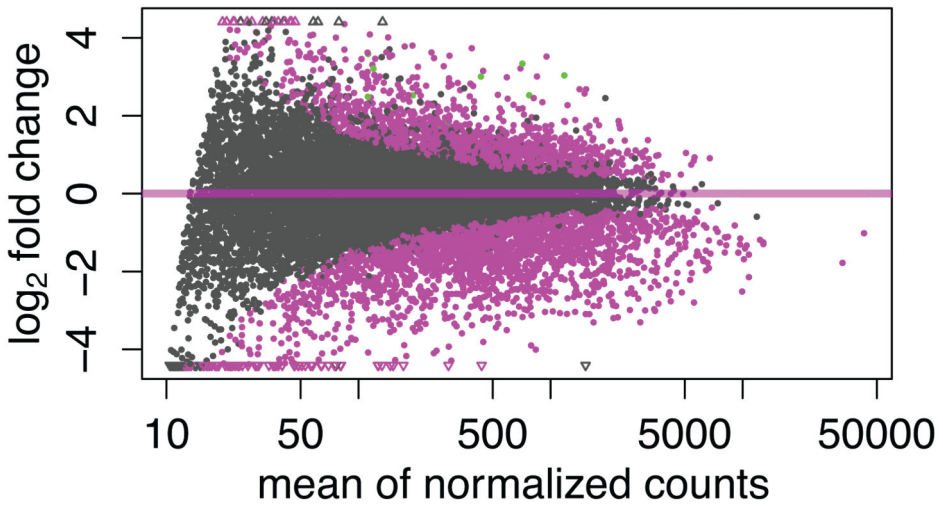

Figure 2.

DESeq Output Plot for Visualizing Differential Expression. Shown is a scatter plot of the entire dataset from a Type I experiment (endothelial and microglia/macrophage cells of postnatal P6 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 thirteen positive control endothelial transcripts are marked in green.