RESOURCE/METHODOLOGY

Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA

Leslie Gay,[1,2,3] Michael R. Miller,[1,2,3] P. Britten Ventura,[2] Vidusha Devasthali,[2] Zer Vue,[4] Heather L. Thompson,[4] Sally Temple,[5] Hui Zong,[2] Michael D. Cleary,[4] Kryn Stankunas,[2] and Chris Q. Doe[1,2,3,6]

[1]Institute of Neuroscience, [2]Institute of Molecular Biology, [3]Howard Hughes Medical Institute, University of Oregon, Eugene, Oregon 97403, USA; [4]School of Natural Sciences, University of California at Merced, Merced, California 95340, USA; [5]Neural Stem Cell Institute, Rensselaer, New York 12144, USA

Transcriptional profiling is a powerful approach for understanding development and disease. Current cell type-specific RNA purification methods have limitations, including cell dissociation trauma or inability to identify all RNA species. Here, we describe "mouse thiouracil (TU) tagging," a genetic and chemical intersectional method for covalent labeling and purification of cell type-specific RNA in vivo. Cre-induced expression of uracil phosphoribosyltransferase (UPRT) provides spatial specificity; injection of 4-thiouracil (4TU) provides temporal specificity. Only UPRT$^+$ cells exposed to 4TU produce thio-RNA, which is then purified for RNA sequencing (RNA-seq). This method can purify transcripts from spatially complex and rare (<5%) cells, such as *Tie2:Cre*$^+$ brain endothelia/microglia (76% validated by expression pattern), or temporally dynamic transcripts, such as those acutely induced by lipopolysaccharide (LPS) injection. Moreover, generating chimeric mice via UPRT$^+$ bone marrow transplants identifies immune versus niche spleen RNA. TU tagging provides a novel method for identifying actively transcribed genes in specific cells at specific times within intact mice.

[*Keywords*: brain; endothelial cells; heart; microglia; nascent RNA; transcriptional profiling]

Supplemental material is available for this article.

Received September 5, 2012; revised version accepted November 19, 2012.

The mammalian body is composed of a complex assembly of distinct cell types. Transcriptional profiling in mouse models can provide insights into how cell types develop and function, communicate with each other, contribute to disease, and are affected by therapeutic treatments. It has proven difficult, however, to monitor acute gene expression changes in a specific cell type within its native environment (e.g., before and after differentiation, cell migration, host/bacterial interactions, or neuronal activity).

The most common methods for isolating cells for transcriptional profiling—fluorescence-activated cell sorting (FACS), laser capture, manual dissection, and panning—require dissociation of the targeted cells from their host tissue (Lobo et al. 2006; Daneman et al. 2010; Bartfai et al. 2012; Guez-Barber et al. 2012; Sanna et al. 2012). While these methods are effective, they have practical and theoretical limitations. In practice, many of these methods are slow or laborious, require expensive equipment, or are not capable of isolating dispersed cell types. In theory, these methods run the risk of losing RNA in fine cellular processes (e.g., axons, dendrites, or glial processes) and can induce nonphysiological changes in gene expression during the dissociation procedure. For example, epithelial cells have apical junctions that act as signaling "hubs," and loss of junctional integrity triggers gene expression changes (Balda and Matter 2009; Stepniak et al. 2009). Mechanical purification of epithelial cells may thus alter gene expression prior to transcriptional profiling. As a response to these problems, several genetic methods have been recently developed for isolating RNA from specific cell types from within intact tissues without the need for cell dissociation (Heiman et al. 2008; He et al. 2012). These methods succeed in avoiding dissociation trauma, but they,

[6]**Corresponding author**
E-mail cdoe@uoneuro.uoregon.edu
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.205278.112. Freely available online through the *Genes & Development* Open Access option.

too, have limitations. Each method only isolates a subset of cellular RNA (messenger RNA [mRNA] or microRNA [miRNA]), each requires overexpression of an endogenous mouse protein that could have deleterious effects, and each provides only limited temporal control of labeling.

Here we describe a method called mouse thiouracil (TU) tagging that is based on our previous work in *Drosophila* (Miller et al. 2009) and pioneering work in cell culture (Cleary et al. 2005). TU tagging is a genetic and chemical intersectional approach that allows covalent labeling of actively transcribed RNAs in specific cell types within intact mice. Spatial specificity is obtained by Cre-induced expression of a transgene encoding uracil phosphoribosyltransferase (UPRT) (Fig. 1A, red). Temporal specificity is via injection of the uracil analog 4-thiouracil (4TU) (Fig. 1A, blue). Only the cell types expressing UPRT will efficiently incorporate 4TU into newly transcribed RNA, thereby covalently labeling cell type-specific nascent RNA. Importantly, production of the thio-RNA occurs within the intact tissue in living mice, thereby preserving normal cell interactions and organismal physiology during the window of RNA labeling (Fig. 1D). The thio-RNA is then in vitro-biotinylated, purified from total RNA, and used for gene expression analyses via next-generation sequencing (RNA-seq). TU tagging has been shown to have a negligible effect on gene expression in cell lines (Cleary et al. 2005), and ubiquitous expression of UPRT has no effect on viability in *Drosophila* (Miller et al. 2009) or mice (this study).

Our work goes beyond previous applications of this method in several important ways. First, we developed genetic tools permitting TU tagging to be used in mice, which is a major advance due to the complexity of mammalian cell types and difficulty in transcriptional profiling of rare or dispersed mammalian cell types. Second, we designed our genetic tools to work with the wide array of existing Cre lines, making the method simple, rapid, and flexible for studying any tissue or developmental stage. Third, we improved the biochemical purification procedure, adding an RNA fragmentation step that reduces background labeling. Fourth, we developed a custom computational pipeline for RNA-seq analysis to account for the high percentage of intron reads present in the purified nascent RNA. TU tagging is complementary to existing RNA purification methods and gives the mouse research community an additional resource for studying spatial and temporal gene expression patterns within intact tissues.

Results

Construction and validation of UPRT transgenes

The TU tagging method requires expression of UPRT in a cell type of interest. To take advantage of the many existing cell-specific and tissue-specific Cre lines, we designed a Cre-inducible UPRT transgene. This transgene incorporated a ubiquitous chicken *β-actin/CMV* (*CA*) promoter driving a *loxP-GFP-3xstop-loxP* cassette followed by a hemagglutinin (HA) epitope-tagged *UPRT* gene (subsequently called *CA>GFPstop>UPRT*) (Fig. 1B). The GFP cassette included three SV40 polyadenylation sequences to prevent transcription of *HA-UPRT* in the absence of Cre; all three were required to prevent read-

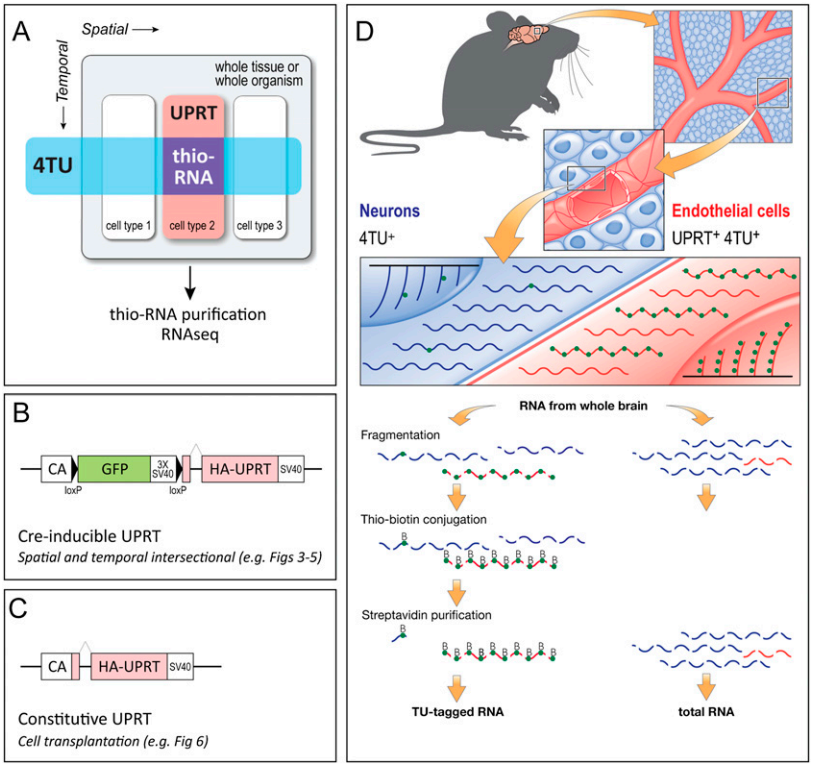


Figure 1. The mouse TU tagging method. (*A*) Schematic of the TU tagging method. Spatial control was provided by cell type-specific expression of UPRT (red), and temporal control was achieved by a pulse of 4TU (blue). Only UPRT$^+$ cells exposed to 4TU will generate thio-labeled newly transcribed RNA, which then can be purified from the intact tissue or organism. (*B*,*C*) Schematic of the UPRT transgenes used in this study. (*D*) Schematic of TU tagging of endothelial RNA within the intact brain. (Green dot) 4TU; (B) biotin.

through transcription. UPRT expression was monitored with an HA antibody and will be called "UPRT expression" for simplicity. In addition, we made a constitutively expressed *CA:HA-UPRT* transgene (subsequently called *CA:UPRT*) (Fig. 1C). We used this transgene for pilot experiments to optimize the method and for cell transplantation of UPRT$^+$ cells into UPRT$^-$ mice. The *CA:UPRT* transgenic line is viable and fertile despite widespread expression of UPRT in all tissues examined.

We next determined whether the *CA>GFPstop>UPRT* transgene was ubiquitously expressed and thus suitable for generating Cre-induced UPRT expression in a broad range of tissues. Control embryonic day 12.5 (E12.5) embryos without the *CA>GFPstop>UPRT* transgene had no GFP fluorescence, as expected (Fig. 2A), whereas *CA>GFPstop>UPRT* transgenic embryos showed widespread GFP expression (Fig. 2B). GFP expression was also observed in all examined organs at E12.5 and postnatal day 6 (P6) (Fig. 2C; data not shown). Thus, the *CA>GFPstop>UPRT* transgene should be useful for Cre-induced UPRT expression in many or all tissues.

To determine the efficiency of Cre-induced UPRT expression, we used *Tie2:Cre* because it is expressed in a well-characterized and distinctive pattern of endothelial cells in all tissues (Kisanuki et al. 2001) as well as in *Tie2:Cre* lineage-derived hematopoietic progenitors that include those giving rise to brain microglia/macrophages (Chen et al. 2010; Tang et al. 2010). First, we tested for *Tie2:Cre*-induced UPRT expression in a P6 brain. Control mice lacking the *Tie2:Cre* transgene showed no detect-

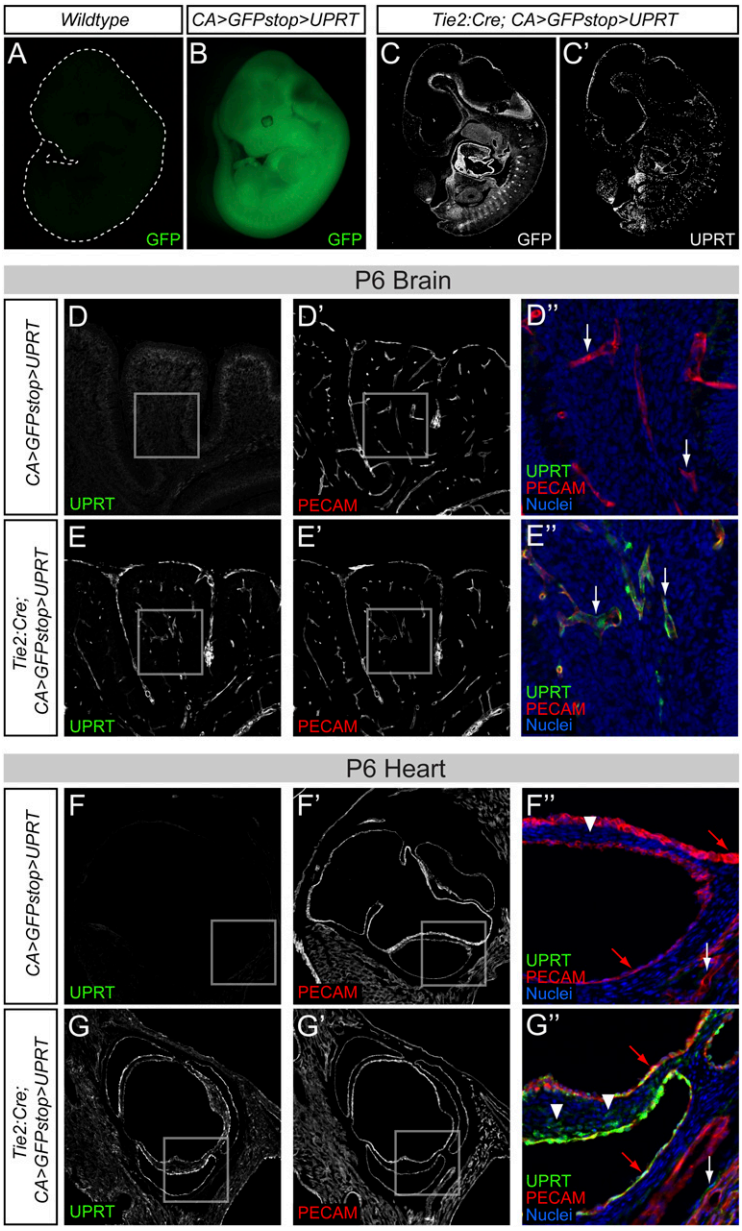


Figure 2. The *CA>GFPstop>UPRT* transgene was ubiquitously expressed and provided high-efficiency Cre-dependent UPRT expression. (*A–C*) E12.5 expression patterns. (*A*) The wild-type embryo has only minimal background autofluorescence. (*B*) The *CA>GFPstop>UPRT* single-transgenic embryo has uniform GFP expression. (*C,C′*) The antibody-stained section of a *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic E12.5 embryo shows persistent GFP expression where *Tie2:Cre* is not expressed and UPRT expression in the characteristic *Tie2:Cre* endothelial pattern. UPRT expression was detected by anti-HA antibody staining of the HA:UPRT fusion protein. (*D–E″*) P6 brain (cerebellum) staining patterns. (*D–D″*) *CA>GFPstop>UPRT* single-transgenic shows no UPRT expression. (*E–E″*) *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic shows robust UPRT expression in the PECAM1$^+$ endothelial cells. White arrows indicate endothelial cells. (*F–G″*) P6 heart (aortic valve region) staining patterns. (*F–F″*) *CA>GFPstop>UPRT* single-transgenic shows no UPRT expression. (*G–G″*) *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic shows robust UPRT expression in PECAM1$^+$ endothelial and endocardial cells and a subset of aortic valve interstitial cells. White arrows indicate endothelial cells, red arrows show aortic valve endocardial cells, and white arrowheads mark aortic valve interstitial cells. Scale: box dimensions, 300 μm.

able UPRT expression in the brain (Fig. 2D), whereas *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic mice showed robust UPRT expression in PECAM1$^+$ (aka CD31) endothelial cells of the cerebellum (Fig. 2E) and all other regions of the brain (e.g., cortex, dentate gyrus, midbrain, choroid plexus, and hypothalamus) (Supplemental Fig. S1). In all brain regions, we observed UPRT expression in ~100% of the PECAM1$^+$ endothelial cells, showing excellent efficiency in Cre-mediated excision of the *GFP:stop* cassette. Next, we tested for *Tie2:Cre*-induced UPRT expression in a P6 heart. Control mice lacking the *Tie2:Cre* transgene showed no detectable UPRT expression in the heart (Fig. 2F), whereas *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic mice showed robust expression of UPRT in most or all PECAM1$^+$ heart endothelial cells (Fig. 2G). As expected, UPRT was also expressed in *Tie2:Cre*-expressing endocardial cells and heart valve interstitial cells of endocardial origin (Fig. 2G). In addition, we observed essentially 100% *Tie2:Cre*-induced UPRT expression in the E11.5 brain and heart (Supplemental Fig. S2).

We tested whether a second Cre line would also direct cell type-specific expression of UPRT using the cerebellar granule neuron precursor (GNP) driver *Math1:Cre* (Matei et al. 2005). Indeed, *Math1:Cre; CA>GFPstop>UPRT* double-transgenic mice showed robust expression of UPRT in GNPs of the P6 brain (Supplemental Fig. S3). We conclude that the *CA>GFPstop>UPRT* transgene provides highly penetrant Cre-inducible expression in response to multiple Cre lines, in multiple cell types, and at all tested stages of development. The homozygous *CA>GFPstop>UPRT* transgenic mouse was viable and fertile alone or in combination with *Tie2:Cre* or *Math1:Cre* transgenes.

TU tagging allows labeling and isolation of endothelial RNA from the postnatal brain

We wanted to know whether TU tagging was sensitive enough to isolate endothelial transcripts from the intact brain, where *Tie2:Cre*$^+$ endothelial cells represent ~5% of all cells (Daneman et al. 2010) and *Tie2:Cre* lineage-derived microglia/macrophages are even less abundant. Prior to the experiment, we selected 13 positive control genes from the literature that had validated widespread endothelial expression at embryonic and postnatal stages: *Cdh5* (*VEcad*), *Cd34*, *Egfl7*, *Emcn*, *Esam*, *Ets1*, *Flt1*, *Kdr* (*Flk1/VEGFR2*), *Nos3*, *Pecam1*, *Tek* (*Tie2*), *Tie1*, and *Thsd1*. We used these 13 positive control genes to evaluate the quality of our endothelial RNA purification. We injected 4TU subcutaneously into *Tie2:Cre; CA>GFPstop>UPRT* double transgenic P6 pups, waited 4 h, then purified total RNA from the intact brain. We used a subset of this total RNA for purification of TU-tagged RNA (presumptive nascent endothelial RNA) (Fig. 3A). We performed RNA-seq on the total RNA and the TU-tagged RNA from two independent biological replicates and observed very high reproducibility between replicates (Pearson correlation coefficients: 0.998 for total RNA replicates and 0.999 for TU-tagged RNA replicates).

For each experiment, we determined the number of reads that aligned to each annotated gene region, the number of reads that aligned to all annotated gene regions, and a "reads per million mapped reads" (RPM) value of each gene (Supplemental Table S1). We averaged the biological replicate RPM and plotted TU-tagged versus total RPM for each annotated gene (Fig. 3B). The vast majority of genes were expressed roughly equally in both endothelial and neural cell types, and their ~1:1 ratio led to a diagonal distribution in the plot (Fig. 3B). Strikingly, 11 of the 13 pan-endothelial control genes were located at the extreme left of the scatter plot, indicating enrichment in the TU-tagged RNA (Fig. 3B). The two positive control transcripts that were poorly enriched, *Tie1* and *Cd34*, may have a lower frequency of transcription initiation (see the Discussion). The 11 most enriched control genes showed an average enrichment of 3.90-fold (range: 2.83–5.64) (Table 1). This strongly suggests that TU tagging is sensitive enough to isolate endothelial transcripts from within the entire brain.

We used the 11 enriched positive control genes to define an enrichment cutoff; genes showing enrichment equal to or greater than these 11 positive control genes were termed "enriched" (n = 130 genes) (Supplemental Table S1) and analyzed further. We performed a gene ontology (GO) analysis of these enriched genes and discovered that they were significantly overrepresented for endothelial and hematopoietic terms (Fig. 3C). To determine whether annotated neuronal transcripts were underrepresented in the TU-tagged endothelial RNA, we performed GO analysis on the 500 genes with the lowest TU-tagged/total RPM ratio. These depleted transcripts were significantly overrepresented for neuronal terms, such as "neuron development" and "neuron projection morphogenesis" (Fig. 3C). These results provided evidence for the specificity of TU tagging: We started with total brain RNA in which 95% of the cells are neural, and yet the TU-tagged RNA was significantly enriched for annotated endothelial transcripts and depleted for annotated neural transcripts. We also observed depletion of transcripts encoding ribosomal proteins (Fig. 3C), most likely because these are stable RNAs that are infrequently transcribed. We conclude that TU tagging is effective at covalently labeling and purifying known endothelial RNAs from the intact brain.

If TU tagging works as intended, we should observe endothelial expression for many of the 130 genes whose transcripts were enriched equal to or greater than our 11 most enriched positive control transcripts. To determine the expression pattern of these genes, we used the E14.5 Eurexpress database, which has comprehensive coverage and high-resolution images (Diez-Roux et al. 2011). The availability of gene expression databases makes it possible to rapidly validate gene expression patterns; indeed, this is the purpose of such databases. Furthermore, the brain endothelial pattern is sufficiently distinctive as to make identification of endothelial-expressed genes unambiguous. We found data for 87 of the 130 genes in the Eurexpress database. Strikingly, 76% showed highly spe-

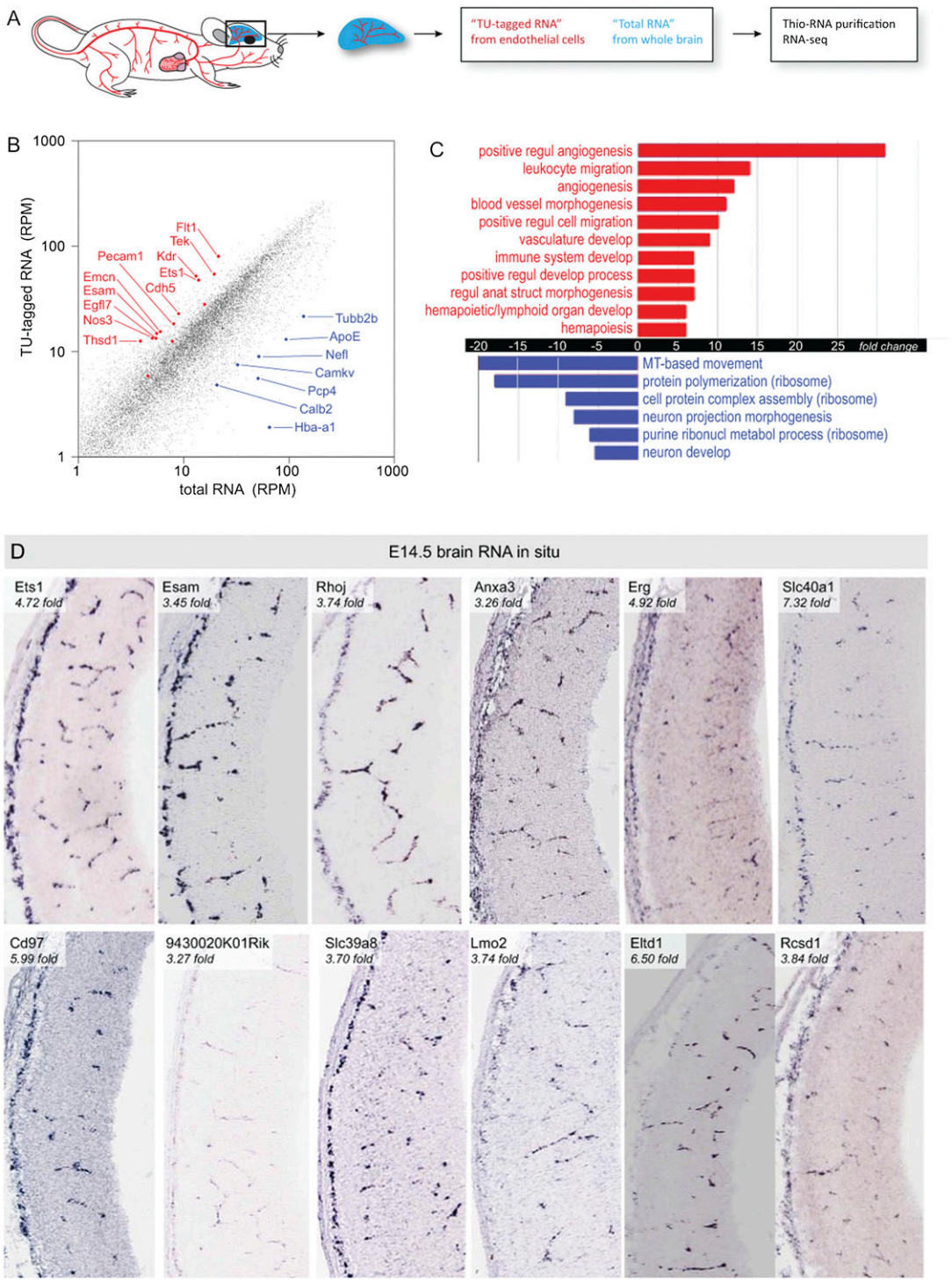


Figure 3. TU tagging of Tie2:Cre$^+$ endothelial cells within the P6 brain. (*A*) Schematic of the experiment. *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic P6 mice were given a 4TU injection subcutaneously and killed after 4 h, and the whole brain was removed. The brain total RNA was isolated ("total RNA"; blue) and a subset was used to purify thio-labeled presumptive endothelial RNA ("TU-tagged RNA"; red) for RNA-seq. (*B*) TU tagging of brain endothelial cells identified known endothelial and vascular genes. RNA-seq analysis of TU-tagged RNA versus total RNA; the average RPM from two biological replicates are shown for each gene. Note that 11 of the 13 pan-endothelial control genes have clear enrichment of their transcripts in the TU-tagged RNA based on their position at the *left* edge of the plot (*Flt1*, *Tek*, *Kdr*, *Ets1*, *Cdh5*, *Pecam1*, *Emcn*, *Esam*, *Egfl7*, *Nos3*, and *Thsd1*); two are not strongly enriched (*Cd34* and *Tie1*). Some of the most depleted transcripts are shown with blue diamonds and include the neuronal transcripts *Tubb2*, *ApoE*, *Nefl*, *Camkv*, *Pcp4*, and *Calb2* and the hemoglobin *Hba-a1* transcript. (*C*) TU tagging of brain endothelial cells identified known endothelial and vascular genes and depleted for known neuronal genes. GO analysis of the most up-regulated genes (red; genes enriched equal to or greater than our 11 most enriched positive control genes were used for GO analysis) (Supplemental Table S1) and the 500 most down-regulated genes (blue) (Supplemental Table S1). Redundant categories were excluded. (*D*) TU tagging of brain endothelial cells identified genes expressed in the *Tie2:Cre*-derived endothelial and microglial/macrophage cells. Gene expression data at E14.5 (Eurexpress, used with permission) are shown for some of the 130 most up-regulated genes from TU tagging in *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic whole-brain tissue. The *top left* pair are two of the pan-endothelial control genes (*Ets1* and *Esam*).

Table 1. *TU tagging of P6 brain endothelial/vasculature RNAs*

Gene symbol[a]	Fold-enriched[b]	Brain vasculature expression[c]	Brain neural expression[d]
C3ar1	8.31	MG/MP (broad)	—
Slc40a1	7.32	Endothelial (broad)	—
Vav1	6.65	MG/MP (subset)	—
Eltd1	6.50	Endothelial (broad)	—
Rasgrp3	6.34	—	Subset (GZ)
Cyyr1	6.25	Endothelial (subset)	—
Apold1	6.17	Endothelial (subset)	—
Fli1	6.16	Endothelial (broad)	—
Cx3cr1	6.10	MG/MP (broad)	—
Cd97	5.99	Endothelial (broad)	—
Kdr	5.64	Endothelial (broad)	—
Zfp366	5.55	Endothelial (subset)	—
Inpp5d	5.41	—	—
Ly75	5.23	—	—
Flt1	5.23	Endothelial (broad)	—
Ikzf1	5.17	—	Subset
Erg	4.92	Endothelial (broad)	Subset
Abcb1a	4.81	Endothelial (broad)**	—
Apbb1ip	4.78	MG/MP (broad)	Subset
Slfn5	4.74	Endothelial (subset)	—
Ets1	4.72	Endothelial (broad)	Subset
Ptprb	4.68	Endothelial (broad)	—
Atp13a5	4.64	MG/MP (broad)	—
Casp8	4.61	Endothelial (broad)	Subset (GZ)
Slco2b1	4.57	Endothelial (broad)	—
Slc2a1	4.52	Endothelial (broad)	Subset
Runx1	4.46	—	Subset
Ly86	4.44	MG/MP (broad)	—
Fzd6	4.41	Endothelial (subset)	—
Slco1c1	4.39	Endothelial (subset)	Subset
Thsd1	4.37	Endothelial (broad)	—
Itga4	4.34	Endothelial (broad)	Subset
Swap70	4.10	—	Broad
Eng	4.09	Endothelial (broad)*	—
Slco1a4	3.99	Endothelial (broad)**	—
Ptprc	3.98	MG/MP (broad)	—
2610019F03Rik	3.94	Endothelial (broad)	—
Rbpms	3.91	—	—
Msrb3	3.89	—	—
Rcsd1	3.84	Endothelial (broad)	Subset (GZ)
Ctsc	3.81	—	—
Slc16a1	3.80	—	Subset
Cd53	3.77	MG/MP (broad)	—
Wasf2	3.75	MG/MP (broad)	—
Lmo2	3.74	Endothelial (broad)	Subset
Rhoj	3.74	Endothelial (broad)	—
Emr1	3.72	MG/MP (broad)	—
Slc39a8	3.70	Endothelial (broad)	—
Tek	3.63	Endothelial (broad)	Subset
Gpr116	3.59	Endothelial (broad)**	—
Ect2	3.58	—	Subset (GZ)
Dab2	3.56	Endothelial (subset)	—
Lcp1	3.54	Endothelial (broad)	—
Osmr	3.51	Endothelial (subset)	—
Palmd	3.50	Endothelial (broad)	Subset (GZ)
Esam	3.45	Endothelial (broad)	—
Ugt1a7c	3.44	—	—
Nos3	3.42	Endothelial (broad)	—
Fzd4	3.42	Endothelial (subset)	Subset
Wwtr1	3.38	Endothelial (broad)	Subset (GZ)
Sgol2	3.33	—	Subset (GZ)
Fyb	3.33	MG/MP (broad)	—
Emcn	3.29	Endothelial (broad)	—
Ocln	3.29	Endothelial (broad)	Subset
9430020K01Rik	3.27	Endothelial (broad)	—
Anxa3	3.26	Endothelial (broad)	—
Pecam1	3.26	Endothelial (broad)	—
Podxl	3.25	Endothelial (subset)	Subset (GZ)
Heg1	3.24	Endothelial (broad)	Subset (GZ)
Slc22a8	3.19	Endothelial (subset)	—
Abcc4	3.11	—	—
Jag1	3.11	Endothelial (subset)	Subset (GZ)
Vrk2	3.11	—	Subset
Egfl7	3.08	Endothelial (broad)	—
Mospd1	2.99	—	—
Ebf1	2.97	—	Subset
Lcp2	2.91	MG/MP (broad)	—
Cobll1	2.91	Endothelial (broad)	—
Mrc1	2.89	—	—
Cd83	2.89	—	Subset
Rassf4	2.89	—	Subset
Fgd5	2.88	Endothelial (broad)	—
Ppfibp2	2.88	—	—
A130022J15Rik	2.87	Endothelial (broad)	—
Arhgap29	2.87	Endothelial (broad)	—
Neil3	2.87	—	—
Cast	2.85	Endothelial (broad)	—
Cdh5	2.84	Endothelial (broad)	—

[a]Genes encoding the 130 most enriched transcripts (Cdh5 and above) from P6 brains of *Tie2:Cre; CA:>stop>UPRT* double transgenics; genes not represented in the Eurexpress E14.5 expression database were excluded. (Bold) Positive control pan-endothelial genes. Note that Ugt1a7c is one of many isoforms sharing common 3′ exons, including Ugt1a9, Ugt1a10, and Ugt1a6b, which have similar enrichment and expression patterns (data not shown).

[b]TU-tagged RNA/total RNA RPM ratio; total and purified combined RPM <27 were excluded to avoid low RPM-biased ratios.

[c](MG/MP) Microglial/macrophage pattern (e.g., similar to the known microglial marker Cx3cr1); both of these cell types derive from Tie2:Cre$^+$ hematopoietic progenitors (Tang et al. 2010). (*) Endothelial only in Allen Brain Atlas developing mouse database; (**) endothelial only in Allen Brain Atlas adult mouse database.

[d]Neural expression from Eurexpress database. (GZ) Germinal zone.

cific expression in the *Tie2:Cre*-derived endothelial or microglial/macrophage cell types of the brain (66 of 87) (Table 1; Fig. 3D). Many of the genes were expressed in all brain endothelial cells (e.g., *Anxa3* and *Rhoj*) (Fig. 3D). Other genes were expressed in subsets of brain endothelial cells (e.g., *Slc22a8*) (Table 1; Supplemental Fig. S4). Importantly, only one of 87 (1%) was expressed broadly in neurons (Table 1), showing that widely expressed neural transcripts do not significantly contaminate the pool of TU-tagged endothelial transcripts. We conclude that TU

tagging can label and purify RNA from the relatively rare population of endothelial cells embedded within the intact brain.

Although unlikely, it is possible 4TU preferentially labels the RNA in endothelial cells, perhaps because of systemic 4TU transport in the blood stream. To test this hypothesis, we injected 4TU into P6 pups containing *Tie2:Cre* but lacking the *CA>GFPstop>UPRT* transgene. We purified TU-tagged RNA from the brain, performed RNA-seq exactly as described above, and performed GO analysis on the top 130 genes ranked by RPM value. The overrepresented GO terms all related to neurogenesis or neural functions (e.g., neuronal differentiation or glutamate receptor signaling) (Supplemental Fig. S4); in contrast, the top 130 most enriched genes from *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic P6 brains were overrepresented for endothelial and hematopoietic terms (Fig. 3C). This shows that there is little or no bias toward background incorporation of 4TU into brain endothelial cells and that our TU tagging of endothelial-specific transcripts is due to the specificity of the *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic genotype.

Having established that TU-tagged RNA from the *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic brains was enriched for endothelial genes (see above), we wanted to know whether this TU-tagged RNA was consistently depleted for neuronal-expressed genes. Such an exclusion of highly expressed neuronal transcripts from TU-tagged "endothelial" RNA would be a stringent validation of the method's specificity. We found that 69 of the 100 most depleted genes (Supplemental Table S1) were present in the Eurexpress database, and nearly 90% (61 of 69) were expressed in neuronal patterns. Many were highly expressed throughout the brain (e.g., *ApoE*, *Camkv*, and *Tubb2b*), while others were strongly expressed in distinct brain regions (e.g., *Calb2*, *Nefl*, and *Pcp4*) (Fig. 3B; Supplemental Fig. S4; Supplemental Table S1). One of the most depleted transcripts was hemoglobin (Fig. 3B), unsurprisingly, as this RNA is abundant in red blood cells of the brain (Supplemental Fig. S4) and thus is highly represented in the total bulk RNA population, but because red blood cells lack nuclei, it should not be present in the TU-tagged actively transcribed RNA population. We conclude that TU tagging is both highly sensitive (able to identify endothelial transcripts) and highly specific (little contamination from pan-neuronal transcripts).

TU tagging of postnatal heart endothelial RNA reveals tissue-specific differences

We next tested whether TU tagging could identify tissue-specific endothelial transcripts, as it has been proposed that each tissue may have a unique endothelial/vasculature transcriptome (Daneman et al. 2010). We used the same 4TU injection protocol as described above, then isolated total RNA from the whole heart of *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic P6 pups. We used a subset of the total RNA for purification of TU-tagged RNA and then performed RNA-seq on both total and TU-tagged RNA populations (Fig. 4A; Supplemental Table S2). Two biological replicates were performed for each experiment, and we observed a high correlation coefficient for total RNA replicates ($r = 0.997$) and TU-tagged RNA replicates ($r = 0.987$).

We averaged the biological replicate RPM and plotted TU-tagged versus total for each annotated gene (Fig. 4B). Similar to the brain endothelial TU tagging experiment, we observed nine of 13 pan-endothelial control genes at the extreme left of the scatter plot, indicating enrichment of known endothelial transcripts in the TU-tagged RNA (Fig. 4B). The nine most enriched control genes showed an average enrichment of 2.65-fold (range: 2.50–2.98). The fold enrichment of the positive control genes was expected to be less in the heart than the brain because *Tie2:Cre*-derived cells represent a much larger fraction of the total cells in the heart, and thus total RNA will contain relatively more endothelial transcripts. Even so, TU tagging was effective at purifying known endothelial transcripts from the intact heart. We also examined the most depleted transcripts and found them to be encoded by many well-known cardiac muscle genes (Fig. 4B, right-shifted blue genes). For example, muscle creatine kinase (*Ckm*) was depleted 24.9-fold, troponin C type 1 (*Tnnc1*) was depleted 15.6-fold, and myosin light chain transcripts (*Myl2*, *Myl3*, and *Myl4*) were depleted 11.5-fold to 16.6-fold (*Myl2* and *Myl3* transcripts are not shown in Fig. 4B because they had RPM values >1000) (Supplemental Table S2). As in the brain, hemoglobin transcripts (*Hba-a1*) were among the most depleted because they were not being actively transcribed (Fig. 4B). We conclude that TU tagging can purify heart *Tie2:Cre* lineage-derived transcripts with little contamination from highly expressed cardiac muscle transcripts.

As a control, we injected 4TU into P6 pups containing *Tie2:Cre* but lacking the *CA>GFPstop>UPRT* transgene. We purified TU-tagged RNA from the heart, performed RNA-seq exactly as described above, and performed GO analysis on the top 130 genes ranked by RPM value. The overrepresented GO terms all related to cardiac biology or muscle development/function (e.g., muscle cell differentiation) (Supplemental Fig. S5). This shows that there is little or no bias toward background incorporation of 4TU into heart endothelial cells and that our TU tagging of heart endothelial transcripts is due to the specificity of the *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic genotype.

To determine how many of the most enriched transcripts are actually expressed in heart endothelial cells, we assayed the 119 genes enriched greater than the nine most enriched positive control genes for endothelial expression using the Eurexpress database (Table 2; Fig. 4C). Of the 81 genes represented in the Eurexpress database, 72% showed expression in *Tie2:Cre*$^+$ endothelial/endocardial cells or *Tie2:Cre* lineage-derived atrioventricular canal cushion mesenchymal cells (58 of 81) (Table 2). Some transcripts were expressed in a broad endothelial pattern, including both endocardium and coronary endothelium, similar to known pan-endothelial genes (e.g., *Gimap6*) (Fig. 4C). Other transcripts were expressed only in endocardial cells (e.g., *Lgals9* and *Ednrb*) (Fig. 4C), coronary endothelium (e.g., *Clec1a*) (Fig. 4C), or cushion

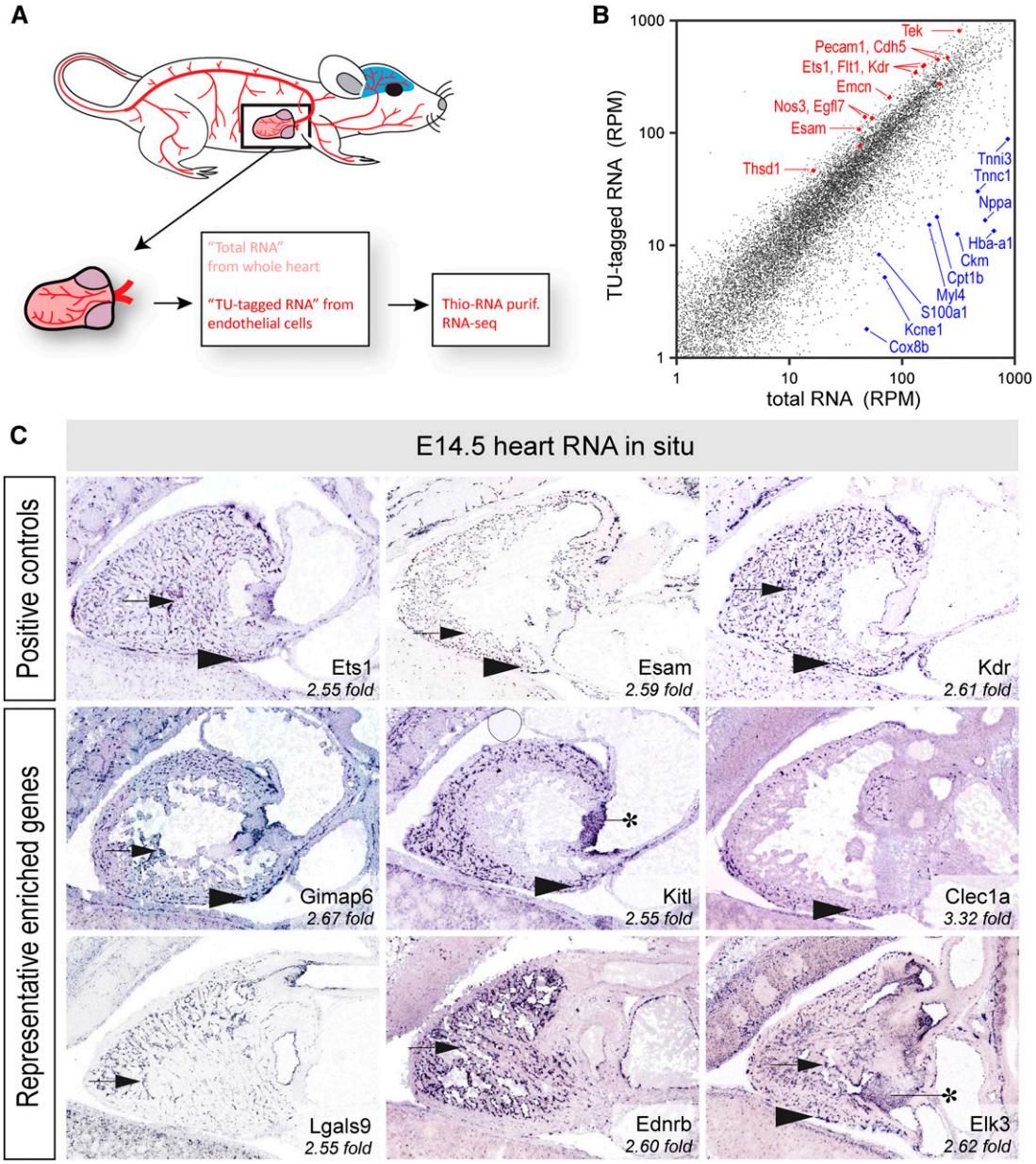


Figure 4. TU tagging of *Tie2:Cre*$^+$ endothelial cells within the P6 heart. (*A*) Schematic of the experiment (see Fig. 3 for explanation). (*B*) TU tagging of heart endothelial cells identified known endothelial-expressed genes. RNA-seq analysis of TU-tagged RNA versus total RNA; average RPM from two biological replicates are shown for each gene. Positive control genes are shown with red diamonds; note that nine of the 13 pan-endothelial control genes have clear enrichment in the TU-tagged RNA based on their position at the *left* edge of the plot (*Flt1*, *Tek*, *Kdr*, *Ets1*, *Emcn*, *Esam*, *Egfl7*, *Nos3*, *and Thsd1*), two others are almost as enriched (*Pecam1* and *Cdh5*), and two are weakly enriched (*Cd34* and *Tie1*). Some of the most depleted transcripts are shown with blue diamonds and include the cardiac muscle transcripts *Tnni3*, *Tnnc1*, *Ckm*, *Cpt1b*, *Myl4*, *S100a1*, and *Kcne1* and the hemoglobin *Hba-a1* transcript. (*C*) TU tagging of heart *Tie2:Cre* lineage-derived cells identified genes expressed in endothelium, endocardium, and valve (endocardial cushion) mesenchyme. Gene expression data at E14.5 (Eurexpress, used with permission) are shown for some of the 119 most up-regulated genes from TU tagging in *Tie2:Cre; CA>GFPstop>UPRT* double-transgenic whole heart tissue. The *top* row shows pan-endothelial control transcripts (*Ets1* and *Esam*). (Arrow) Ventricle endocardial cells; (arrowhead) coronary endothelial cells; (black asterisk) *Tie2:Cre*-derived atrioventricular canal cushion mesenchyme.

mesenchymal cells (e.g., *Serpini1*) (Fig. 4C). The expression patterns of additional known and novel endothelial/endocardial transcripts identified by TU tagging are shown in Supplemental Figure S5. Interestingly, 19 genes showed tissue-specific expression in endothelial cells of the heart but not the brain (Table 2, right column). Most of the enriched genes showed minimal or no expression in non-endothelial lineage cells of the heart, further highlighting

Table 2. *TU tagging of P6 heart endothelial/vasculature RNAs*

Gene symbol[a]	Fold-enriched[b]	Heart expression[c]	Brain endothelia[d]
She	3.82	Endothelial$^{EC/CE}$	+
Gbp6	3.48	—	—
Clec1a	3.32	Endothelial^{CE}	+
Inpp5d	3.22	CM	+
Nova2	3.21	Endothelial^{CE*}	—
Lin7a	3.17	—	—
Ppp1r16b	3.11	Endothelial$^{CE/EC\star}$	—
Prkcc	3.06	CM	—
Sox7	3.04	Endothelial^{CE}	+
Slc28a2	3.03	—	—
Sh2d3c	3.01	Endothelial$^{EC/CE}$, CM	—
9930013L23Rik	2.98	—	—
Nos3	2.98	Endothelial$^{EC/CE}$	+
3110035E14Rik	2.94	—	—
Nfatc1	2.93	Endothelial^{EC*}	—
St8sia4	2.86	Endothelial^{CE*}	—
Bmx	2.84	Endothelial^{CE*}	—
Thsd1	2.84	Endothelial$^{EC/CE}$	+
Apold1	2.82	Endothelial^{CE}	+
F11r	2.81	Endothelial$^{EC/CE}$	(+)
Eltd1	2.80	Endothelial$^{EC/CE}$	+
Slfn5	2.80	Endothelial^{EC}	+
Cyyr1	2.80	Endothelial^{CE}	+
Stil	2.80	—	—
Fli1	2.79	Endothelial^{CE}	+
Sipa1	2.78	Endothelial^{CE*}	+
Sncaip	2.78	Endothelial^{EC*}	—
Ptpre	2.76	Endothelial^{EC*}	+
Rasgrp3	2.75	—	—
Trp53i11	2.75	—	—
Mcf2l	2.74	—	—
Serpini1	2.74	Endothelial^{CE*}, CM	—
Nrp2	2.74	—	—
Shroom2	2.73	—	—
F13a1	2.70	—	(+)
Ablim3	2.69	Endothelial^{EC}, CM	—
Ptprb	2.68	Endothelial^{CE}, CM	+
Meox2	2.68	Endothelial^{CE}	—
Myct1	2.68	—	—
Emcn	2.68	Endothelial$^{EC/CE}$	+
Pparg	2.67	Endothelial$^{EC/CE}$, CM	+
Gimap6	2.67	Endothelial$^{EC/CE}$	+
Plcb1	2.67	—	—
Ggta1	2.67	Endothelial$^{EC/CE}$, CM	—
Sema6a	2.67	Endothelial^{EC*}	—
Kcnc2	2.66	—	—
Erg	2.65	Endothelial$^{EC/CE}$	+
Cd200	2.64	Endothelial^{EC*}	—
Dcbld1	2.63	Endothelial^{EC}, CM	—
Cit	2.63	—	—
Lims2	2.63	—	(+)
Elk3	2.62	Endothelial$^{EC/CE}$, CM	+
Kdr	2.61	Endothelial$^{EC/CE}$	+
Cdc42ep4	2.60	CM	+
2610021K21Rik	2.60	—	—
Ednrb	2.60	Endothelial^{EC*}	—
Chrm3	2.60	—	—
Esam	2.59	Endothelial$^{EC/CE}$	+
St6galnac3	2.59	—	—
Palm	2.57	Endothelial^{EC}, CM	—
Notch1	2.57	Endothelial$^{EC/CE}$, CM	+
Podxl	2.57	Endothelial$^{EC/CE\star}$	+
Ccdc85a	2.57	—	—
Cd38	2.56	Endothelial$^{EC/CE}$	—
Arhgef15	2.56	Endothelial$^{EC/CE}$	+
Flt1	2.56	Endothelial$^{EC/CE}$	+
Casp8	2.55	Endothelial^{CE}, CM	+
Kitl	2.55	Endothelial$^{EC\ CE}$, CM	+
Myo1d	2.55	—	—
Ets1	2.55	Endothelial$^{EC/CE}$	+
Lgals9	2.55	Endothelial^{EC}	—
Dab2	2.55	Endothelial^{EC*}	+
Coro2b	2.55	—	—
Prkd2	2.54	Endothelial$^{EC/CE}$, CM	+
Dach1	2.54	Endothelial^{CE*}	—
Fgd5	2.54	Endothelial$^{EC/CE}$, CM	+
1190002H23Rik	2.53	Endothelial^{CE}	+
9430020K01Rik	2.53	Endothelial^{CE}	+
Clec14a	2.53	Endothelial^{EC}	(+)
Tek	2.52	Endothelial$^{EC/CE}$	+
Egfl7	2.50	Endothelial$^{EC/CE}$	+

[a]Genes encoding the 119 most enriched transcripts (positive control Egfl7 and above) from P6 hearts of *Tie2:Cre; CA:>stop>UPRT* double transgenics (TU-tagged/total RNA); genes not represented in the Eurexpress database were excluded. (Bold) Positive control endothelial genes.

[b]TU-tagged RPM/total RPM ratio; total RPM <15 and total + purified combined RPM <62 were excluded to avoid low RPM-biased ratios.

[c](EC) Endocardial cells; (CE) coronary endothelial cells; (CM) *Tie2:Cre* lineage-positive cushion mesenchyme; (*) expression in a subset of the indicated cell type.

[d]Neural expression from Eurexpress database; (+) expressed in brain endothelia.

the specificity of the method. Conversely, genes with robust expression in nonendothelial cells of the heart were not enriched, showing little contamination from nonendothelial transcripts. For example, *cardiac troponin T2* (*Tnnt2*), *cardiac troponin I3* (*Tnni3*), and *myosin light polypeptide 4* (*Myl4*) were highly expressed in nonendothelial cells of the heart (Supplemental Fig. S5) but were depleted in our TU tagging experiment (Supplemental Table S2). This highlights the cell type specificity of the method. We conclude that TU tagging can be used to purify cell type-specific transcripts from multiple tissues as well as to identify novel tissue-specific endothelial transcripts.

TU tagging of embryonic brain endothelial cells

Developmental studies would particularly benefit from TU tagging given the vast number of tightly coordinated processes involving gene expression changes that must occur in small cell populations and during narrow time windows. Using TU tagging to label RNA within the developing embryo requires injecting 4TU into

the pregnant female at sufficient levels for 4TU to pass the placental barrier and become incorporated into embryonic RNA. To determine whether this could be done, we injected 4TU into females carrying E15.5 embryos 6 h prior to harvesting embryos. We then isolated total RNA from the brains of *Tie2:Cre;* CA>*GFPstop*>*UPRT* double-transgenic E15.5 embryos (Fig. 5A). We used a subset of this total RNA for purification of TU-tagged endothelial RNA. We performed RNA-seq on both total and TU-tagged RNA (Supplemental Table S3). Two biological replicates were performed, and we observed a high correlation coefficient for total RNA replicates (r = 0.998) and for TU-tagged RNA replicates (r = 0.994).

To determine the most enriched genes, we averaged the two total RNA replicates and the two TU-tagged RNA replicates and then, for each gene, we calculated the ratio of TU-tagged/total RNA RPM (Fig. 5B; Supplemental Table S3). Six of the 13 positive control genes were clearly clustered among the most enriched transcripts (top red oval in Fig. 5B), two other control genes were adjacent but less enriched, and the remaining five positive control genes were also less enriched and clustered in a different area of the plot (bottom red circle in Fig. 5B). To determine the expression patterns of the unknown enriched genes, we searched the Eurexpress E14.5 database for the expression patterns of the 289 genes enriched equal to or greater than the six most enriched positive control genes (red oval in Fig. 5B). We found expression data for 92 genes and found 30 (33%) genes expressed in *Tie2:Cre*-derived endothelial cells or microglia/macrophages (Supplemental Table S3).

The relatively low percentage of endothelial genes that we observed in the embryonic brain experiments led us to try a different approach. We reasoned that if we purified TU-tagged RNA from UPRT-positive brains and UPRT-negative brains that were both exposed to 4TU, the background 4TU incorporation would be equal across all brains, and thus transcripts uniquely detected in the UPRT-positive cell types would stand out. To do this experiment, we compared TU-tagged purified RNA from UPRT-positive E15.5 brains (*Tie2:Cre;* CA>*GFPstop*>*UPRT*) and TU-tagged purified RNA from UPRT-negative E15.5 brains (CA>*GFPstop*>*UPRT*). We performed replicates for each condition (correlation coefficients: TU-tagged UPRT$^+$, r = 0.994; TU-tagged UPRT$^-$, r = 0.821); a lower correlation coefficient is expected for the UPRT$^-$ experiments, where we are purifying very low levels of background-labeled transcripts. We identified 39 genes that were enriched equal to or greater than the positive control gene *Pecam1* (4.06-fold enriched), and 30 of these genes were present in the Eurexpress database. A full 60% (18 of 30) of these enriched transcripts were expressed in endothelial cells or *Tie2:Cre* lineage-derived macrophages/microglia (Table 3). This is almost double the percentage obtained by comparing TU-tagged RNA and total RNA. Thus, we found that comparing TU-tagged RNA from UPRT$^+$ and UPRT$^-$ tissues is an effective method for identifying transcripts specific to the UPRT$^+$ cell type. In some cases, this approach may be better than comparing TU-tagged RNA and total RNA from the same sample (see the Discussion). We conclude that TU tagging can be used to label and purify cell type-specific RNA from within the developing embryo.

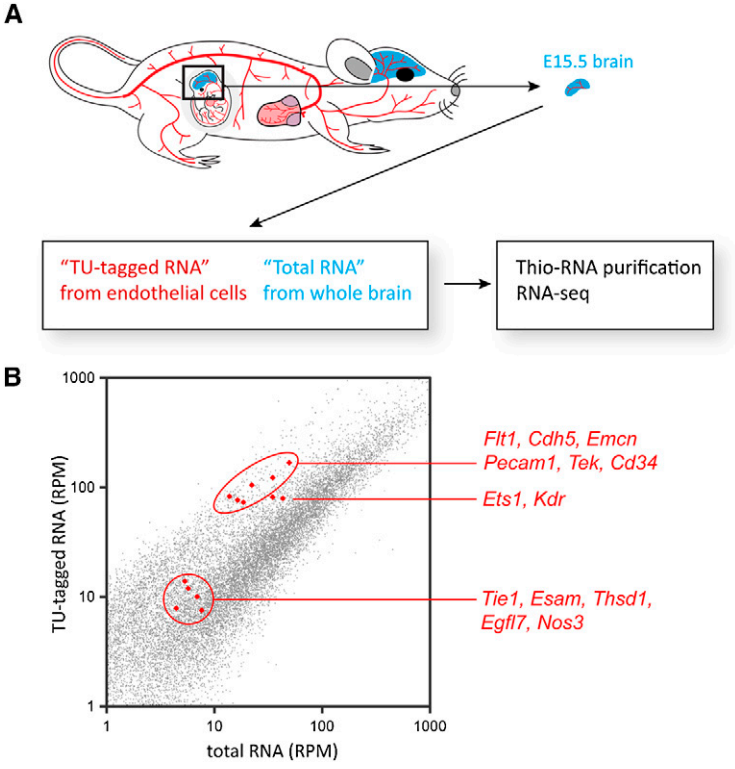


Figure 5. TU tagging of *Tie2:Cre*$^+$ embryonic brain endothelial cells. (*A*) Schematic of the experiment (see Fig. 3 for explanation). (*B*) TU tagging of E15.5 brain endothelial cells identified known endothelial and vascular genes. RNA-seq analysis of TU-tagged RNA versus total RNA; an average of two replicates for each gene. Six of the 13 pan-endothelial control genes cluster at the *left* edge of the plot (red oval: *Flt1*, *Tek*, *Cdh5*, *Pecam1*, *Emcn*, and *Cd34*), two are adjacent but less enriched (*Kdr* and *Ets1*), and five cluster closer to the center of the plot (red circle: *Thsd1*, *Esam*, *Nos3*, *Tie1*, and *Egfl7*).

Table 3. *TU tagging of E15.5 brain endothelial/vasculature RNAs*

Gene symbol[a]	Fold-enriched (exptl[b]/neg[c])	Brain vasculature expression[d]
Sepp1	11.65	Endothelial (broad)
Nt5c2	7.39	—
Slc40a1	7.12	Endothelial (broad)
Stab1	6.59	Endothelial (broad)
Cav1	6.56	Endothelial (broad)
Tpm4	6.06	Endothelial (broad)
Mest	5.86	Endothelial (subset)
Tspan6	5.30	—
Gpr116	5.20	Endothelial (broad)
Cd164	5.14	Endothelial (broad)
Kctd12	4.96	—
Ccng1	4.94	—
F2r	4.91	—
Cntnap3	4.70	—
Cdh5	4.60	Endothelial (broad)
Ap1s2	4.48	—
Rgs5	4.40	Endothelial (broad)
Ctsb	4.37	MG/MP
Ets1	4.28	Endothelial (broad)
Mrc1	4.26	MG/MP
Cd34	4.26	Endothelial (broad)
Lamp1	4.25	—
Tmsb4x	4.23	Endothelial (subset)
Snx5	4.20	Endothelial (broad)
Hmgb1	4.18	—
Lamp2	4.13	—
Prkrir	4.12	—
Gas6	4.10	Endothelial (broad)
Rbm3	4.09	—
Pecam1	4.04	Endothelial (broad)

[a]Genes encoding the 39 most enriched transcripts (positive control Pecam1 and above) from *Tie2:Cre; CA>GFPstop>UPRT* E15.5 brains; genes not represented in the Eurexpress database were excluded. (Bold) Positive control endothelial genes.
[b]RPM of TU-tagged RNA purified from E15.5 experimental double-transgenic brains (*Tie2:Cre; CA>GFPstop>UPRT*).
[c]RPM of TU-tagged RNA purified from E15.5 negative control Cre-minus single-transgenic brains (*CA>GFPstop>UPRT*). Genes with RPM <10 were excluded to avoid low RPM-biased ratios. Several female-specific transcripts were highly enriched (e.g., Tsix; ratio of 111.97), probably due to fact that the double-transgenic embryos were female and the negative control embryos were male; these genes have been excluded.
[d]Expression patterns from Eurexpress database, with enriched genes not represented excluded. (MG/MP) Microglial/macrophage pattern (e.g., similar to the known microglial marker Cx3cr1).

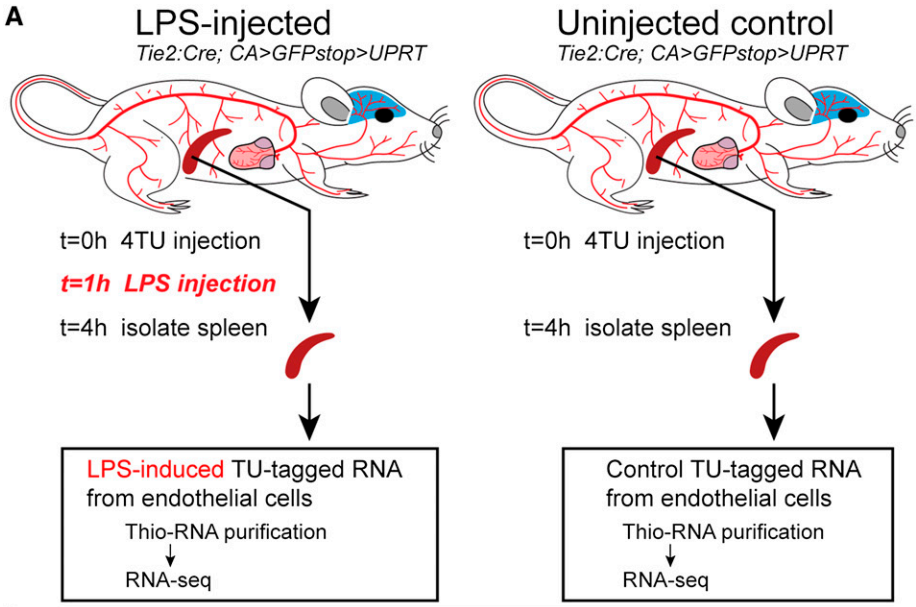
TU tagging to detect temporal changes in gene expression

TU tagging has the potential to identify rapid changes in gene expression due to its ability to purify nascent transcripts, thus improving detection of early response genes by ignoring previously transcribed RNAs. To test the ability of TU tagging to identify newly induced transcripts, we assayed the spleen endothelial gene expression response to lipopolysaccharide (LPS) injection. This assay is ideal because several innate immune response genes are known to be induced by LPS injection (Buttini and Boddeke 1995; Qin et al. 2007; Benicky et al. 2009), and microarray data exist for spleen transcriptomes following LPS injection (Hammer et al. 2006). We injected 4TU into *Tie2:Cre; CA>GFPstop>UPRT* adult mice and subsequently injected half of these mice with LPS after 1 h; 3 h later, we collected intact spleen tissue from both LPS-injected and uninjected mice (Fig. 6A). TU-tagged RNA was isolated, and RNA-seq was performed. We determined the RPM ratio for TU-tagged RNA from LPS-injected spleens to that of TU-tagged RNA from uninjected spleens (Supplemental Table S4) and performed GO analysis on the most significantly enriched transcripts (LPS-induced/uninduced) (see the Materials and Methods for cutoff criteria). We found that all of the most overrepresented categories were related to the immune response, such as "defense response" ($P = 2.9 \times 10^{-24}$), "innate immune response" ($P = 4.1 \times 10^{-24}$), and "response to cytokine stimulation" ($P = 1.9 \times 10^{-9}$) (Fig. 6B). In addition, we compared our most enriched transcripts with those previously identified as LPS-induced spleen transcripts by microarray analysis (Hammer et al. 2006). We found 24 transcripts in common, and overall, our TU tagging method provided greater fold enrichment than that observed by microarray analysis (Fig. 6C). Thus, TU tagging can identify rapidly induced transcripts following systemic LPS injection, and many of these transcripts are annotated as immune response genes. We conclude that TU tagging is an effective method for identifying temporal changes in gene expression within specific cell types in intact tissue.

TU tagging in chimeric mice following bone marrow transplants

TU tagging has already been shown to work in transplantation experiments in which UPRT$^+$ *Toxoplasma gondii* protozoans are injected into UPRT$^-$ murine hosts (Cleary et al. 2005). Here, we extend these experiments using UPRT transgenic mice. We transplanted bone marrow from *CA:UPRT* mice into sublethally irradiated UPRT$^-$ hosts (Fig. 7A). We analyzed gene expression in the spleen, where the UPRT$^-$ mesenchymal, vascular endothelial, and smooth muscle cells surround UPRT$^+$ hematopoietic cells. In these chimeras, UPRT$^+$ leukocytes expressed the CD45.2 allele, while host leukocytes expressed the CD45.1 allele, and all other cells in the spleen would be CD45$^-$ niche cells. We performed bone marrow transplantations with varying amounts of CD45.2$^+$ cells, and subsequent quantification of CD45.2, CD45.1, and CD45$^-$ populations by flow cytometry allowed calculation of UPRT$^+$ cell percentages in the spleen (Fig. 7B). Flow cytometry also revealed that niche cells consistently make up ~30% of the spleen in both control and chimeric mice. We prepared TU-tagged RNA from the intact spleen and found that RNA yields correlated with the percentage of UPRT$^+$ cells in the spleen, approaching negative control yields only when <1% of the spleen cells were UPRT$^+$ (Fig. 7C). This suggests that TU tagging has the potential to isolate cell type-specific RNA from 1% or less of the total cell number in a tissue.

We purified TU-tagged RNA from chimeric spleens and from *CA:UPRT* spleens and performed RNA-seq (Supple-



B

GO terms of transcripts enriched after LPS-injection

Term	p-value	Selected genes in category
Defense response	2.9e-24	*Nfkbiz, Mx1, Cxcl10, Tnip3, Il2ra, Dhx58, Irgm1*
Innate immune response	4.1e-24	*Ifih1, Irf1, Tnip3, Nfkbia, Ccl5, Mx2, Oasl2*
Immune system process	3.5e-19	*Cxcl11, Gbp5, Herc6, Akirin2, Prkcq, Isg20, Myc*
Immune effector process	1.3e-16	*Oas2, Isg20, Oas1b, Pml, Cxcl10, H2-T23*
Response to cytokine stim.	1.9e-09	*Irgm2, Il2ra, Gbp5, Serpina3g, Pyhin1, Il1rn*
Response to interferon-beta	1.7e-07	*Ifit1, Ifit3, Pyhin1, Iigp1, Plscr1*
Toll-like receptor signaling	2.5e-05	*Cd86, Tnip3, Tnfaip3, Nfkbia, Rsad2, Irf1*
Response to interferon-alpha	1.5e-04	*Ifit1, Ifit3, Ifit2, Plscr1*
Leukocyte activation	3.7e-03	*Cd86, Prkcq, Il2ra, Ccl5, Il15ra, Bcl3, Treml2*

C

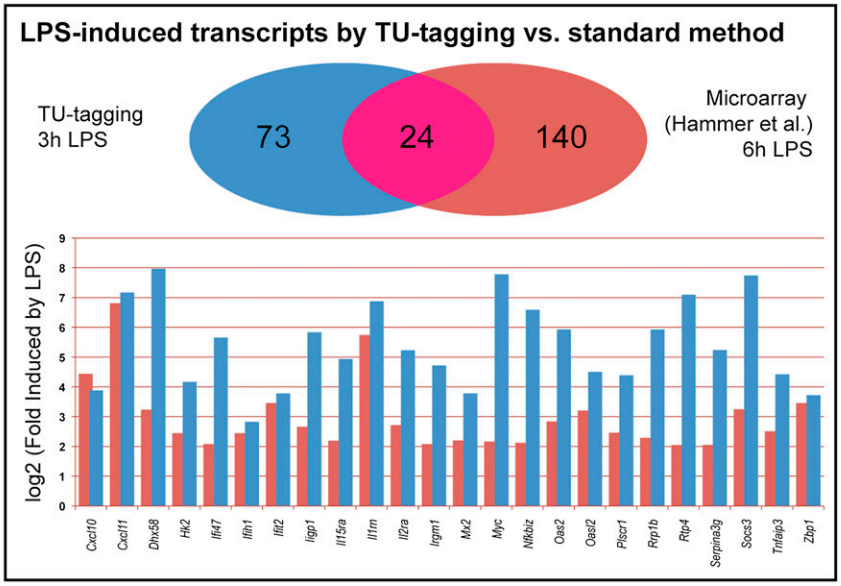


Figure 6. TU tagging of LPS-induced transcripts in the *Tie2:Cre*$^+$ adult spleen. (*A*) Schematic of the experiment. TU-tagged RNA was purified from the intact spleen of LPS-induced mice (*left*) or uninduced control mice (*right*). (*B*) GO analysis of overrepresented categories found in the TU-tagged spleen transcripts determined to be significantly up-regulated upon LPS treatment. Selected GO terms shared by those genes were ranked by significance. A subset of LPS-induced spleen genes in each category is noted. (*C*) TU tagging of LPS-induced adult spleen transcripts identified known LPS-induced genes. The Venn diagram compares the set of significantly up-regulated TU-tagged transcripts following a 3-h LPS treatment with those identified in a published microarray study of the mouse spleen transcriptome response to a 6-h LPS exposure (fourfold up and higher). See the text for details. The bar graph shows the fold increase (log2 scale) of the 24 transcripts shared in common between the two data sets in their respective LPS induction studies. The colors match the experiments shown in the Venn diagram.

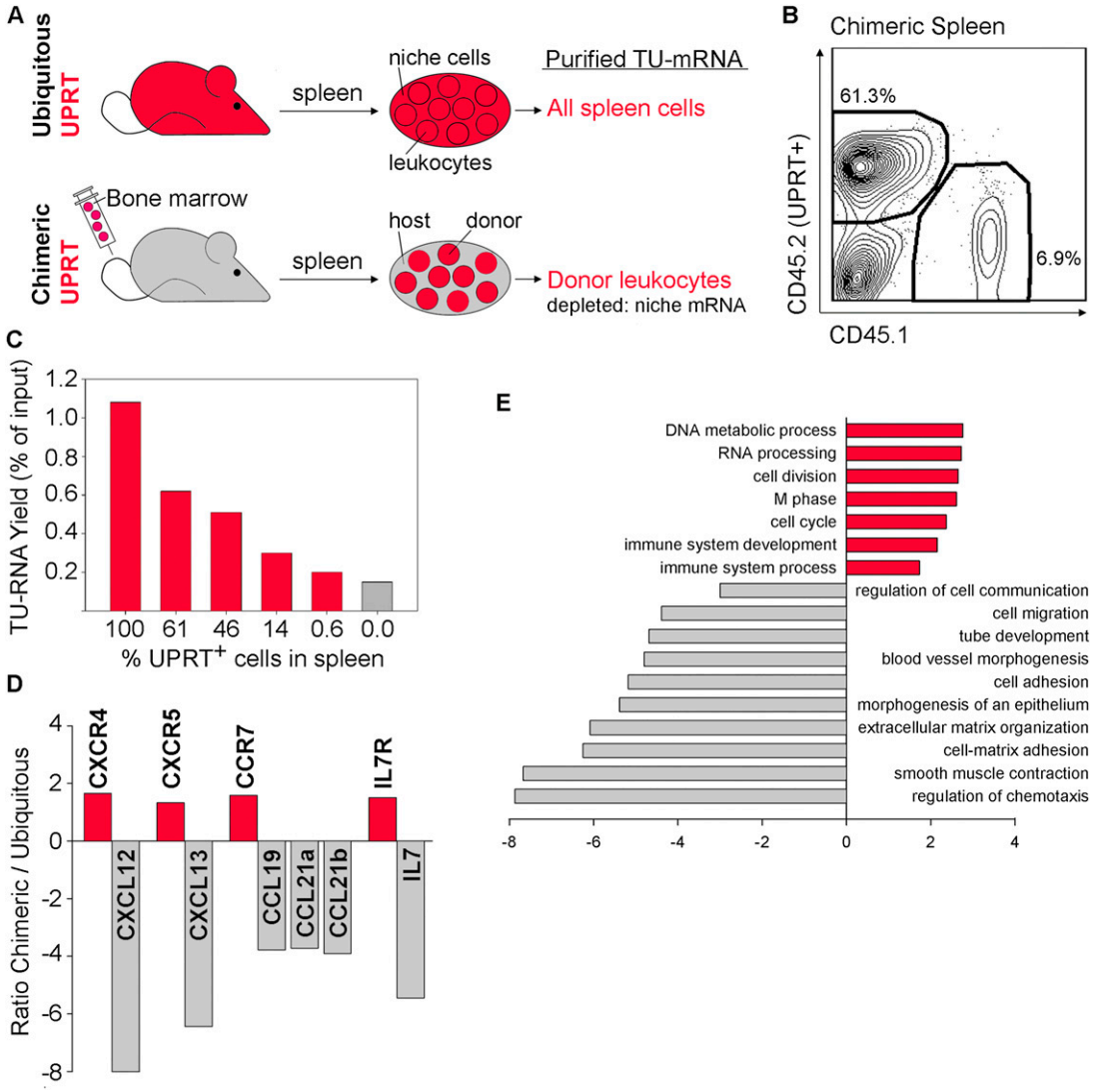


Figure 7. TU tagging of transplanted bone marrow cells in an unlabeled host. (*A*) Schematic of experiments comparing gene expression in whole spleen (ubiquitous UPRT) to gene expression in spleen leukocytes (chimeric UPRT). (Red) UPRT$^+$ cells. (*B*) FACS quantification of UPRT$^+$ leukocytes in the spleen of a chimeric mouse. Bone marrow donor mice express the CD45.2 allele on leukocytes, while the host leukocytes express the CD45.1 allele. Percentages are based on total cell counts in the spleen (excluding erythrocytes). The remaining 31.8% of cells in this spleen are the CD45-negative niche cells. These data produced the bar labeled "61" in *C*. (*C*) TU-tagged RNA yields as a function of UPRT$^+$ cell frequency. FACS analysis was used to quantify UPRT$^+$ CD45.2 cells following injection of decreasing amounts of UPRT$^+$ donor bone marrow in multiple transplantation experiments. Representative yields are shown. One-hundred percent of UPRT is from a ubiquitous UPRT mouse (no bone marrow transplant) and 0.0% of UPRT is from a wild-type mouse (no bone marrow transplant). (*D*) Relative abundance of transcripts encoding leukocyte-specific cell surface receptors and the cognate niche-specific ligands. Receptor–ligand pairs are adjacent. The CCR7 receptor has three ligands: CCL19, CCL21a, and CCL21b. The chimeric/ubiquitous ratio is the RPM value for chimeric spleen TU-tagged RNA divided by RPM value for ubiquitous spleen RNA (negative values are inverse of the ratio). (*E*) GO categories of genes that are enriched in the chimeric spleen/ubiquitous spleen >1.33-fold (red bars) and genes that are depleted in the chimeric spleen/ubiquitous spleen less than twofold (gray bars). Representative enriched categories with statistically significant differences ($P < 0.0001$) are shown. Redundant categories are excluded.

mental Table S5). We compared transcript abundance of mRNAs encoding known leukocyte-specific cell surface receptors and mRNAs encoding their niche-specific ligands. As expected, receptor transcripts were enriched in the chimeric spleen, and the ligand transcripts were depleted from the chimeric spleen (Fig. 7D). GO analysis of TU-tagged enriched transcripts from the chimeric spleen revealed a significant overrepresentation of genes in categories associated with immune system development and cell division (Fig. 7E). The enrichment of genes in cell cycle-related categories correlates with the high degree of leukocyte proliferation that occurs in the spleen. GO analysis of transcripts that were depleted in the TU-tagged RNA population revealed a significant overrepresentation of genes in categories associated with the niche cells, such as cell adhesion, chemotaxis, and

smooth muscle contraction (Fig. 7E). We conclude that transplantation of cells from a *CA:UPRT* donor into a UPRT$^-$ recipient allows identification of both donor-specific transcripts (enriched) and host-specific transcripts (depleted) from the mosaic tissues.

Discussion

Here, we describe the mouse TU tagging system, a chemical/genetic intersectional method for covalently labeling and purifying actively transcribed (nascent) RNA from specific cell types within intact mouse tissues. The isolation of endothelial RNA from both the brain and the heart has revealed several interesting findings. First, we identified a number of tissue-restricted endothelial genes that are expressed in the heart but not the brain (Table 2). Second, we failed to identify any novel transcripts with pan-endothelial expression similar to our 13 endothelial control transcripts. This suggests that the core set of transcripts common to all endothelial cells may be relatively limited, although it is also possible that additional core endothelial transcripts are not readily identified because they are long-lived transcripts or transcripts that are infrequently initiated. Third, we observed substantial regional heterogeneity in endothelial gene expression within an organ. For example, we identified transcripts expressed in coronary endothelial cells but not in endocardial cells at E14.5 (e.g., *Fli1*, *Meox2*, *Sox7*, and *Bmx*, all known endothelial transcripts) (Melet et al. 1996; Ekman et al. 1997; Takash et al. 2001; Patel et al. 2005). Regional expression within brain endothelial cells was also observed (e.g., *Slc22a8* is low in E14.5 forebrain endothelia but high in more caudal brain endothelial cells). Our results highlight the importance of gene expression differences in endothelial heterogeneity and provide a powerful resource for understanding their developmental genetic origins.

Another useful application of TU tagging is the generation of chimeric UPRT-expressing mice. Here, we transplanted *CA:UPRT* bone marrow cells to generate chimeric mice in which donor leukocytes are UPRT$^+$. The design of these experiments resulted in sensitive detection of niche-specific mRNAs because chimeric spleens completely lack UPRT$^+$ niche cells (thus, niche RNAs are underrepresented in the TU-tagged RNA). Several of these niche-specific genes have not been shown previously to function in immune development within the spleen. Of particular interest are genes that encode regulators of cell migration and morphogenesis that were originally identified for their role in neural development, such as *Sema3d*, *Sema3a*, and *Robo2*. The role of these genes or related family members in other tissues allows some prediction of their function in the spleen. For example, Robo4 is necessary for hematopoietic stem cell localization to bone marrow niches (Smith-Berdan et al. 2011). Sema3a is expressed in lymphatic endothelial cells and regulates dendritic cell migration into lymph nodes (Takamatsu et al. 2010). We suggest that Robo2 and Sema3a/d may play similar roles in the spleen.

What are the key variables that must be optimized for TU tagging to be successful? First, it is essential to keep background 4TU incorporation as low as possible in UPRT$^-$ cells. The amount of 4TU incorporation can be assayed following the biotinylation step using streptavidin-HRP Northern blots (see Miller et al. 2009). While this approach can detect TU-tagged RNA enrichment over background in heart experiments, it is not sensitive enough to detect enrichment over background in our brain experiments (data not shown) despite our success in validating 76% of the enriched transcripts from the brain as being expressed in endothelial cells. We are currently developing more sensitive methods for quantifying thio-labeled RNA. Second, it is important to minimize purification of background-labeled RNA from UPRT-negative cell types. This can be accomplished by the fragmentation step of the protocol, which lowers background by reducing the fraction of each weakly labeled transcript that is purified (see Fig. 1D; Supplemental Fig. S6). Third, there are different TU tagging experimental designs, and each should be tested to see which is best for a particular experiment. The design we used for most of the experiments in this study was to compare TU-tagged purified RNA from UPRT$^+$ tissues to total RNA from the same sample (Figs. 2–4). This method should work well for tissues with a low percentage of UPRT$^+$ cells. An alternative method is to compare TU-tagged RNA from UPRT$^+$ and UPRT-negative tissues that were both exposed to 4TU. This method may work well when there are a relatively high percentage of UPRT$^+$ cells in the tissue, and it was a superior method for our embryonic tissue experiments.

TU tagging is complementary to cell dissociation methods such as FACS, panning, or laser capture. These methods can be used to isolate the entire RNA population of a cell, both nascent and mature RNAs, whereas TU tagging only isolates nascent RNA. Thus, each approach may be useful for different applications. However, TU tagging has several practical and theoretical advantages over dissociation-based methods. TU tagging works on dispersed cell populations and cell types with fine cellular processes (e.g., neurons and glia) that are impossible to purify by laser capture or manual dissection. In addition, it is well known that loss of epithelial junctions can induce gene expression changes (Balda and Matter 2009; Stepniak et al. 2009); thus, dissociation methods may result in biased transcriptional profiles. TU tagging avoids these limitations by covalently labeling RNA within intact tissue in the living organism, allowing gene expression profiles to be determined in the native cell environment.

It is informative to compare our data with those of Daneman et al. (2010), where FACS was used to purify *Tie2:GFP*$^+$/PDGFRβ$^-$ cells from the P7 brain, although the experiments are done on different ages of mice (P6 vs. P7). Only a few genes were identified by both methods (present in Table 1 of Daneman et al. 2010 and this study): *Flt1*, *Ets1*, *Eng*, *Abcb1a*, *Slc39a8*, *Gpr116*, *Eltd1*, *Ly75*, and *Cyyr1*; the remainder of the validated endothelial genes were uniquely identified by each study. The most likely explanations for the difference in the two studies are (1) our study examines actively transcribed RNA, whereas theirs looks at the entire transcriptome, or (2)

our study identified transcripts from all cells that ever expressed *Tie2:Cre*, whereas their study looked at cells acutely expressing *Tie2:GFP*. It is also possible that some of the endothelial genes uniquely found by TU tagging were down-regulated in response to cell dissociation or processing during the FACS purification. Nevertheless, the ability of each experiment (TU tagging and FACS) to identify a different pool of validated endothelial transcripts highlights the complementary nature of the methods.

TU tagging is also complementary to existing genetic methods for RNA purification: TRAP (translating ribosome affinity purification) (Doyle et al. 2008; Heiman et al. 2008), Ribo tag (Sanz et al. 2009), ribosome profiling (Ingolia et al. 2011; Brar et al. 2012), INTACT (isolation of nuclei tagged in cell types) (Deal and Henikoff 2011), and miRAP (miRNA affinity purification) (He et al. 2012) (Table 4). First, TU tagging has the potential to purify mRNA, noncoding RNA (ncRNA), miRNA, and ribosomal RNA (rRNA); each of the other methods purifies a more restricted population of RNA. Second, TU tagging requires only crossing the *CA>GFPstop>UPRT* line to any existing Cre line. In contrast, TRAP currently requires generating a distinct bacterial artificial chromosome (BAC) transgenic line for every cell type to be studied, which is a slow and expensive bottleneck. Third, TU tagging purifies RNA from all parts of the cell, whereas INTACT purifies only nuclear RNA. Fourth, TU tagging uses an exogenous soluble enzyme (UPRT) that has little or no effect on viability or fertility in mice or *Drosophila* (Miller et al. 2009; this study), the transcriptome of cultured human cells (Cleary et al. 2005), or the translation efficiency of thio-RNA (Cleary et al. 2005). In contrast, most other genetic methods require overexpression of an endogenous protein, such as a ribosomal structural protein, which may have deleterious effects due to altering protein complex stoichiometry. Nevertheless, each method has its strengths (for example, TU tagging is not suitable for bulk RNA purification, and the other methods are not designed to isolate nascent RNAs), highlighting the complementary nature of each method.

A major advantage of TU tagging is its ability to covalently label and purify nascent transcripts, especially for detecting dynamic changes in gene expression. This feature has been well documented in TU tagging experiments on cell lines (Cleary et al. 2005; Friedel and Dolken 2009), and we provide two lines of evidence for supporting this in mouse tissue. (1) We observed strong enrichment for intronic sequences in our RNA-seq analysis of TU-tagged RNA relative to total RNA (TU-tagged RNA: 75.3% intronic reads; total RNA: 55.9% intronic reads; brain endothelial experiments), which shows that we are enriching for nascent transcripts; purifying RNA sooner after 4TU injection (e.g., 1 or 2 h post-injection) may further increase the percentage of nascent RNA purified. (2) We found hemoglobin RNA to be among the most depleted heart or brain transcripts (Figs. 3B, 4B); the anucleate red blood cells do not transcribe the hemoglobin gene, showing that hemoglobin transcription is required for TU tagging and that mature hemoglobin RNA is not a substrate for 4TU incorporation. Overall, the ability of TU tagging to monitor active transcription makes it a powerful method to study the dynamics of gene expression.

How can TU tagging be used in the future? Our *CA:UPRT* transgene can be used for cell transplantation or coculture experiments, and UPRT can be delivered into specific cell types by viral transfection. Transplantation of $UPRT^+$ donor cells could include in vitro manipula-

Table 4. *Cell type-specific RNA purification methods in mice*

Method	Unique features	References
TU tagging	Labels within intact organism Spatial control: Cre line, transplantation, viral delivery Temporal control: 4TU pulse Labels RNAs transcribed during 4TU pulse Purifies RNA in all cellular processes Labels all species of RNA	This study
TRAP	Labels within intact organism Spatial control: requires BAC construction and validation Poor temporal resolution (transcription on/off) Labels translated mRNA only	Doyle et al. 2008
miRAP	Labels within intact organism Spatial control: Cre line Poor temporal resolution (transcription on/off) Labels miRNAs only	He et al. 2012
INTACT	Labels within intact organism Spatial control: Cre line Poor temporal resolution (transcription on/off) Labels nuclear RNA only	Deal and Henikoff 2011
FACS, panning, laser capture	Requires mechanical dissociation Removes target cells from endogenous niche May lose RNA in cellular processes Poor temporal resolution (transcription on/off) Purifies all RNA species	

tions prior to transplantation, thus allowing transgene expression or shRNA knockdown of target genes in the UPRT$^+$ cells. In addition, the CA>*GFPstop*>*UPRT* transgene can be used with any of the existing, well-validated Cre lines for defining cell type-specific gene expression programs. This method can be used for identifying cell type-specific transcripts (as we did here for endothelial cells) and should be particularly useful for developmental studies, where the vast majority of developmental processes arise from poorly characterized gene expression changes that frequently occur in small pools of cells.

Materials and methods

Generation of transgenic mice

The *CA:HA-UPRT* construct was made by digesting the vector pStec-1 (Transgenic Animal Facility, University of Iowa Carver College of Medicine) with HindIII and PstI for insertion of the ubiquitously expressing *CA* promoter 5′ to the pStec-1 mini exon/intron sequences and digesting with XbaI and SalI to insert the HA-UPRT coding sequence. We showed previously that addition of an HA epitope tag has little effect on UPRT enzymatic activity (Miller et al. 2009). The *CA:loxGFPStoplox-HA-UPRT* construct (called *CA>GFPstop>UPRT* in the text) was made by digesting the *CA:HA-UPRT* construct with Pst1 for insertion of a *lox:GFP:3xSV40:lox* cassette between the *CA* promoter and the first exon. Both transgenes were digested with HindIII and BamHI for purification from the pStec vector and injected into B6D2F1 oocytes (Transgenic Mouse Facility, University of Oregon); all subsequent generations were backcrossed to C57BL/6J. All offspring were genotyped via PCR analysis using primer set #1 for 35 cycles (20 sec at 95°C, 30 sec at 57°C, and 20 sec at 72°C) and primer set #2 for 35 cycles (20 sec at 95°C, 30 sec at 57°C, and 40 sec at 72°C). Primer set #1 was 5′-AGTGACAACCCCTCTGG ATG-3′ and 5′-CATCGGATCTAGCAGCACA-3′ (both in the UPRT coding region), and primer set #2 was 5′-TGGTTTCTAAAGGCGA GGAA-3′ in the intron and 5′-TTGCCAGTGGTGCAGATAAA-3′ in the UPRT coding region. The Animal Care and Use Committee at the University of Oregon approved all mouse procedures.

Immunofluorescent staining

Embryos were harvested in ice-cold phosphate-buffered saline (PBS), and whole-mount bright-field and fluorescent images were captured. Embryos were subsequently fixed overnight at 4°C in 4% paraformaldehyde, which was then replaced with PBS. To process for sectioning, embryos were dehydrated using an ethanol series followed by immersion in xylenes and then molten paraffin. Paraffin-embedded embryos were sectioned at 7 μm using a microtome (Leica) and mounted on microscope slides.

Prior to immunofluorescent staining, sections were rehydrated and subjected to the appropriate antigen retrieval method (described below). Sections were incubated in 100 mM glycine and then blocked with 5% normal goat serum in PBS. Anti-HA antibodies (Bethyl Laboratories, #A190-108A), anti-GFP antibodies (AVES, # GFP-1020), and anti-endomucin antibodies (eBiosciences, #14-5851-82) were used as primary antibodies diluted in blocking buffer. For anti-HA and anti-endomucin costaining, epitope retrieval was accomplished by boiling sections in Ready-to-Use Target Retrieval solution (Dako) for 10 min in a pressure cooker. For anti-HA and anti-GFP costaining, epitope retrieval was accomplished by incubating sections in 0.25% Trypsin-EDTA (Life Technologies) for 15 min at 37°C. Alexa-conjugated secondary antibodies (Life Technologies) were used to visualize antibody-bound antigens, and nuclei were stained using Hoechst (Life Technologies). Slides were mounted with VectaShield hard-set mounting medium (Vector Laboratories).

Immunofluorescent staining of postnatal tissues was performed on frozen sections. Mice at P6 were anesthetized and then perfused with cold 4% paraformaldehyde following standard procedures. Brains were isolated and post-fixed overnight at 4°C, followed by cryoprotection in 30% sucrose overnight at 4°C. The tissue was then embedded into optimal cutting temperature (OCT) embedding medium prior to cryosectioning. Slides were allowed to air dry before nonspecific binding of primary antibodies were blocked using 5% normal donkey serum. In addition to the anti-HA and anti-GFP antibodies described previously, PECAM was detected using an anti-PECAM antibody (BD Pharmingen, #553370). Alexa-conjugated secondary antibodies were used to visualize antibody-bound antigens, and nuclei were stained with DAPI. All imaging was performed on a Leica wide-field microscope except Supplemental Figure S3, which was imaged on an Olympus FV-1000 confocal microscope and analyzed with Fluoview 1000 software.

Bone marrow transplants and flow cytometry

Recipient B6.SJL-*Ptprc*^{a} *Pep3*^{b}/BoyJ mice (strain number 002014) were lethally irradiated with 950 rads and injected with donor bone marrow between 4 and 12 h post-irradiation. Bone marrow was aseptically harvested from C57BL/6 *CA:UPRT* mice carrying the CD45.2 allotype or from B6.SJL-*Ptprc*^{a} *Pep3*^{b}/BoyJ congenic mice carrying the CD5.1 allotype (Spangrude 2008). Cells were transplanted via retro-orbital injection at various ratios of the C57BL/6 *CA:UPRT* to B6.SJL-*Ptprc*^{a} *Pep3*^{b}/BoyJ cells. The cell ratios were confirmed post-transplantation by flow cytometry using antibodies specific for CD45.1 (CD45.1 FITC clone A20, Biolegend, #110705) and CD45.2 (CD45.2 PerCP-Cy5.5, clone 104, Biolegend, #110705). Each mouse received 10^6 cells. Following transplantation, peripheral blood was analyzed for chimerism, and after eight weeks, the spleens were harvested for Trizol extraction of RNA. A sample of each spleen was also analyzed by flow cytometry using the above antibodies to determine the percentage of live cells that were CD45.1$^+$, CD45.2$^+$, and CD45$^-$. The Animal Care and Use Committee at the University of California at Merced approved all mouse procedures.

4TU and LPS delivery

We dissolved 90 mg of 4TU (Aldrich, #440736) in 0.5 mL of DMSO by heating to 50°C with vortexing, and the solution was then diluted 1:25 in 50°C 0.05 M Tris-HCl (pH 8.8) and kept in the dark at 50°C until use. We delivered ~430 mg/kg body weight by intraperitoneal injections for adults and by subcutaneous neck injections for pregnant females and postnatal pups. As initial experiments showed progressively more TU-tagged RNA in livers harvested at 2 h, 4 h, and 6 h following 4TU injection, we harvested mice 4 or 6 h after 4TU exposure.

For LPS experiments, 4TU was prepared and injected as described above into 6-wk-old adult *Tie2:Cre; CA>GFPstop>UPRT* mice (t = 0 h). The LPS$^+$ experimental mice were injected intraperitoneally with LPS (Millipore, #LPS25) at 5 mg/kg 1 h after 4TU injection (t = 1 h). Control mice were not injected with LPS, and spleen tissue was isolated from experimental and control mice 4 h after 4TU injection (t = 4 h) and used for RNA isolation and sequencing as described below. LPS RNA-seq data were analyzed using the DESeq Bioconductor/R package (Anders and Huber 2010) to determine whether *Tie2:Cre* lineage spleen transcripts statistically significantly changed in expression upon LPS treatment. TU-tagged spleen RNA was compared between

two LPS-treated mice and one untreated mouse using count data generated using the Bowtie2/SAMtools workflow described below. An adjusted *P*-value (Benjamini-Hochberg) of <0.10 was set as the significance cutoff for differential expression.

RNA purification

For *CA:UPRT* chimera experiments, spleen tissue was homogenized in 1 mL of Trizol and incubated for 5 min at room temperature. Chloroform (200 μL) was added to the mix, and tubes were vortexed for 15 sec, incubated for 2–3 min at room temperature, and centrifugedat 12,000*g* for 15 min at 4°C . The aqueous phase was add to 0.5 mL of isopropyl alcohol, incubated for 10 min at room temperature, and centrifuged at 12,000*g* for 10 min at 4°C. Supernatant was removed, and the RNA pellets were rinsed with 75% ethanol and centrifuged for 5 min at 4°C. Supernatant was removed, tubes were inverted for 2–3 min, and RNA pellets were resuspended in 30 μL of RNase-free H_20. Concentration was determined, and 260 of 280 rations of >2.0 were verified.

For *Tie2:Cre; CA>GFPstop>UPRT* experiments, 50 μg of total RNA was purified as described above, fragmented using NEBNext Magnesium RNA Fragmentation Module (New England BioLabs, #E6150S) for 4 min at 94°C, purified by RNeasy minikit (Qiagen, #74104), eluted in 50 μL of RNase-free water, processed through Ribo-Zero Magnetic kit (Epicenter, #MRZH116) for ribosomal RNA removal, purified by RNeasy minikit, and eluted in 22 μL of RNase-free water. The RNA concentration was determined by Qubit, and a 2-μL sample was reserved as the total RNA control. Remaining RNA (18 μL) was biotinylated using 25 μL of EZ-Link Biotin HPDP (Pierce Biotechnology, #21341) dissolved to 1 mg/mL in N,N-dimethylformamide in total reaction volume of 100 μL of 1× TE (pH 8.0) for 3 h. The labeled RNA was purified by RNeasy minikit and eluted in 20 μL of RNase-free water. Biotinylated RNA was then isolated from nonlabeled RNA using a μMacs streptavidin kit (Miltenyi Biotec, #130-074-101) following the manufacturer's directions with the exception of elution from the column in 100 mM β-mercaptoethanol. Eluted RNA was purified by RNeasy minielute kit columns (Qiagen, #74204) and eluted in 14 μL of RNase-free EB buffer.

RNA-seq

The ScriptSeq version 2 RNA-Seq Library Preparation kit (Epicenter, #SSV21106) was used for RNA-seq library preparation. The streptavidin-purified RNA (9.5 μL of the 14 μL above) was processed according to the manufacturer's directions for fragmented RNA starting at Appendix steps 5.1.A and 5.1.B and then continuing at step 4C with the standard kit procedure. Agencourt AMPure XP bead (BeckmanCoulter, #A63880) purification was used as recommended. The library was amplified for 15 cycles using ScriptSeqIndex PCR primers (Epicenter, #RSBC10948). The final DNA was eluted in 20 μL of water. A small aliquot from each RNA sample processed was run on an agarose gel for ethidium bromide visualization and used for Qubit quantification. The resulting indexed Illumina libraries from four separate RNA samples were mixed at equal concentrations and sequenced together in one lane of an Illumina HiSeq 2000 using version 3 sequencing reagents.

Initial characterization of the RNA-seq data revealed that a high percentage of reads from TU-tagged RNA aligned to intronic regions, presumably because nascent TU-tagged RNA is enriched for pre-mRNAs. We therefore developed a custom analysis pipeline that included all reads that mapped between the start of the first exon and the end of the last exon of each gene. For each experiment, the *s*equence reads were aligned against the University of California at Santa Cruz (UCSC) mm9/NCBI build 37 genome sequence using Bowtie2 (Langmead and Salzberg 2012) in "sensitive-local" mode, and the number of reads mapping to each gene region was determined using the SAMtools (Li et al. 2009) "view" command. The coordinates of each gene region were determined from a gene transfer format (GTF) annotation file downloaded from UCSC on January 27, 2011. RPM expression values were calculated by normalizing the number of reads that mapped to each gene region to the total number of reads that mapped to all gene regions. RPM values from biological replicate experiments were averaged.

Acknowledgments

We thank Richard Goodman (Vollum Institute), Cris Niell (University of Oregon), and Karen Guillemin (University of Oregon) for comments on the manuscript; Curt Sigmund (University of Iowa) for the pSTEC plasmid; and Ute Hostick and the Oregon Transgenic Mouse Facility (Eugene, Oregon) for help making the transgenic mice. Funding was provided by NIH/NHLBI 5R00HL087598 and a March of Dimes Basil O'Connor Award (to K.S.), Department of Defense Peer-Reviewed Cancer Research Program CA100469 and The Pew Charitable Trust (to H.Z.), an NRSA predoctoral fellowship (to P.B.V.), California Institute for Regenerative Medicine RT1-01052-1 (to M.D.C.), Empire State Stem Cell Fund from New York State Department of Health Contract C024352 (to S.T.), and the Howard Hughes Medical Institute (to C.Q.D.), where C.Q.D. is an Investigator.

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11:** R106.

Balda MS, Matter K. 2009. Tight junctions and the regulation of gene expression. *Biochim Biophys Acta* **1788:** 761–767.

Bartfai T, Buckley PT, Eberwine J. 2012. Drug targets: Single-cell transcriptomics hastens unbiased discovery. *Trends Pharmacol Sci* **33:** 9–16.

Benicky J, Sanchez-Lemus E, Pavel J, Saavedra JM. 2009. Anti-inflammatory effects of angiotensin receptor blockers in the brain and the periphery. *Cell Mol Neurobiol* **29:** 781–792.

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. 2012. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335:** 552–557.

Buttini M, Boddeke H. 1995. Peripheral lipopolysaccharide stimulation induces interleukin-1β messenger RNA in rat brain microglial cells. *Neuroscience* **65:** 523–530.

Chen SK, Tvrdik P, Peden E, Cho S, Wu S, Spangrude G, Capecchi MR. 2010. Hematopoietic origin of pathological grooming in Hoxb8 mutant mice. *Cell* **141:** 775–785.

Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. 2005. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* **23:** 232–237.

Daneman R, Zhou L, Agalliu D, Cahoy JD, Kaushal A, Barres BA. 2010. The mouse blood-brain barrier transcriptome: A new resource for understanding the development and function of brain endothelial cells. *PLoS ONE* **5:** e13741.

Deal RB, Henikoff S. 2011. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat Protoc* **6:** 56–68.

Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D, Magen A, Canidio E, Pagani M, Peluso I, et al. 2011. A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol* **9:** e1000582.

Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, et al. 2008. Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135:** 749–762.

Ekman N, Lymboussaki A, Vastrik I, Sarvas K, Kaipainen A, Alitalo K. 1997. Bmx tyrosine kinase is specifically expressed in the endocardium and the endothelium of large arteries. *Circulation* **96:** 1729–1732.

Friedel CC, Dolken L. 2009. Metabolic tagging and purification of nascent RNA: Implications for transcriptomics. *Mol Biosyst* **5:** 1271–1278.

Guez-Barber D, Fanous S, Harvey BK, Zhang Y, Lehrmann E, Becker KG, Picciotto MR, Hope BT. 2012. FACS purification of immunolabeled cell types from adult rat brain. *J Neurosci Methods* **203:** 10–18.

Hammer M, Mages J, Dietrich H, Servatius A, Howells N, Cato ACB, Lang R. 2006. Dual specificity phosphatase 1 (DUSP1) regulates a subset of LPS-induced genes and protects mice from lethal endotoxin shock. *J Exp Med* **203:** 15–20.

He M, Liu Y, Wang X, Zhang MQ, Hannon GJ, Huang ZJ. 2012. Cell-type-based analysis of microRNA profiles in the mouse brain. *Neuron* **73:** 35–48.

Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suarez-Farinas M, Schwarz C, Stephan DA, Surmeier DJ, et al. 2008. A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135:** 738–748.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:** 789–802.

Kisanuki YY, Hammer RE, Miyazaki J, Williams SC, Richardson JA, Yanagisawa M. 2001. Tie2-Cre transgenic mice: A new model for endothelial cell-lineage analysis in vivo. *Dev Biol* **230:** 230–242.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lobo MK, Karsten SL, Gray M, Geschwind DH, Yang XW. 2006. FACS-array profiling of striatal projection neuron subtypes in juvenile and adult mouse brains. *Nat Neurosci* **9:** 443–452.

Matei V, Pauley S, Kaing S, Rowitch D, Beisel KW, Morris K, Feng F, Jones K, Lee J, Fritzsch B. 2005. Smaller inner ear sensory epithelia in Neurog 1 null mice are related to earlier hair cell cycle exit. *Dev Dyn* **234:** 633–650.

Melet F, Motro B, Rossi DJ, Zhang L, Bernstein A. 1996. Generation of a novel Fli-1 protein by gene targeting leads to a defect in thymus development and a delay in Friend virus-induced erythroleukemia. *Mol Cell Biol* **16:** 2708–2718.

Miller MR, Robinson KJ, Cleary MD, Doe CQ. 2009. TU-tagging: Cell type-specific RNA isolation from intact complex tissues. *Nat Methods* **6:** 439–441.

Patel S, Leal AD, Gorski DH. 2005. The homeobox gene Gax inhibits angiogenesis through inhibition of nuclear factor-κB-dependent endothelial cell gene expression. *Cancer Res* **65:** 1414–1424.

Qin L, Wu X, Block ML, Liu Y, Breese GR, Hong JS, Knapp DJ, Crews FT. 2007. Systemic LPS causes chronic neuroinflammation and progressive neurodegeneration. *Glia* **55:** 453–462.

Sanna PP, Repunte-Canonigo V, Guidotti A. 2012. Gene profiling of laser-microdissected brain regions and individual cells in drug abuse and schizophrenia research. *Methods Mol Biol* **829:** 541–550.

Sanz E, Yang L, Su T, Morris DR, McKnight GS, Amieux PS. 2009. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc Natl Acad Sci* **106:** 13939–13944.

Smith-Berdan S, Nguyen A, Hassanein D, Zimmer M, Ugarte F, Ciriza J, Li D, García-Ojeda ME, Hinck L, Forsberg EC. 2011. Robo4 cooperates with CXCR4 to specify hematopoietic stem cell localization to bone marrow niches. *Cell Stem Cell* **8:** 72–83.

Spangrude GJ. 2008. Assessment of lymphocyte development in radiation bone marrow chimeras. *Curr Protoc Immunol* **81:** 4.6.1–4.6.9.

Stepniak E, Radice GL, Vasioukhin V. 2009. Adhesive and signaling functions of cadherins and catenins in vertebrate development. *Cold Spring Harb Perspect Biol* **1:** a002949.

Takamatsu H, Takegahara N, Nakagawa Y, Tomura M, Taniguchi M, Friedel RH, Rayburn H, Tessier-Lavigne M, Yoshida Y, Okuno T, et al. 2010. Semaphorins guide the entry of dendritic cells into the lymphatics by activating myosin II. *Nat Immunol* **11:** 594–600.

Takash W, Canizares J, Bonneaud N, Poulat F, Mattei MG, Jay P, Berta P. 2001. SOX7 transcription factor: Sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling. *Nucleic Acids Res* **29:** 4274–4283.

Tang Y, Harrington A, Yang X, Friesel RE, Liaw L. 2010. The contribution of the Tie2$^+$ lineage to primitive and definitive hematopoietic cells. *Genesis* **48:** 563–567.

Figure S1. *Tie2:Cre* induced UPRT endothelial expression in the P6 brain.
Tie2:Cre; CA>GFPstop>UPRT double transgenic shows robust UPRT expression in PECAM1$^+$ endothelial cells in all brain regions. (A) Low magnification image of the P6 brain stitched together from high magnification panels; PECAM1, red; DAPI$^+$ nuclei, blue. (B-K) Indicated regions from A are shown at high magnification for PECAM1 and UPRT (single label panels) together with a DNA counterstain (PECAM1, red; UPRT, green). UPRT expression is detected by anti-HA antibody staining of the HA-UPRT fusion protein. Scale bar, 200 μm.

Figure S2. *Tie2:Cre* induced UPRT endothelial expression at embryonic day 11.5.
E11.5 embryos antibody stained for the indicated markers in the indicated tissues.
(A-D) *CA>GFPstop>UPRT* single transgenic. Note absence of UPRT staining (green).
(E-H) *Tie2:Cre; CA>GFPstop>UPRT* double transgenic. Note robust UPRT staining (green) in Endomucin$^+$ endothelial and endocardial cells (red). UPRT is also expressed, as expected, in all atrioventricular canal cushion mesenchymal cells and a subset of mesenchymal cells of the outflow tract cushions in the heart (G). UPRT expression is detected by anti-HA antibody staining of the HA-UPRT fusion protein. Red blood cells (white) are detected by their strong auto-fluorescence. Nuclei (blue) are stained with Hoechst.

Figure S3. *Math1:Cre* induced UPRT expression in the P6 brain.
(A) *Math1:Cre; CA>GFPstop>UPRT* double transgenic postnatal day 7 brain sagittal section. Note the mutually exclusive expression of UPRT (red) and GFP (green). UPRT is detected in the cerebellum (top box, enlarged in B), brainstem (bottom box, enlarged in C) and pons (arrowhead); all regions known to express *Math1:Cre* (Wang et al. 2005). (B) UPRT+ (red) granule neuron progenitors in the inner granule layer (IGL) and external granule layer (EGL) but not in the GFP-positive (green) Purkinje cell layer (PCL) or meninges. (C) UPRT-positive (red) cochlear neurons (CN) in brainstem adjacent to GFP+ cells (green). In all panels, UPRT expression is detected by HA antibody staining of the HA:UPRT fusion protein. Scale bars, 50 μm.

Figure S4. Expression patterns of some of the most-enriched and most-depleted transcripts from the P6 brain TU-tagging experiment.
Enriched (A) and depleted (B) transcripts. All images are from the Eurexpress E14.5 database. Gene symbol and fold enrichment indicated. *Slc22a8* is regionally expressed in endothelial cells of the midbrain and more caudal CNS (bracket) but not in the forebrain (arrow). Cx3cr1 shows the microglial pattern (Imai et al. 1997; Harrison et al. 1998); transcripts with this

pattern are likely to be identified based on Tie2:Cre expression in their progenitors, resulting in UPRT expression in these microglial cells.
(C) Background TU-tagging does not preferentially label endothelial/vasculature transcripts. GO terms and P values for the top 130 genes (based on RPM values) from TU-tagged P42 brain tissue in *Tie2:Cre* mice lacking a UPRT transgene shows over-representation of neural terms but not endothelial/vasculature terms.

Figure S5. Expression patterns of a subset of the most-enriched heart endothelial/vasculature transcripts, and one depleted transcript.
Enriched (A) and depleted (B) transcripts. All images are RNA in situ hybridizations of E14.5 hearts taken from the Eurexpress transcriptome atlas database. Gene symbol and fold enrichment indicated. We observed enrichment of all genes shown with the exception of the bottom row of four myocardial-expressed transcripts (*Tnnt2*, *Tnni3*, *Myl4*, and *Nppa*) that were depleted from the TU-tagged heart endothelial RNA. The first four transcripts (*Egfl7*, *Emcn*, *Tek*, and *Nos3*) are enriched positive control transcripts showing both endocardial and coronary endothelial expression. *Eltd1*, *F11r*, *She*, *Prkd2*, and *Ppp1r16b* are expressed in at least a subset of both endocardial and coronary endothelial cells. *Apold1*, *Meox2*, *Sipa1*, and *Cyyr1* are detected only in coronary endothelium and not endocardial cells. *Ptprb* is expressed in coronary endothelium and atrioventricular canal (AVC) cushion (developing valve) mesenchyme. *Cdc42ep4* is preferentially found in AVC cushion mesenchyme. *Pecam1* is another, less-enriched, positive control endothelial transcript whose protein product was detected by antibody staining in other Figures. Arrows mark ventricular endocardial cells, arrowheads denote coronary endothelial cells, and the asterisk indicates AVC cushion mesenchyme.
(C) Background TU-tagging does not preferentially label endothelial/vasculature transcripts. GO terms and P values for the top 130 genes (based on RPM values) from TU-tagged P42 heart tissue in *Tie2:Cre* mice lacking a UPRT transgene shows over-representation of cardiac terms but not endothelial/vasculature terms.

Figure S6. RNA fragmentation prior to streptavidin purification reduces background.
To improve our published TU-tagging protocol, we added an RNA fragmentation step (see methods). (Left) Brain endothelial TU-tagging experiment without fragmentation step. Only 5 of 11 control pan-endothelial genes (red) are enriched compared to the majority of genes. (Right) Brain endothelial TU-tagging experiment with the fragmentation step. 11 of the 13 positive control pan-endothelial genes (red) are enriched compared to the majority of genes.

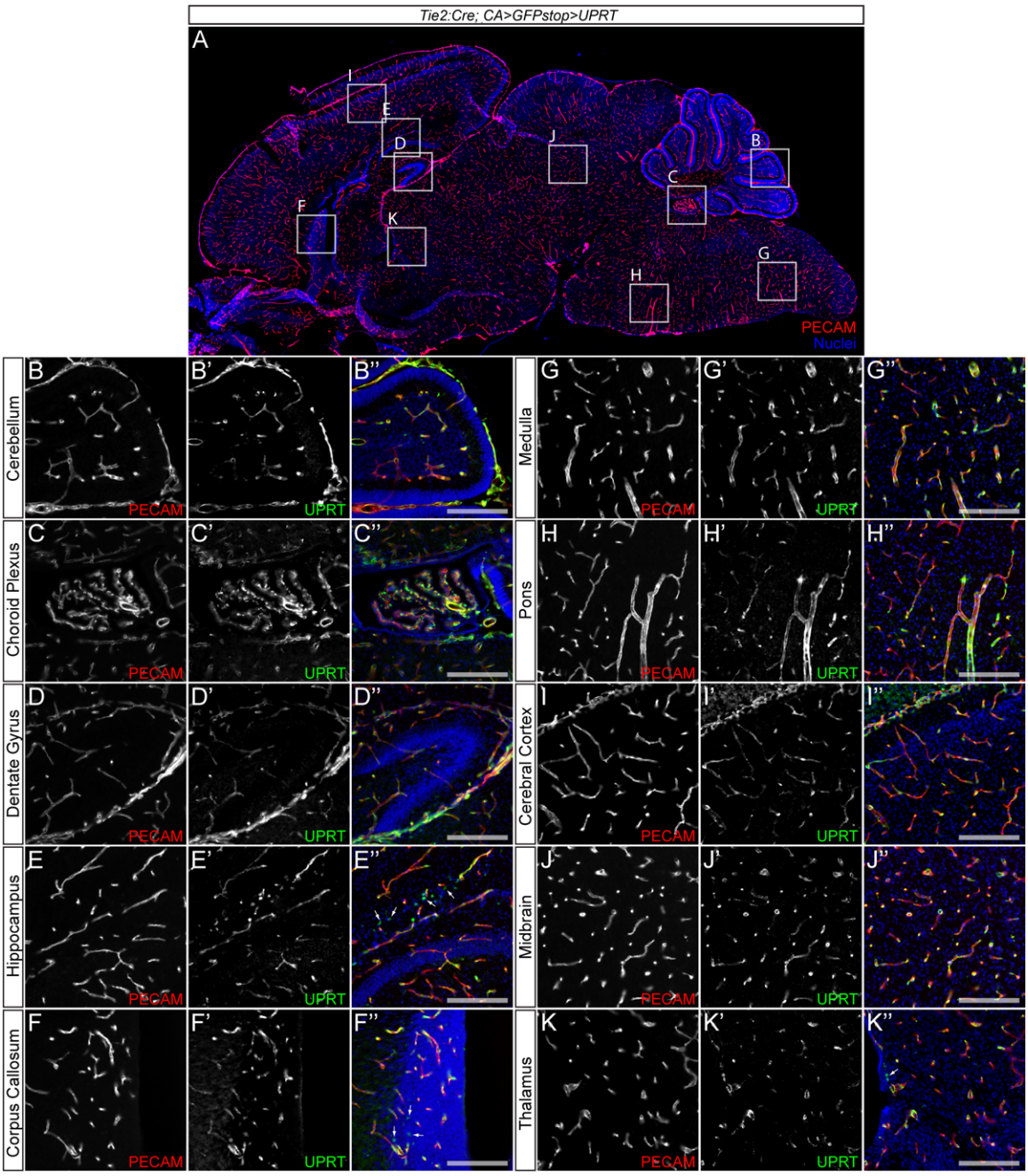
Table S1. RNA-seq data for P6 *Tie2:Cre; CA>GFPstop>UPRT* total and TU-tagged brain RNA.

Table S2. RNA-seq data for P6 *Tie2:Cre; CA>GFPstop>UPRT* total and TU-tagged heart RNA.

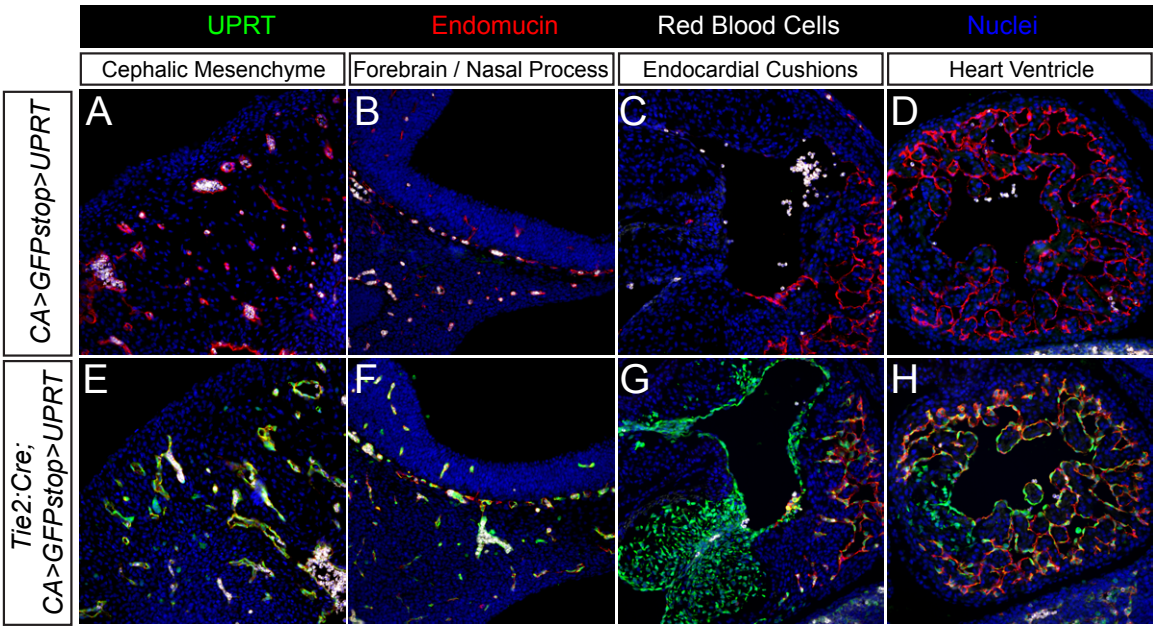
Table S3. RNA-seq data for E15.5 *Tie2:Cre; CA>GFPstop>UPRT* total and TU-tagged brain RNA.

Table S4. RNA-seq data for TU-tagged spleen RNA from LPS injected or uninjected P42 *Tie2:Cre; CA>GFPstop>UPRT* adult mice.

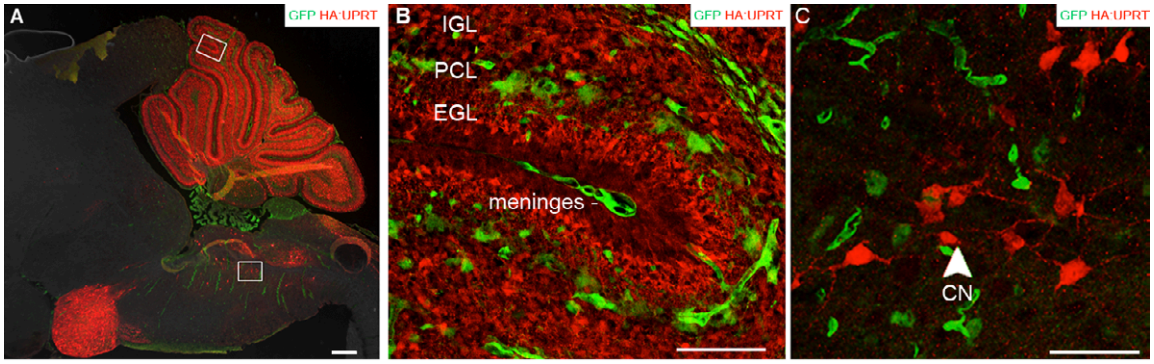
Table S5. RNA-seq data for TU-tagged RNA from *CA:UPRT* donor bone marrow transplanted into unlabeled host spleen compared to total spleen RNA.



Gay et al., Supplemental Figure 1

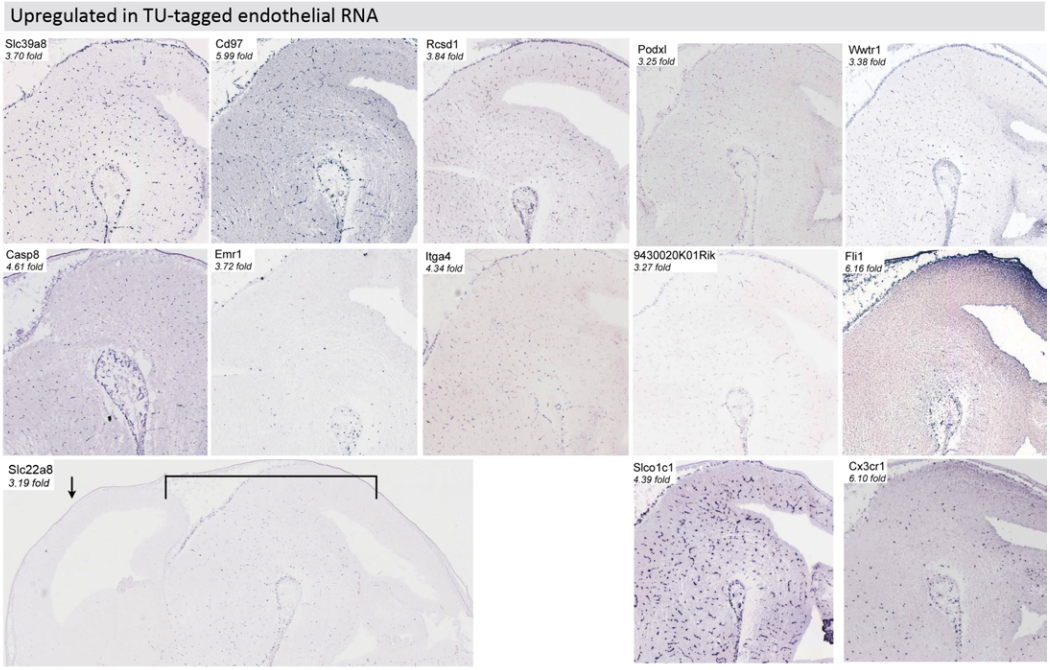


Gay et al. Supplemental Figure 2

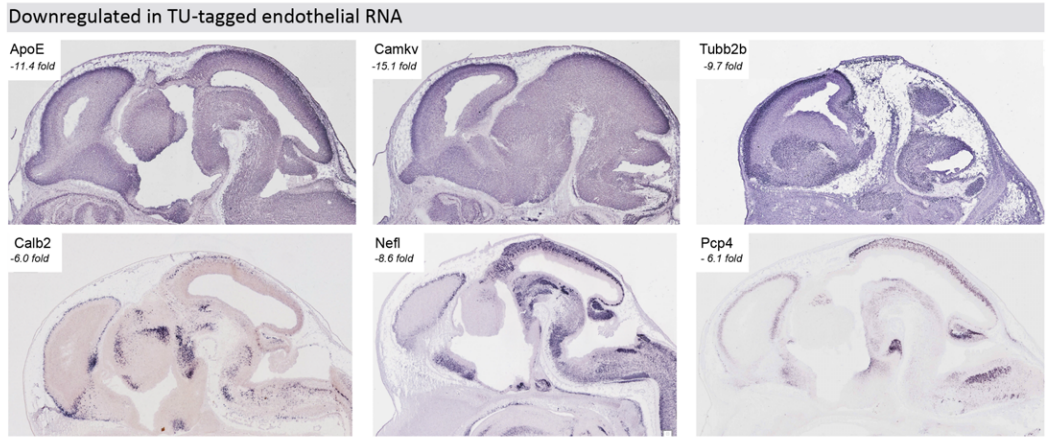


Gay et al., Supplemental Figure 3

A



B

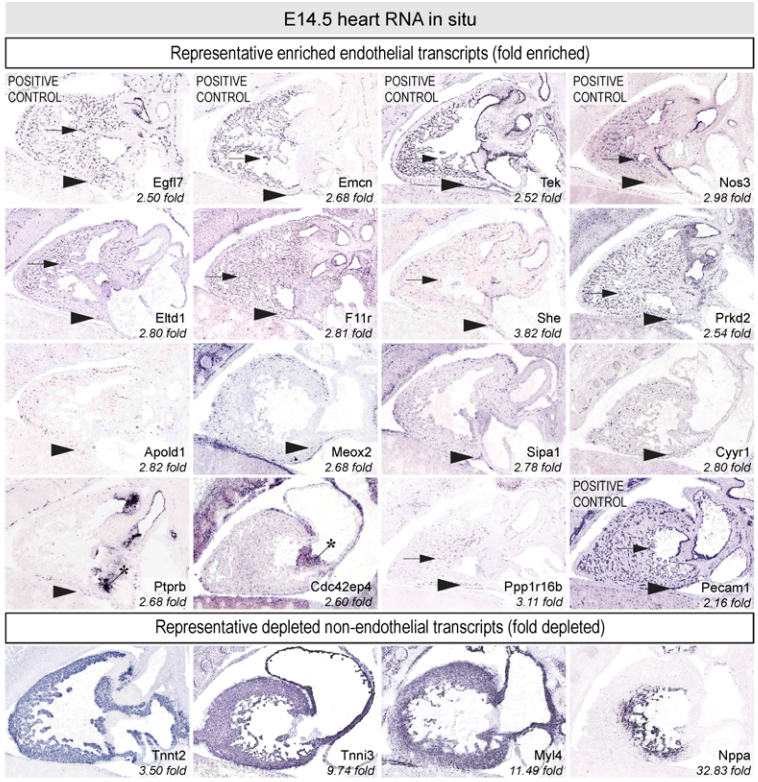


C

GO term	P value
multicellular organismal sig.	3.57 e-20
synaptic transmission	9.98 e-19
cell-cell sig.	3.13 e-15
nervous system dev	4.47 e-14
neurogenesis	1.94 e-13
neuron differentiation	1.38 e-12
neuron projection dev.	5.12 e-11
regul. of localization	8.57 e-11
multicell. organism process	1.99 e-10
regul. of transport	4.59 e-10
regul. nervous system dev.	8.27 e-10
anatomical struct. dev.	1.02 e-09
regul. cell dev.	4.64 e-09
glutamate receptor sig. p'way	1.81 e-08
behavior	5.17 e-08

Gay et al., Supplemental Figure 4

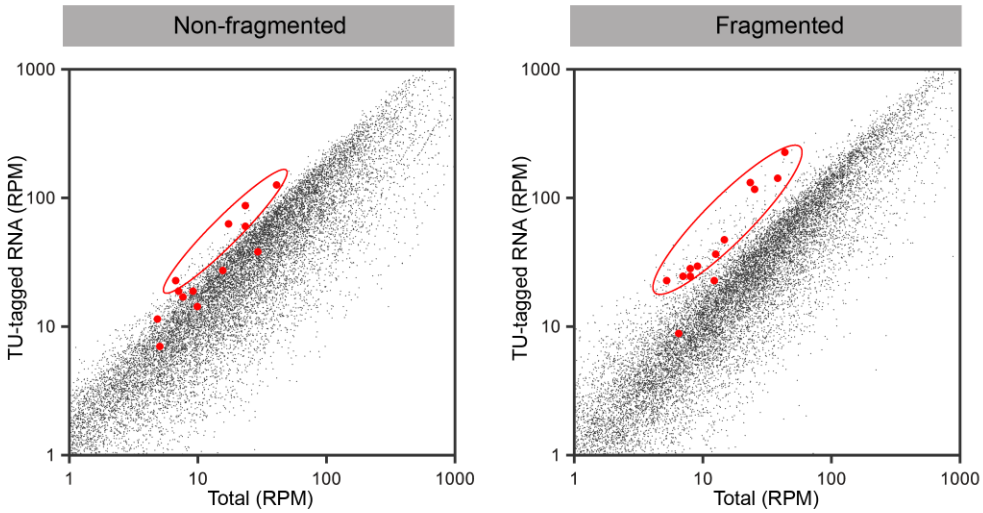
A



C

GO term	P value
muscle system process.	4.71 e-13
cardiovascular system dev.	6.98 e-13
cell dev.	9.63 e-13
muscle structure dev.	3.10 e-12
muscle cell differentiation	6.60 e-12
heart dev.	8.36 e-12
anatomical structure morph.	2.72 e-10
muscle contraction	3.68 e-10
circulatory system process	7.27 e-10
cell differentiation	1.60 e-09
muscle cell dev.	2.65 e-09
cell. dev. process	2.99 e-09
cell component morph.	3.33 e-09
striated muscle tissue dev.	3.37 e-09
muscle tissue dev.	6.71 e-09

Gay et al. Supplemental Figure 5


Non-fragmented
Fragmented
TU-tagged RNA (RPM)
Total (RPM)
1000
100
10
1
1
10
100
1000