Sequencing cell type-specific transcriptomes with SLAM-ITseq

Wayo Matsushima[1,2,3], **Veronika A Herzog**[4], **Tobias Neumann**[5], **Katharina Gapp**[1,2,3], **Johannes Zuber**[5], **Stefan L Ameres**[4], **Eric A Miska**[1,2,3]

Wayo Matsushima: wm286@cam.ac.uk; Veronika A Herzog: veronika.herzog@imba.oeaw.ac.at; Tobias Neumann: tobias.neumann@imp.ac.at; Katharina Gapp: kg7@sanger.ac.uk; Johannes Zuber: johannes.zuber@imp.ac.at; Stefan L Ameres: stefan.ameres@imba.oeaw.ac.at

[1]Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK

[2]Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK

[3]Wellcome Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK

[4]Institute of Molecular Biotechnology, Vienna Biocenter Campus, Vienna, Austria

[5]Research Institute of Molecular Pathology, Vienna Biocenter Campus, Vienna, Austria

Abstract

Analysis of cell type-specific transcriptomes is vital for understanding the biology of tissues and organs in the context of multicellular organisms. Here, we describe a protocol to metabolically label RNA with a uracil analogue, 4-thiouracil, in specific cell types *in vivo*, and a pipeline to detect the labelled transcripts by a novel RNA-seq method, SLAMseq. By injecting 4-thiouracil to transgenic mice that express cell-specific UPRT, an enzyme required for 4-thiouracil incorporation, only cells expressing UPRT synthesise thiol-containing RNA. Total RNA isolated from a tissue of interest is then sequenced with SLAMseq, which introduces T>C conversions at the sites of the incorporated 4-thiouracil. The resulting sequencing reads are then mapped with the T>C aware alignment software, SLAM-DUNK (http://t-neumann.github.io/slamdunk/). The number of T>C per transcript determined is further analysed to identify which transcripts are synthesised in the UPRT-expressing cells. Thus, our method, SLAM-ITseq, enables cell-specific transcriptomics without laborious cell or RNA sorting steps required in the previously developed

methods such as FACS/RNA-seq or TU-tagging. In the murine tissues we assessed previously, this method identified around 5,000 cell-specific genes. Any laboratories with access to a high-throughput sequencer and high power computing can adapt this protocol with ease and the entire pipeline can be completed in less than five days.

Keywords

RNA-seq; 4-thiouracil; transcriptomics; transgenics; SLAMseq; SLAM-ITseq

Introduction

Life is supported by organ systems and each organ is formed of numerous highly specialised cells. Thus, it is essential to understand how each cell dynamically alters its gene expression in response to surrounding environmental changes. A simple yet effective way to study cell-specific transcriptome is to isolate a specific cell type of interest from an intact tissue, extract RNA from the isolated cells, and perform high-throughput gene expression analysis such as RNA-seq. Fluorescent-activated cell sorting (FACS) is often used to sort the cells tagged with one or more fluorescent markers from a dissociated tissue. Although this method has been commonly used in many research fields, it has a number of limitations: (i) An extensive optimisation is needed to effectively dissociate a tissue while maintaining cell viability and transcriptome [1]; (ii) A cell sorter is not always available and its run time is often extensive; (iii) A cell type of interest often consists of a minor proportion of a tissue; hence, use of a larger number of animals is sometimes necessary to obtain sufficient cells for a transcriptomic analysis.

Thus, a method that does not rely on physical cell isolation procedures had long been awaited for an easier and more accurate quantification of the transcriptome in a specific cell type.

Development of the protocol

An elegant alternative approach to study cell-specific transcriptome was proposed by Gay *et. al.*, who used a uracil analogue, 4-thiouracil, to label RNA in a specific cell type [2–4]. While many animals lack a pathway to convert 4-thiouracil to 4-thiouridine monophosphate (4-thio-UMP), which can be used as a substrate of RNA polymerases to synthesise RNAs, uracil phosphoribosyltransferase (UPRT) from *Toxoplasma gondii* (*T. gondii*) can convert 4-thiouracil to 4-thio-UTP [5]. Gay *et. al.* generated '*UPRT* mice', which have a transgene that consists of a constitutively active chicken beta-actin promoter, followed by floxed GFP with stop codons, and a UPRT-coding region downstream of these elements [2] (Fig. 1a). When this mouse is crossed with a mouse expressing Cre under a cell type-specific promoter, double-transgenic mice that express UPRT only in Cre-expressing cells can be obtained. Thus, by exposing these double-transgenic mice to 4-thiouracil systemically, only the cells expressing UPRT label RNA, while the rest of the cells do not. Subsequently, RNA from a tissue of interest is extracted and thiol-specific biotinylation followed by streptavidin pull-down is performed to isolate the thiol-containing RNA. Both this pulled-down fraction and input are

quantified by RNA-seq, and the labelled RNA can be identified by finding transcripts that are more enriched in the pulled-down fraction compared to the input fraction.

Although this method, i.e. TU-tagging, has already been used in a few studies [6–8], it faces significant limitations that prevent its broader applications. Among all, the use of the biotin-streptavidin based RNA isolation to identify thiol-containing RNA is the most challenging step. This method requires to sort thiol-containing RNA from total RNA followed by separate sequencing of these two fractions to bioinformatically identify genes that are enriched in the sorted fraction compared to input. However, it is not appropriate to simply apply an existing RNA-seq analysis pipeline to analyse these data, since most differential gene expression analysis softwares employ a normalisation method that assumes that the provided samples have similar proportions of RNA species apart from the genes that are differentially expressed, which is not the case when comparing pulled-down and input fractions. While the use of well-designed spike-ins for normalisation presents a potential solution [9], determining an appropriate sample-to-spike-in ratio is not always easy: too much spike-ins may result in the low depth of sample RNAs, whereas too low spike-in read count would not be sufficient for an accurate normalisation. Even if spike-in controls are introduced, it is still difficult to apply TU-tagging to cells that consist of either very minor or major proportion of a tissue [2]. If the cell number is quite limited, the RNA yield from pull-down eluates would not suffice for a high-throughput transcriptome analysis, while if a tissue has substantial proportion of labelled RNA, it is difficult to identify genes that are enriched in the eluate over input.

Here, we present an enrichment-independent approach to identify cell type-specific gene expression programs *in vivo*. Instead of sorting labelled RNA for a separate sequencing library, we employ thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAMseq), an RNA sequencing method that identifies the thiol-containing base by introducing thymine to cytosine (T>C) mismatch at the position of thiol-containing uridine. In this sequencing method, a carboxyamidomethyl group is added to the thiol group through a nucleophilic substitution reaction with iodoacetamide (IAA), and, in the following reverse transcription step, a guanine (G), instead of a adenine (A), is base-paired to this alkylated U resulting in T>C in the RNA-seq reads (Fig.1b) [10,11]. With this improvement, we now circumvent the tedious and noise-prone biochemical isolation step, and overcome the limitation in the range of the proportion of UPRT-expressing cells allowed in a given tissue.

Also, since mammalian cells could potentially have a trace of activity to incorporate exogenous uracil [12–14], background 4-thiouracil incorporation in non-UPRT-expressing cells has been a potential problem in previously developed methods. Especially, relying on enrichment ratio of eluate to flow-through, TU-tagging lacks a subjective threshold to tell the signal from noise. Therefore, we have redesigned the experiment to include an improved control to account for UPRT-independent RNA labelling and developed a bioinformatic pipeline that quantitatively identifies the labelled RNA based on T>C count (Fig.1b).

With this method, we term SLAMseq *in tissue* (SLAM-ITseq), we have successfully identified three different cell-specific transcriptomes in mouse, with little optimisation and

minimal hands-on time [15]. SLAM-ITseq can be applied in diverse cell types in different tissues with ease.

Applications of the method

In this protocol, we present a method for identifying cell-specific polyadenylated (polyA) RNA using a 3' mRNA sequencing kit from Lexogen. However, SLAM-ITseq can be applied using different RNA-seq library preparation kits as long as a cDNA synthesis process is involved. In fact, we have also used illumina TruSeq total RNA kit, which sequences all the non-ribosomal RNA species that are longer than 200 nt, and NEXTflex small RNA sequencing kit (Bioo Scientific) for RNA shorter than 200 nt. Both methods have successfully identified cell-specific long RNA and small RNA, respectively. Thus, SLAM-ITseq can be applied to study not only polyA RNA but also non-polyA RNAs including nascent transcripts and small RNA species such as microRNA (miRNA).

Another benefit of using metabolic labelling in transcriptomic analysis is that it provides temporal information about the sequenced RNA by labelling newly synthesised RNAs in a given time. This approach significantly differs from classic RNA-seq methods that can only measure steady-state RNA abundance. An *in vitro* experiment with SLAMseq successfully discovered direct targets of BRD4-MYC pathway [16]. Thus, when combined with a transient perturbation to animals, one can study dynamic transcriptional change of a specific cell type in response to stimuli in a unique time-controlled manner.

Also, this method could potentially be used to discover RNA species that work in a non-cell-autonomous manner as has been hinted by the small RNA mobility shown in *Caenorhabditis elegans* and plants [17]. A recent publication using 4-thiouracil showed data that is suggestive of potential intercellular transport of miRNA, though it was not possible to identify individual mobile miRNA species possibly due to its low concentration in the recipient cells [18].

Comparison with other methods

A clear advantage of SLAM-ITseq over classic FACS-based approaches is that it does not require any cell isolation steps, and could thus be used to study cells that are difficult to isolate from a tissue or that are vulnerable to any chemical and/or physical stimuli. It should, however, be noted that SLAM-ITseq relies on the specificity of a single promoter that drives Cre expression to achieve cell-specific RNA labelling. Hence, where identity of cells depends on expression of multiple genes, FACS might still be suitable (e.g. some subtypes of immune cells can only be distinguished based on the expression of multiple surface markers).

Another FACS-independent, yet pull-down dependent approach to study cell-specific RNA is to overexpress a tagged RNA-binding protein in a specific cell type to co-purify the RNA that is bound to it. By using different RNA-binding proteins to "fish out" RNAs, one can study different cell-specific properties: mRNAs being translated can be found with a tagged ribosome pull-down [19]; alternative polyadenylation events can be studied with polyA-binding protein (PABP) [20]. These methods and SLAM-ITseq are complementary to each other and enable investigation of RNA dynamics from different perspectives.

There are also alternative RNA-seq-based methods to detect thiol-containing RNA: thiouridine-to-cytidine sequencing (TUC-seq) [21] and TimeLapse-seq [22]. Instead of introducing T>C during an RT step as in SLAMseq, these two methods directly convert 4-thiouridine to cytidine in RNA by using osmium tetroxide (OsO_4) and ammonium in TUC-seq, and 2,2,2-trifluoroethylamine (TFEA) and sodium periodate ($NaIO_4$) in TimeLapse-seq. SLAMseq and TUC-seq have a comparative conversion rate (>90%) per each 4-thiouridine introduced. TimeLapse-seq only provides semi-quantitative data about its conversion rate with a restriction digestion assay, showing it has around 80% conversion rate per 4-thiouridine (also reviewed in [23]). Both methods could potentially be used instead of SLAMseq to determine labelled transcripts depending on one's experimental aim (e.g. direct RNA sequencing is desired). A direct comparison using the same starting material might be needed to conclude which method is the most sensitive and how significant the difference is.

Limitations

Sensitivity—Although we have confirmed sufficient labelling of RNA in endothelial cells, which represents a minor proportion of brain cells [15], SLAM-ITseq might not be sensitive enough to detect transcripts from an even less abundant cell population. The sensitivity could be improved by increasing sequencing depth, but it comes with an increased sequencing cost per sample.

A potential workaround is to enrich for the labelled RNA by the biotin-streptavidin isolation method, followed by SLAMseq on the purified RNA to identify which pulled-down RNA contains the thiol group. Although this particular combination has not been tested yet, TimeLapse-seq has used this strategy (TT-TimeLapse-seq) and detected more reads originated from transient transcripts (i.e. pre-mRNA) from cells exposed to 4-thiouridine for a limited time [22]. The benefits of combining these two methods are: 1) the labelled RNA is more deeply sequenced with unlabelled RNA removed, which might increase the sensitivity; 2) the unlabelled RNA that is non-specifically pulled down in the purification step can be identified based on the number of T>C in the reads. This combined method, however, has not been assessed for its recovery rate (i.e. how much labelled RNA is lost) and still implies all the drawbacks of using a purification step discussed in the "Development of the protocol" section, such as the requirement of spike-ins for downstream analysis; therefore, this prompts a thorough comparison to deem this approach is superior to increasing sequencing depth.

Also, accessibility for circulating 4-thiouracil in the blood might be limited for certain cell types (e.g. neurons), which may have lower labelling level. Increasing the dose could potentially have negative impacts on animal wellbeing so too much deviation from the reported dose is not recommended.

Effects on RNA metabolism—Since our knowledge about the effects of RNA modifications on the biochemical properties of RNA is still limited, it is hard to predict if 4-thiouracil-containing RNA retains the same properties as unmodified RNA. In an *in vitro* experiment, it was reported that the high dose of 4-thiouridine inhibits translation by inhibiting rRNA synthesis [24]. Although the concentration of 4-thiouracil we can achieve *in*

vivo with SLAM-ITseq is much lower than this *in vitro* experiment, its potential effects on RNA metabolism are unknown. It is recommended to routinely compare transcriptomes between Cre+ and Cre-. If the expression levels of the two highly correlate, it suggests that the 4-thiouracil incorporation at the given concentration has little impact on RNA metabolism.

Leaky Cre expression—As has been reported in several publications, the expression of promoters is not always as highly specific and, in some cases, stochastic [25]. Thus, minor non-specific expression of Cre and UPRT in a given tissue can occur, though this issue is relevant to all methods that depend on cell-specificity of a promoter. It is strongly advisable to use a well-studied Cre line or to conduct additional validations on the specificity of the Cre expression, tailored to one's specific experimental design. Since the transgenic UPRT expressed in the mouse has haemagglutinin (HA) attached, its cellular expression pattern can be validated by a standard immunohistochemistry method against HA [2].

Experimental design

Mouse breeding— *UPRT* mice and *Cre* mice are crossed to generate the double-transgenic mice with UPRT transgene, and Cre transgene under a cell-specific promoter (Cre + mice) (Fig. 1b). As a control, mice only with UPRT transgene (Cre-) should also be generated. For this purpose, cross between homozygous *UPRT* mice (*uprt/uprt*) and hemizygous *Cre* mice (*cre*/0) is recommended as 1:1 ratio of Cre+ and Cre- mice can be obtained in the same litter. Minimum of two mice for each condition are necessary to perform the statistical test used to identify labelled genes in the downstream bioinformatic analysis. We have used both male and female C57BL/6NTac mice with the age between 8-10 weeks.

4-thiouracil administration—Both Cre+ and Cre- animals should be given the same 4-thiouracil solution in a consistent way. Although, in this protocol, we describe the administration method with intraperitoneal (i.p.) injection [2,3,15], we have confirmed that subcutaneous (s.c.) injection also worked with maintaining the same labelling efficiency. It is recommended to allow enough time between injections so that each mouse is exposed to the solution for approximately the same duration upon culling. Also, the exposure time to 4-thiouracil may be adjusted according to one's experimental aims. Though we have confirmed that the majority of mRNA was significantly labelled in three different cell types by 4-hour exposure to 4-thiouracil, a longer exposure time may be employed when studying a cell that is known to have a less active transcriptional state or RNAs that have a longer turnover time.

Tissue harvest—Care must be taken to avoid RNA degradation when collecting tissues. Although we have used RNAlater for tissue preservation, any other standard tissue storage methods that are compatible with a subsequent RNA extraction step could be used (e.g. snap freezing).

RNA extraction—Although we have used an organic extraction method using TRIsure, other standard RNA extraction methods should be applicable. To avoid oxidation of the thiol

group, it is important to add dithiothreitol (DTT) to a final concentration of 100 μM when performing the isopropanol precipitation step and dissolve the obtained RNA pellet in 1 mM DTT solution.

RNA quality check with Bioanalyzer RNA chip or equivalent should be performed. Although QuantSeq kit is compatible with degraded RNA, all the samples should have a similar RNA integrity number (RIN) for consistent results.

To confirm the transgene expression, it is recommended to perform reverse transcription followed by quantitative PCR (RT-qPCR) against UPRT (Box 1).

Iodoacetamide (IAA) treatment—IAA solution must be made freshly just before a reaction is performed. It is recommended to include a reaction that alkylates pure 4-thiouracil solution as a positive control and to confirm successful alkylation by a spectrophotometer (details are described in the step 26).

RNA-seq library preparation—We have followed the standard protocol of QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen). Before performing a sequencing run, library quality check with Bioanalyzer high-sensitivity chip or equivalent should be performed. We have multiplexed up to 6 libraries per lane on a flow cell. Multiplexing with more samples could potentially be achieved, where higher depth is available.

Bioinformatic analysis—After a set of basic quality checks of the reads with FastQC are performed, T>C aware alignment followed by T>C counting is conducted by the software, SLAM-DUNK, which was specifically developed for SLAMseq data analyses. Although various analyses could be performed on the obtained T>C count table depending on the purpose of an experiment, we have used beta-binomial regression [26] to identify which genes have significantly higher T>C rate in Cre+ compared to Cre-.

Expertise needed to implement the protocol—Since this protocol involves generation of transgenic mice, an appropriate mouse facility is needed to maintain a small number of mice. Either i.p. injection or s.c. injection must be performed by a person who is able to inject a small volume of solution constantly without injuring surrounding organs.

SLAMseq data analysis with SLAM-DUNK requires a high performance computing system with a Linux operating system. Although basic command line usage is required to run SLAM-DUNK pipelines, and, for the downstream analyses, knowledge about a language such as R or Python is beneficial, Lexogen now offers a software on a web-server for SLAMseq analyses (http://www.bluebee.com/lexogen/). Hence, basic analyses could be performed with little or no programming skill.

Materials

REAGENTS

- B6;D2-Tg(CAG-GFP,-Uprt)985Cdoe/J (*UPRT* mice; The Jackson Laboratory, stock no. 021469)

- Cre driver line (*Cre* mice)

  CAUTION: All the mouse procedures must be conducted at an appropriate animal facility and both governmental and institutional laws and guidelines must be followed. All the animal experiments involved in the development of this protocols were conducted in accordance with UK Home Office regulations, UK Animals (Scientific Procedures) Act of 1986 under a UK Home Office license that approved this work (PF8733E07), which was reviewed regularly by the Wellcome Sanger Institute Animal Welfare and Ethical Review Body.

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

  CAUTION: 4-thiouracil is harmful by inhalation, in contact with skin, and if swallowed.

  Gloves and a lab coat must be worn when handling.

- Dimethyl sulfoxide (DMSO; Sigma-Aldrich, cat. no. D1435-500ML)

- Corn oil (Santa Cruz Biotechnology, cat. no. sc-214761)

- RNAlater (Thermo Fisher Scientific, cat. no. AM7021)

- TRIsure (Bioline, cat. no. BIO-38033)

  CAUTION: TRIsure is toxic and corrosive. Always handle with gloves, a lab coat, and protective glasses in a chemical fume hood.

- Chloroform – isoamyl alcohol mixture (Sigma-Aldrich, cat. no. 25666-100ML)

  CAUTION: Chloroform is toxic if inhaled and a potential carcinogen. Always handle with gloves, a lab coat, and protective glasses in a chemical fume hood.

- Glycogen, Molecular Biology Grade (Roche, cat. no. 10899232103)

- Dithiothreitol (DTT; Promega, cat. no. V3151)

  CAUTION: DTT is harmful if swallowed and causes skin irritation upon direct contact.

  Gloves and a lab coat must be worn when handling.

- Nuclease-Free Water, not DEPC-Treated (Thermo Fisher Scientific, cat. no. AM9932)

- Ethanol absolute 99.8% (VWR, cat. no. 20821.321)

- Qubit RNA BR Assay kit (Thermo Fisher Scientific, cat. no. Q10210)

- Agilent RNA 6000 Nano Kit (Agilent, cat. no. 5067-1511)

- TURBO DNA-free Kit (Thermo Fisher Scientific, cat. no. AM1907)

- RNA Clean & Concentrator-5 (Zymo Research, cat. no. R1013)

- SuperScript II Reverse Transcriptase (Thermo Fisher Scientific, cat. no. 18064014)

- dATP Solution, 100 mM (Thermo Fisher Scientific, cat. no. R0141)

- dCTP Solution, 100 mM (Thermo Fisher Scientific, cat. no. R0151)

- dGTP Solution, 100 mM (Thermo Fisher Scientific, cat. no. R0161)

- dTTP Solution, 100 mM (Thermo Fisher Scientific, cat. no. R0171)

- Random Hexamers (50 μM; Thermo Fisher Scientific, cat. no. N8080127)

- PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, cat. no. A25779)

- Iodoacetamide (IAA; Sigma-Aldrich, cat. no. I1149-5G)

  CAUTION: IAA is toxic if swallowed and causes skin burn and irritation. Gloves and a lab coat must be worn when handling.

- Sodium phosphate monobasic (Sigma-Aldrich, cat. no. S0751)

- Sodium Acetate Solution (3 M), pH 5.2 (Thermo Fisher Scientific, cat. no. R1181)

- QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, cat. no. 015.24)

- Bioanalyzer DNA High Sensitivity kit (Agilent, cat. no. 5067-4626)

- Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, cat. no. Q32851)

EQUIPMENT

- Filtered pipette tips

CRITICAL: To prevent contamination of RNase present in environment, filtered tips should be used when handling solution containing RNA.

- DNase-free microcentrifuge tubes (1.5 ml and 2 ml)

- 8-strip PCR tubes (0.2 ml)

CRITICAL: To avoid RNA degradation, any equipment that directly contacts with RNA should be nuclease-free.

- 1ml Syringe Plastipak Luer Slip (BD, cat. no. 303172)

- Terumo AGANI Needle 25G Orange x 5/8" (Terumo, cat. no. AN-2516R)

- Stainless steel beads, 7 mm (Qiagen, cat. no. 69990)

- TissueLyser Single-Bead Dispenser, 7 mm (Qiagen, cat. no. 69967)

- TissueLyser LT (Qiagen)

- Nanodrop 1000 (Thermo Fisher Scientific)

- Qubit assay tubes (Thermo Fisher Scientific, cat. no. Q32856)

- Qubit 2.0 fluorometer (Thermo Fisher Scientific)

- Agilent Bioanalyzer 2100 (Agilent)

- MicroAmp™ Fast Optical 96-Well Reaction Plate with Barcode, 0.1 mL (Thermo Fisher Scientific, cat. no. 4346906)

- StepOnePlus (Thermo Fisher Scientific)

- Illumina Hiseq 2000 (Illumina)

SOFTWARE

- Linux/Unix operating system

- FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

- SLAM-DUNK: http://t-neumann.github.io/slamdunk/

- MultiQC: http://multiqc.info/

- Rstudio: https://www.rstudio.com/

REAGENT SETUP

4-thiouracil solution—Prepare 200 mg/ml stock solution in DMSO, which can be stored at -20 °C for one year. On the day of injection, dissolve the stock solution in corn oil in 1:4 ratio. Shake vigorously just before injection to achieve homogeneity. Since 4-thiouracil is UV-sensitive, exposure to light should be kept minimal.

80% (vol/vol) ethanol—Always dilute absolute ethanol in nuclease-free water freshly.

Glycogen solution (20 mg/ml)—Dissolve glycogen in nuclease-free water to 20 mg/ml. The solution should be aliquoted into sterile tubes (~100 μl per tube) to avoid excess freeze-thaw cycles. This solution can be stored at -20 °C for one year.

dNTP mix (10 mM)—Mix one volume of 100 mM dATP, dTTP, dGTP, dCTP solution each and add 6 volume of nuclease-free water. The solution should be aliquoted into sterile tubes (~100 μl per tube) to avoid excess freeze-thaw cycles. The mix can be stored at -20 °C for one year.

DTT—Prepare 1 M stock solution in nuclease-free water and store it at -20 °C up to one year. The solution should be aliquoted into sterile tubes (~100 μl per tube) to avoid excess freeze-thaw cycles. On the day of use, dissolve the stock solution in nuclease-free water to make either 100 mM or 1 mM solution for reactions.

IAA—Freshly prepare 100 mM solution in absolute ethanol on the day of use.

1 M Sodium phosphate (pH 8)—Dissolve 11.998 g sodium phosphate monobasic in 100 ml ultrapure water. Autoclave the solution and adjust its pH using sodium hydroxide and hydrochloric acid. Alternatively, use sodium phosphate dibasic to adjust the pH of the buffer.

The solution can be stored at room temperature (20 °C) for one year.

Procedure

4-thiouracil administration (TIMING: 5 h)

1. Prepare 50 mg/ml 4-thiouracil solution in DMSO and corn oil (see REAGENT SETUP for further details).

    CAUTION: 4-thiouracil is harmful by inhalation, in contact with skin, and if swallowed.

    Gloves and a lab coat must be worn when handling.

    CRITICAL STEP: 4-thiouracil is UV-sensitive, and thus its exposure to UV should be limited. It is advised to cover the tube with a foil if the solution could be exposed to the sunlight.

    ?TROUBLESHOOTING

2. Weigh mice and calculate the volume to inject to achieve the dose of 400 mg/kg body weight.

3. I.p. inject the 4-thiouracil solution using a 25G x $\frac{5}{8}$" needle.

    CRITICAL STEP: shake the solution vigorously just before each injection to homogenise the solution since the solution can easily get phase-separate.

    CRITICAL STEP: Allow enough time between injections to each mouse so that each mouse is exposed to the solution for the same duration at the time of culling.

4. Return the mouse cages to the original holding space and allow minimum of 4 h exposure to the solution.

    CRITICAL STEP: Longer exposure time and repeated injections might increase the labelling level of RNA especially the ones with a longer half life.

Tissue harvest (TIMING: 2 h)

5. Prepare 1.5-ml tubes to collect tissues and fill these with 1 ml RNAlater.

6. Cull the mice one by one and dissect out the tissues of interest and submerge in RNAlater.

    CRITICAL STEP: Cut the tissues into pieces so a thickness of any one dimension does not exceed 5 mm to make sure RNAlater spread effectively.

PAUSE POINT: Tissues in RNAlater can be stored at 4 °C for one month. A longer storage can be achieved at -20 °C and below after storing at 4 °C overnight to allow the RNAlater to diffuse into the cells.

RNA extraction (TIMING: 4 h)

7. Dispense a 7-mm bead in a 2-ml tube using a TissueLyser bead dispenser.

8. Take a piece of tissue out of RNAlater and remove residual RNAlater using a clean piece of paper (e.g. Kimwipe).

9. Cut into a small piece so it does not exceed 50 mg and put the piece in the 2-ml tube prepared in the step 7.

   CRITICAL STEP: Maximum weight of a tissue allowed can be varied depending on the tissues. Refer to the instruction of TRIsure to adjust it accordingly.

10. Dispense 1 ml of TRIsure in the 2-ml tube, load the tubes in a TissueLyser, and run the TissueLyser following the manufacturer's instruction. Incubate at room temperature for 5 min after the homogenisation.

    CAUTION: TRIsure is toxic and corrosive. Always handle with gloves, a lab coat, and protective glasses in a chemical fume hood.

    CRITICAL STEP: From this step up to the step 28, the solution containing thiolated RNA should be protected from light, since the thiol group is UV-sensitive.

11. Transfer the solution to a new 1.5-ml tube and add 200 μl chloroform-isoamyl alcohol mix.

    CAUTION: Chloroform is toxic if swallowed and a potential carcinogen. Always handle with gloves, a lab coat, and protective glasses in a chemical fume hood.

12. Shake vigorously for 15 s and incubate for 3 min at room temperature.

13. Centrifuge at 12,000g for 15 min at 4 °C.

14. Carefully transfer the upper aqueous phase to a new 1.5-ml tube.

15. Add 1 μl glycogen solution (20 mg/ml), 1 μl 100 mM DTT, and 500 μl cold isopropanol. Mix vigorously.

    CAUTION: DTT is harmful if swallowed and causes skin irritation upon direct contact.

    Gloves and a lab coat must be worn when handling.

    CRITICAL STEP: Always add freshly prepared DTT solution to prevent oxidation of the thiol group introduced to the RNA.

16. Incubate at -20 °C for at least 2 h.

    PAUSE POINT: Alternatively, the incubation can be extended indefinitely.

17. Centrifuge at 12,000g for 30 min at 4 °C.

18. Remove the supernatant and wash the pellet with 1 ml cold 80% (vol/vol) ethanol.

19. Centrifuge at 12,000g for 10 min at 4 °C.

20. Remove the supernatant and air dry the pellet for 4 min at room temperature.

21. Resuspend the pellet in 1 mM DTT.

    CRITICAL STEP: Always use freshly prepared DTT solution to prevent oxidation of the thiol group introduced to the RNA.

22. Confirm RNA concentration and RNA integrity with Qubit BR kit and Bioanalyzer Nano kit. Continue with IAA treatment or store the RNA solution at -80 °C.

PAUSE POINT: The RNA solution can be stored at -80 °C indefinitely.

?TROUBLESHOOTING

DNase treatment (TIMING: 1 h)

23. Set up the reaction following the instruction of TURBO DNA-free Kit.

24. Incubate for 20 min at 37 °C.

25. Clean up RNA with RNA Clean & Concentrator-5. Add DTT to the eluted RNA solution to the final concentration of 1 mM. Successful transgene expression can be confirmed using this RNA solution (BOX 1).

CRITICAL STEP: Always add freshly prepared DTT solution to prevent oxidation of the thiol group introduced to the RNA.

PAUSE POINT: The RNA solution can be stored at -80 °C indefinitely.

Iodoacetamide (IAA) treatment (1 h)

26. Set up following IAA reaction mix in a 1.5-ml tube.

Component	Amount per reaction (µl)	Final concentration
RNA (~1 µg)	x	
DMSO	25	50% (vol/vol)
100 mM IAA in ethanol	5	10 mM
500 mM sodium phosphate, pH 8	5	50 mM
Nuclease-free water	15 - x	
Total	50	

CRITICAL STEP: Always prepare IAA solution freshly to have an efficient alkylation reaction.

CRITICAL STEP: To confirm successful alkylation, it is strongly recommended to prepare the following reaction containing 4-thiouracil instead of RNA as a positive control and check its absorbance peak. Dilute the resulting solution to 1/10 and load 2 µl for a spectrophotometer assay. Successful alkylation can be confirmed with the shift of an absorbance peak from 330-335 nm (4-thiouracil) to 295 nm (alkylated 4-thiouracil) by a spectrophotometer (e.g. UV-Vis mode on Nanodrop) (Fig. 2).

Component	Amount per reaction (μl)	Final concentration
10 mM 4-thiouracil solution	1	
DMSO	5	50% (vol/vol)
100 mM IAA (or EtOH for negative control)	1	10 mM
500 mM sodium phosphate, pH 8	1	50 mM
Nuclease-free water	2	
Total	10	

?TROUBLESHOOTING

27. Incubate for 15 min at 50 °C on a heat block.

28. Stop the reaction by adding 1 μl 1 M DTT. Mix well.

29. Add 1 μl glycogen solution (20 mg/ml), 5 μl sodium acetate (3M, pH 5.2), and 180 μl cold ethanol absolute. Mix vigorously.

30. Incubate for 2 h at -20 °C.

    PAUSE POINT: Alternatively, the incubation can be extended indefinitely.

31. Centrifuge at 12,000g for 30 min at 4 °C.

32. Remove the supernatant and wash the pellet with 1 ml cold 80% (vol/vol) ethanol.

33. Centrifuge at 12,000 g for 10 min at 4 °C.

34. Remove the supernatant and air dry the pellet for 4 min at room temperature.

35. Resuspend the pellet in nuclease-free water. Continue with RNA-seq library preparation or store the alkylated RNA solution at -80 °C.

    PAUSE POINT: The alkylated RNA solution can be stored at -80 °C indefinitely.

RNA-seq library preparation (4.5 h, 1 h 45 min hands-on time)

36. Follow the instruction of QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen) to prepare RNA-seq libraries. Use different indices for different samples for multiplexing.

    CRITICAL STEP: Although the minimum total RNA input for QuantSeq is 500 pg, it is recommended to use higher amount of starting RNA to have more RNA molecules sequenced for a sensitive detection of labelled transcripts. We routinely start with around 500 ng total RNA.

37. Check library quality by Bioanalyzer high sensitivity kit.

High-throughput sequencing (1 d)

38. Multiplex equimolar amount of each library and perform high-throughput sequencing.

PAUSE POINT: RNA-seq libraries can be stored at 4 °C for a week or at -20 °C indefinitely.

Software installation (15 min)

39. Miniconda downloader installation.

```
wget https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh
```

40. Install miniconda. Follow the command prompt.

```
bash Miniconda3-latest-Linux-x86_64.sh
```

41. Set up channels where the necessary softwares are located.

```
conda config --add channels defaults
conda config --add channels conda-forge
conda config --add channels bioconda
```

42. Install softwares required for the analyses into a custom environment.

```
conda create --name slamseq slamdunk fastqc multiqc
```

43. Activate the environment.

```
source activate slamseq
```

44. Download Rstudio from: https://www.rstudio.com/

Genomic files download (15 min)

45. Go to ensembl (https://www.ensembl.org/Mus_musculus/Info/Index) and download mouse primary genome assembly from "Download DNA sequence" link. Download the file ending with "dna.primary_assembly.fa.gz".

46. Go to NCBI Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables).

47. Set the parameters as following:

- clade: Mammal
- genome: Mouse
- assembly: *the same version as the one downloaded in the step 46
- group: Genes and Gene Predictions
- track: *select the one that fits with your experimental purpose
- region: genome
- output format: BED - browser extensible data

48. Click on "get output".

49. Select "Create one BED record per: 3' UTR Exons".

50. Click on "get BED" to download mouse 3' UTR annotation.

Bioinformatic analysis (5 h)

51. Generate FASTQ files from the row sequencing data and demultiplex based on the indices used.

52. Check sequence quality with FastQC.

```
fastqc path/to/fastq
```

53. Run SLAM-DUNK on high performance computing. The details of the parameters are outlined on http://t-neumann.github.io/slamdunk/docs.html.

    CRITICAL STEP: The run requires 5 threads and 20 GB of memory per sample and takes around 2 h to complete all the processes of SLAM-DUNK pipeline.

```
slamdunk all -r path/to/fasta -b path/to/bed -t 5 -5 12 -n 100 -m -mv 0.2 -c
2 -rl 100 path/to/fastq
```

54. Assess overall conversion rate by `utrrates` function.

```
alleyoop utrrates -b path/to/bed path/to/filtered_bam
```

55. Create summary graphs by MultiQC.

```
multiqc path/to/utrrates_output
```

56. Analyse SLAM-DUNK output data using RStudio. Especially, T>C count table is available as `count/<sample_name>_tcount.tsv`. Use ibb package's `bb.test()` function to identify differentially labelled genes. Assign the "ConversionsOnTs" column to "x" in bb.test() and the "CoverageOnTs" column to "tx".

    CRITICAL STEP: It is recommended to always confirm expression between Cre + and Cre- is highly correlated to ensure 4-thiouracil incorporation itself does not affect RNA metabolism.

    ?TROUBLESHOOTING

Timing

Steps 1-4, 4-thiouracil administration: 5 h

Steps 5+6, Tissue harvest: 2 h

Steps 7-22, RNA extraction: 4 h

Steps 23-25, DNase treatment: 1 h

Steps 26-35, Iodoacetamide (IAA) treatment: 1 h

Steps 36+37, RNA-seq library preparation: 4.5 h

Step 38: High-throughput sequencing: 1 d

Steps 39-44, Software installation: 15 min

Steps 45-50, Genomic files download: 15 min

Steps 48-56, Bioinformatic analysis: 5 h

Box 1, RT-qPCR: 1.5 h

Troubleshooting

Step	Problem	Possible reason	Solution
1	4-thiouracil does not dissolve	Temperature of the solution is too low.	Warm up the solution to 37 °C and shake vigorously using Vortex.
22	RIN lower than 8	RNA degradation	Store tissues immediately after collection and make sure to always wear gloves when handling.
Box 1	UPRT expression not confirmed	Inappropriate cross setup	Check parental genotypes to see if an appropriate pair was chosen for the cross.
26	Shift in absorption peak not observed	Alkylation was unsuccessful	Confirm all the reagents were prepared properly. Especially, IAA solution should be prepared freshly.
56	Low expression correlation between Cre+ and Cre-	Too high 4-thiouracil concentration used	Repeat the experiment with different doses of 4-thiouracil to determine the highest possible concentration.

Anticipated results

RNA obtained from tissues through this RNA extraction method should have RNA integrity number (RIN) of 8 or higher. Lower RIN value could mean mishandling of samples when harvesting tissues, but some tissues generally have a relatively lower RIN value compared with the other tissues (e.g. intestine). In the RT-qPCR results, we observed ⌂Ct (UPRT - Hprt) of 3-4 in Cre- tissues. ⌂Ct in Cre+ varied in different tissues depending on the proportion of Cre-expressing cells in the tissue. The IAA treatment step should have a minimum impact on RNA quantity and quality. By following QuantSeq kit protocol, each resulting library should be more than 10 nM in concentration and have a single peak around 250-300 bp region in Bioanalyzer results.

When high-throughput sequencing is performed by multiplexing 6 QuantSeq libraries per lane, each library should yield around 50 million reads, and 200-300 median mapped reads per gene. Although T>C rate might be varied depending on the ratio of Cre-expressing cells to the rest of the cells in a tissue from which RNA is extracted, we have observed the median T>C rate of 0.1% (endothelial cells in brain) and 1% (epithelial cells in intestine) [15]. Typical results from multiQC reports about conversion rate per 3'UTR are attached (Supplementary

data 1). Beta-binomial test identified around ~5,000 endothelial genes in mouse brain (Fig. 3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Richardson GM, Lannigan J, Macara IG. Does FACS perturb gene expression? Cytometry. 2015; 87:166–175. [PubMed: 25598345]

2. Gay L, et al. Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes Dev. 2013; 27:98–115. [PubMed: 23307870]

3. Gay L, Karfilis KV, Miller MR, Doe CQ, Stankunas K. Applying thiouracil tagging to mouse transcriptome analysis. Nat Protoc. 2014; 9:410–420. [PubMed: 24457332]

4. Miller MR, Robinson KJ, Cleary MD, Doe CQ. TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat Methods. 2009; 6:439–441. [PubMed: 19430475]

5. 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. Nat Biotechnol. 2005; 23:232–237. [PubMed: 15685165]

6. Chatzi C, Zhang Y, Shen R, Westbrook GL, Goodman RH. Transcriptional Profiling of Newly Generated Dentate Granule Cells Using TU Tagging Reveals Pattern Shifts in Gene Expression during Circuit Integration. eNeuro. 2016; 3

7. Tomorsky J, DeBlander L, Kentros CG, Doe CQ, Niell CM. TU-Tagging: A Method for Identifying Layer-Enriched Neuronal Genes in Developing Mouse Visual Cortex. eNeuro. 2017; 4

8. Erickson T, Nicolson T. Identification of sensory hair-cell transcripts by thiouracil-tagging in zebrafish. BMC Genomics. 2015; 16:842. [PubMed: 26494580]

9. Duffy EE, et al. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Mol Cell. 2015; 59:858–866. [PubMed: 26340425]

10. Herzog VA, et al. Thiol-linked alkylation of RNA to assess expression dynamics. Nat Methods. 2017; 14:1198–1204. [PubMed: 28945705]

11. Herzog VA, Reichholf B, Ameres S. Thiol-linked alkylation for the metabolic sequencing of RNA (SLAMseq). Protocol Exchange. 2017; doi: 10.1038/protex.2017.105

12. Watanabe S, Hino A, Wada K, Eliason JF, Uchida T. Purification, cloning, and expression of murine uridine phosphorylase. J Biol Chem. 1995; 270:12191–12196. [PubMed: 7744869]

13. Cheng N, Payne RC, Traut TW. Regulation of uridine kinase. Evidence for a regulatory site. J Biol Chem. 1986; 261:13006–13012. [PubMed: 3020019]

14. Carter D, Donald RG, Roos D, Ullman B. Expression, purification, and characterization of uracil phosphoribosyltransferase from Toxoplasma gondii. Mol Biochem Parasitol. 1997; 87:137–144. [PubMed: 9247925]

15. Matsushima W, et al. SLAM-ITseq: sequencing cell type-specific transcriptomes without cell sorting. Development. 2018; 145

16. Muhar M, et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science. 2018; 360:800–805. [PubMed: 29622725]

17. Sarkies P, Miska Ea. Small RNAs break out: the molecular cell biology of mobile small RNAs. Nat Rev Mol Cell Biol. 2014; 15:525–535. [PubMed: 25053358]

18. Sharma U, et al. Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. Dev Cell. 2018; 46:481–494.e6. [PubMed: 30057273]

19. Hupe M, Li MX, Gertow Gillner K, Adams RH, Stenman JM. Evaluation of TRAP-sequencing technology with a versatile conditional mouse model. Nucleic Acids Res. 2014; 42:e14. [PubMed: 24165879]

20. Hwang H-W, et al. cTag-PAPERCLIP Reveals Alternative Polyadenylation Promotes Cell-Type Specific Protein Diversity and Shifts Araf Isoforms with Microglia Activation. Neuron. 2017; 95:1334–1349.e5. [PubMed: 28910620]

21. Riml C, et al. Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. Angew Chem Int Ed Engl. 2017; 56:13479–13483. [PubMed: 28817234]

22. Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat Methods. 2018; 15:221–225. [PubMed: 29355846]

23. Baptista MAP, Dölken L. RNA dynamics revealed by metabolic RNA labeling and biochemical nucleoside conversions. Nat Methods. 2018; 15:171–172. [PubMed: 29489745]

24. Burger K, et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. 2013; 10:1623–1630. [PubMed: 24025460]

25. Heffner CS, et al. Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource. Nat Commun. 2012; 3

26. Pham TV, Piersma SR, Warmoes M, Jimenez CR. On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. Bioinformatics. 2010; 26:363–369. [PubMed: 20007255]

BOX 1

RT-qPCR (TIMING: 1.5 h)

1. Dilute RNA to 200 ng/µl in nuclease-free water.

2. Set up following reaction mix in an 8-strip tube.

Component	Amount per reaction (µl)	Final concentration
RNA	2	
dNTP mix (10 mM)	1	0.50 mM
Random hexamers (50 µM)	1	2.5 µM
Nuclease-free water	8	
Total	12	

3. Heat mixture to 65 °C for 5 min and quick chill on ice.

4. Add 4 µl of 5X First-Strand Buffer, 2 µl of 0.1 M DTT, and 1 µl of RNaseOUT (40 units/µl) to each reaction mix. Mix well and incubate at 25 °C for 2 min.

5. Add 1 µL (200 units) of SuperScript II reverse transcriptase. Mix well by pipetting up and down.

6. Run following reaction on a thermal cycler.

Temperature (°C)	Time (min)
25	10
42	50
70	15
4	Forever

7. Add 30 μl of nuclease-free water to the cDNA mix to make it 50 μl.

   PAUSE POINT: cDNA solution can be stored at -20 °C indefinitely.

8. Assemble following reaction mix in optical 96-well plate. Use primers listed in Supplementary table 1.

Component	Amount per reaction (µl)	Final concentration
Nuclease-free water	3	
2X PowerUp SYBR Green master Mix	5	1X
10 µM forward primer	0.5	1 µM
10 µM reverse primer	0.5	1 µM
Total	9	

9. Add 1 µl cDNA (~8 ng) to each reaction mix. Preparing at least three technical replicates per sample is recommended.

10. Run StepOnePlus in "Fast" mode.

CRITICAL STEP: Perform a melt curve analysis at least for the first run to confirm a single amplicon is generated.
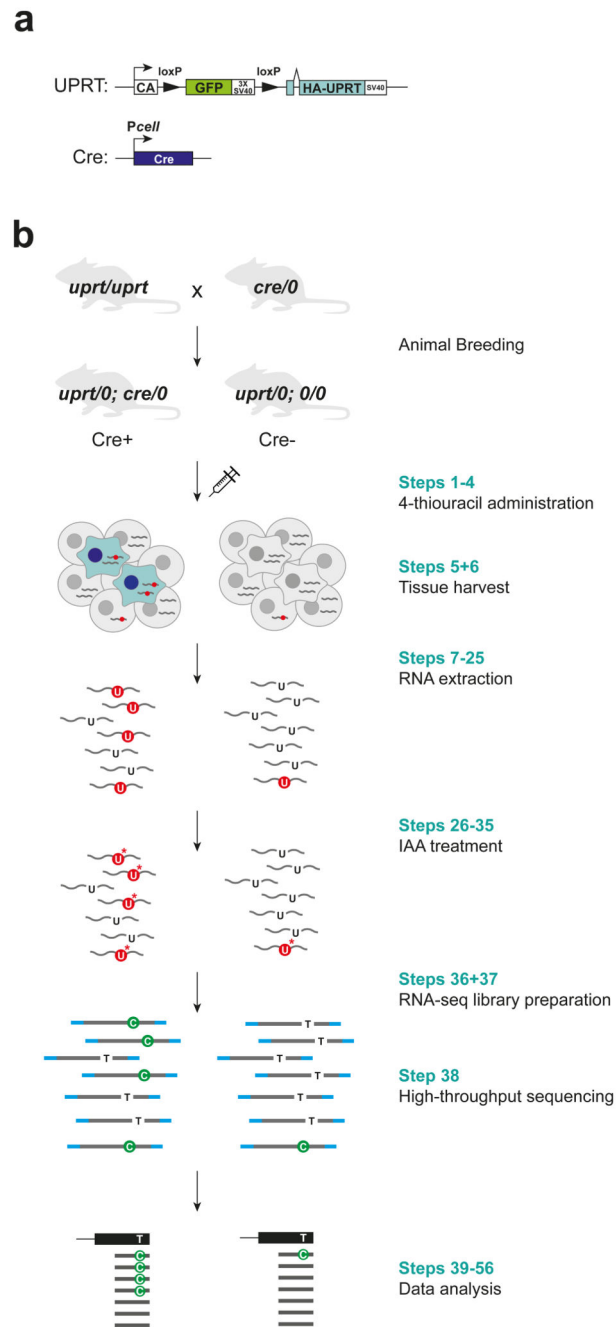
?TROUBLESHOOTING

Figure 1. Flow chart of SLAM-ITseq protocol

CA represents chicken beta actin promoter; P*cell*, cell-specific promoter; wavy line, RNA; red circle, thiolated uracil; asterisk, carboxyamidomethyl group.
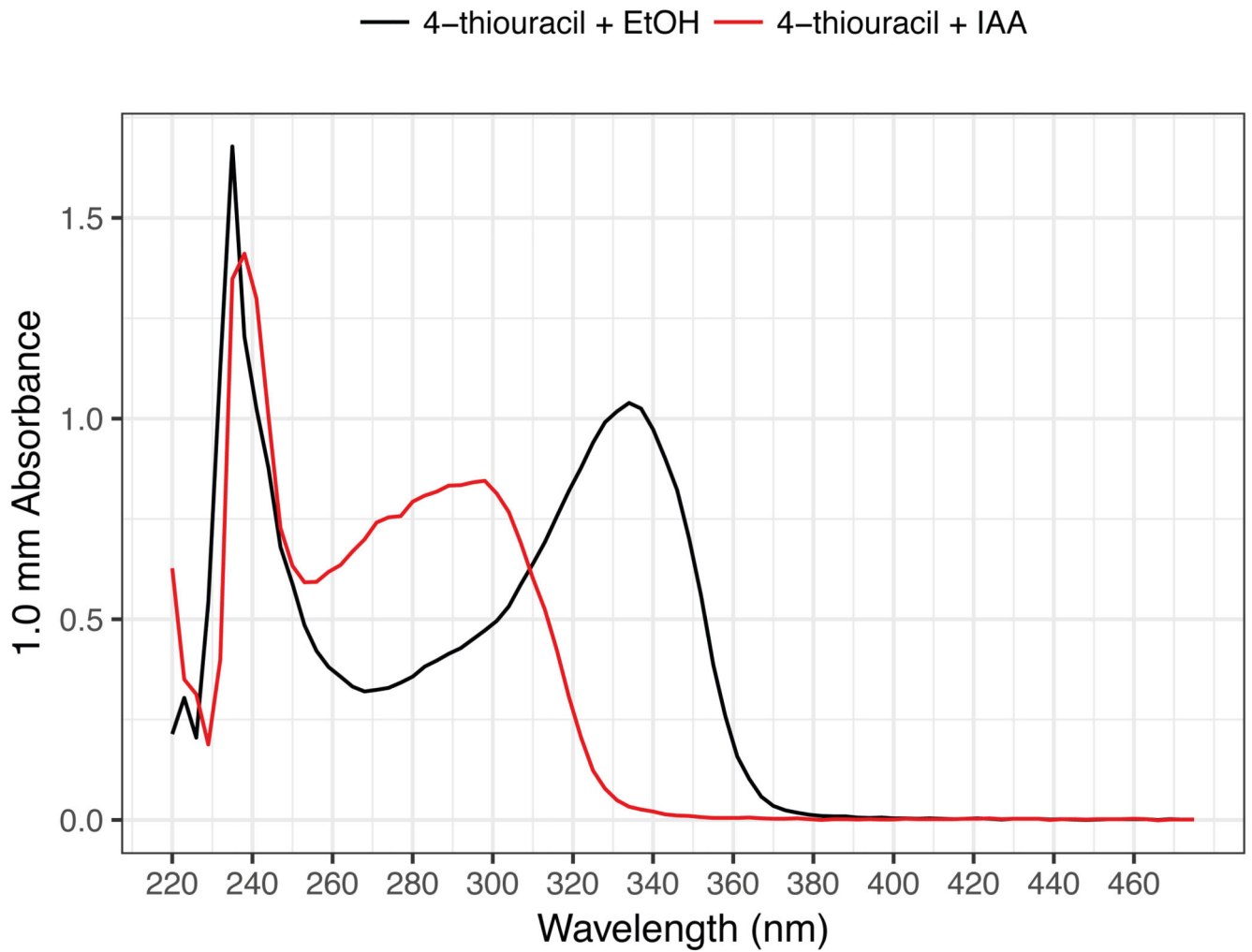
Figure 2. Typical spectrophotometer results of 4-thiouracil before and after IAA treatment
A shift in the peak absorbance from 335 nm to 300 nm is observed after IAA treatment.
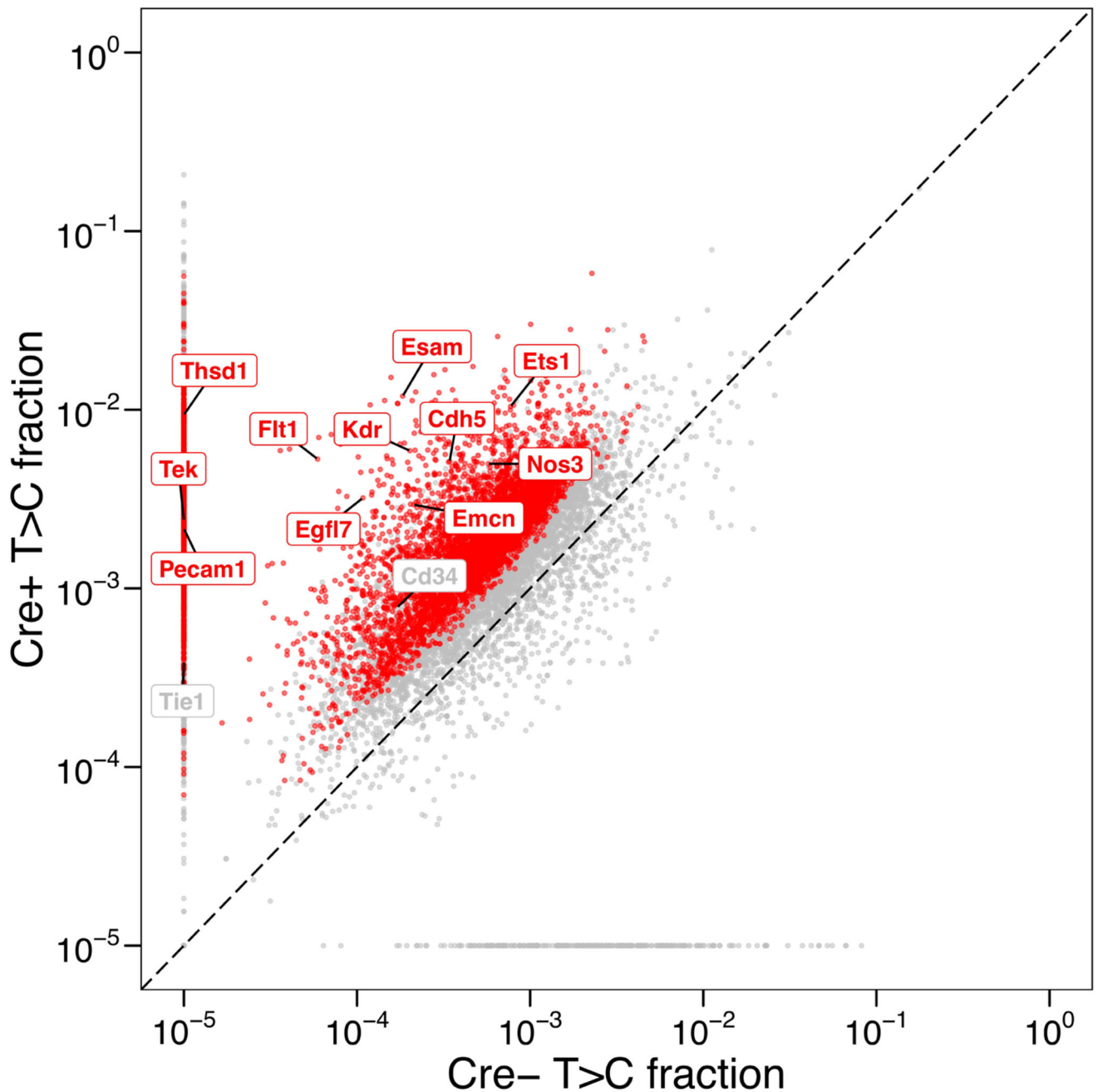
Figure 3. Typical T>C fraction results in Cre+ and Cre- mice
Each point represents T>C fraction of a single gene in two mouse genotypes, Cre+ and Cre-.
Red points represent genes significantly more labelled in Cre+ compared to Cre- (FDR <
0.05, beta-binomial test). Known endothelium-specific genes are labelled with their gene
names. This plot is adopted from Matsushima et al[12]. This animal experiment was conducted
in accordance with UK Home Office regulations, UK Animals (Scientific Procedures) Act of
1986 under a UK Home Office license that approved this work (PF8733E07), which was

reviewed regularly by the Wellcome Sanger Institute Animal Welfare and Ethical Review Body.