Massively parallel, time-resolved single-cell RNA sequencing with scNT-Seq

Qi Qiu[1,2,#], Peng Hu[1,2,#], Kiya W. Govek[1,3], Pablo G. Camara[1,3], Hao Wu[1,2,*]

[1]Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104
[2]Penn Epigenetics Institute, University of Pennsylvania, Philadelphia, PA 19104
[3]Institute of Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104

[#]These authors contributed equally to this work
*Correspondence should be addressed to H.W. (haowu2@pennmedicine.upenn.edu).

ABSTRACT

Single-cell RNA sequencing offers snapshots of whole transcriptomes but obscures the temporal dynamics of RNA biogenesis and decay. Here we present single-cell new transcript tagging sequencing (scNT-Seq), a method for massively parallel analysis of newly-transcribed and pre-existing RNAs from the same cell. This droplet microfluidics-based method enables high-throughput chemical conversion on barcoded beads, efficiently marking metabolically labeled newly-transcribed RNAs with T-to-C substitutions. By simultaneously measuring new and old transcriptomes, scNT-Seq reveals neuronal subtype-specific gene regulatory networks and time-resolved RNA trajectories in response to brief (minutes) versus sustained (hours) neuronal activation. Integrating scNT-Seq with genetic perturbation reveals that DNA methylcytosine dioxygenases may inhibit stepwise transition from pluripotent embryonic stem cell state to intermediate and totipotent two-cell-embryo-like (2C-like) states by promoting global RNA biogenesis. Furthermore, pulse-chase scNT-Seq enables transcriptome-wide measurements of RNA stability in rare 2C-like cells. Time-resolved single-cell transcriptomic analysis thus opens new lines of inquiry regarding cell-type-specific RNA regulatory mechanisms.

INTRODUCTION

Dynamic changes in RNA levels are regulated by the interplay of RNA transcription, processing, and degradation[1, 2]. Understanding the mechanisms by which the transcriptome is regulated in functionally diverse cell-types within multi-cellular organisms thus requires cell-type-specific measurements of the kinetics of RNA biogenesis and decay. Recent advances in single-cell RNA sequencing (scRNA-Seq) technologies are leading to a more complete understanding of heterogeneity in cell types and states[3]. However, standard scRNA-Seq methods capture a mixture of newly-synthesized ("new") and pre-existing ("old") RNAs without being able to temporally resolve RNA dynamics.

Accurately capturing changes in the nascent transcriptome over time at single-cell resolution is of particular interest because measuring newly-transcribed RNAs can reveal immediate regulatory changes in response to developmental, environmental, metabolic, and pathological signals[4]. Commonly used approaches for distinguishing newly-transcribed from pre-existing RNAs of the same population of transcripts rely on metabolic labeling that employs thiol-labeled nucleoside analogs such as 4-thiouridine (4sU) and subsequent biochemical enrichment of metabolically labeled RNAs[2]. Although these methods have yielded unprecedented insights into the regulation of RNA dynamics, they require ample starting material and present challenges for enrichment normalization. Several approaches have recently been developed to chemically convert 4sU into cytidine analogs, yielding uracil-to-cytosine (U-to-C) substitutions that label newly-transcribed RNAs after reverse transcription[5-7]. These chemical nucleotide conversion methods allow for direct measurement of temporal information about cellular RNAs in a sequencing experiment without biochemical enrichment. Recent studies integrated standard plate-based single-cell RNA-Seq method with one of these chemical methods (i.e. thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM)-Seq)[8, 9], demonstrating the feasibility of studying newly-transcribed transcriptomes at single-cell levels. However, these plate-based single-cell SLAM-Seq methods suffer from several limitations. First, they are costly, and the associated library preparation steps are time-consuming, prohibiting it for large-scale single-cell analysis of highly heterogeneous cell populations. Second, these methods lack unique molecular identifiers (UMIs), preventing accurate quantification of the newly-transcribed transcript levels.

To overcome these constraints, we developed single-cell new transcript tagging sequencing (scNT-Seq), a high-throughput and UMI-based scRNA-Seq method that simultaneously measures both new and old transcriptomes from the same cell. In scNT-Seq, integration of metabolic RNA labeling, droplet microfluidics, and chemically induced recoding of 4sU to cytosine analog permits highly scalable and time-resolved single-cell analysis of cellular RNA dynamics. We demonstrate that the method is easy to set up and substantially improves the time and cost. We show scNT-Seq enables more detailed characterization of gene regulatory networks and temporal RNA trajectories than single-cell whole-transcriptome measurements alone.

RESULTS

Development and validation of scNT-Seq. To enable newly-transcribed and pre-existing transcripts from the same cell to be analyzed in a scalable manner, we focused on the Drop-Seq platform because its unique barcoded bead design affords both immobilization of RNAs for efficient chemical conversion reactions and UMI-based high-throughput scRNA-Seq analysis, and this droplet microfluidics platform has been widely adopted[10-14]. The scNT-Seq consists of the following key steps (**Fig. 1a**): (1) metabolically labeling of cells with 4sU for a temporally defined time period; (2-3) co-encapsulating each individual cell with a barcoded oligo-dT primer coated bead in a nanoliter-scale droplet which captures both newly-transcribed and pre-existing RNAs; (4) performing 4sU chemical conversion on pooled barcoded beads in one reaction after droplet breakage; (5-8) reverse transcription, cDNA amplification, tagmentation, indexing PCR, and sequencing; and (9) using a UMI-based statistical model to analyze T-to-C substitutions within transcripts and infer the new transcript fraction at single-cell level.

To identify the optimal reaction conditions on barcoded beads, we first explored two independent chemical conversion methods (TimeLapse-Seq: 2,2,2-trifluoroethylamine (TFEA)/sodium periodate ($NaIO_4$)-based 4sU conversion; SLAM-Seq: iodoacetamide (IAA)-based conversion) and validated their performance with species-mixing experiments using cultured mouse embryonic stem cells (mESCs) and human K562 cells. This analysis indicates that TFEA/$NaIO_4$-based scNT-Seq substantially outperforms the IAA-based assay in terms of mRNA recovery rates (**Fig. 1b**). Furthermore, the collision rate is comparable between TFEA/$NaIO_4$-based scNT-Seq and standard Drop-Seq (**Fig. 1b**), demonstrating the specificity of scNT-Seq in analyzing single-cell transcriptomes. Shallow sequencing of mESCs under different treatment conditions showed that TFEA/$NaIO_4$-based scNT-Seq identified a similar number of genes or UMIs per cell compared to standard Drop-Seq (**Supplementary Fig. 1a**). Aggregated single-cell new or old transcriptomes were highly correlated between replicates (**Supplementary Fig. 1b**). Further analysis of K562 cells revealed that only 4sU labeling and TFEA/$NaIO_4$ treatment resulted in a substantial increase in T-to-C substitution (**Fig. 1c**) and in UMIs that contained one or more such substitutions (**Fig. 1d, e**). Notably, TFEA/$NaIO_4$ chemical treatment works efficiently with both freshly isolated and cryo-preserved (methanol-fixed) cells (**Supplementary Fig. 1c**), demonstrating the versatility of scNT-Seq. Collectively, these data indicate that TFEA/$NaIO_4$-based scNT-Seq is capable of efficiently detecting newly synthesized transcripts at single-cell resolution.

Application of scNT-Seq to study neuronal activity-dependent transcription and regulatory networks.

Neuronal activity induces expression of hundreds of activity-regulated genes (ARGs) in the vertebrate brain, leading to new protein synthesis and epigenetic changes necessary for short- and long-term memories of experiences. Thus, the coupling of synaptic activity to nascent transcription in the nucleus allows neurons to both respond dynamically to their immediate environment, and to store information

97 stably[15]. Recent studies suggest that different patterns of neuronal activity could induce a distinct set of
98 ARGs[16]. To investigate cell-type-specific transcription in response to distinct activity-duration patterns
99 (brief versus sustained stimulation), we applied scNT-Seq to primary cortical neuronal cultures, derived
100 from mouse cortex (embryonic day 16 (E16)), which contains heterogeneous populations of neuronal
101 subtypes and undifferentiated neural progenitor cells.

102 To detect activity-regulated immediate transcriptional responses, we metabolically labeled primary mouse
103 cortical cultures (200 μM 4sU) for two hours and stimulated the cells with different durations of neuronal
104 activity (0, 15, 30, 60 and 120 min of potassium chloride (KCl)-mediated membrane depolarization) (**Fig.**
105 **2a**). After filtering low-quality cells and potential doublets, we retained 20,547 single-cell transcriptomes
106 from five time-points (**Fig. 2b** and **Supplementary Table 1**). After sample integration, principal
107 component analysis (PCA), graph-based clustering, and visualization by Uniform Manifold Approximation
108 and Projection (UMAP), we identified nine major cell-types based on known marker genes: *Neurod6*+
109 cortical excitatory neurons (Ex, 68.5%), four *Gad1*+ inhibitory neuronal subtypes (Inh1-4, 13.9% in total),
110 *Dlx1/Dlx2*+ inhibitory neuronal precursors (Inh-NP: 1.7%), two sub-populations of *Nes/Sox2*+ excitatory
111 neuronal precursors (Ex-NP1/2: 10.4% in total), and *Nes/Aldh1l1*+ radial glia (RG: 5.5%) (**Fig. 2b** and
112 **Supplementary Fig. 2b, c**).

113 We next sought to distinguish newly-transcribed from pre-existing RNAs by counting and statistically
114 modeling T-to-C substitutions in transcripts (UMIs), an approach that overcomes the problem of
115 incomplete 4sU labeling of new transcripts (up to 50% of all reads originated from new RNAs may not
116 contain T-to-C substitutions)[17]. The quantification accuracy is further improved by UMI-based analysis,
117 which substantially increases the number of uridines or T-to-C substitutions covered in each transcript
118 compared to the analysis of individual sequencing reads (**Supplementary Fig. 3a, b**). After statistical
119 correction, we obtained reliable measurements of newly-transcribed RNA faction for both activity-induced
120 genes (e.g., *Fos*, ~90% new/total) and slow turnover house-keeping genes (e.g., *Mapt*, <5% new/total) in
121 excitatory neurons (**Supplementary Fig. 3c, d**).

122 Furthermore, sub-clustering of newly-transcribed (upper) or pre-existing (lower) transcriptomes derived
123 from randomly sampled excitatory neurons could readily separate new, but not old single-cell
124 transcriptomes in an activity-pattern dependent manner (right panels in **Fig. 2b**). In contrast, new
125 transcriptomes of undifferentiated excitatory neural precursors (Ex-NP/RG) did not exhibit similar
126 distributions (**Supplementary Fig. 4a**). Next, we directly examined how transcription of classic ARGs are
127 induced in different cell-types in response to distinct activity-patterns. While some ARGs, such as *Jun* and
128 *Btg2*, were specifically induced in Ex neurons upon activation, other ARGs (e.g., *Egr1*, *Fos*, and *Npas4*),
129 were broadly induced in many cell-types including non-neuronal cells, albeit with different magnitudes and
130 response curves (**Fig. 2c, Supplementary Fig. 4b**). There was little to no change at pre-existing levels
131 upon activation of Ex neurons (**Fig. 2c**). Together, these results suggest scNT-Seq can accurately detect

cell-type-specific, activity-induced immediate transcriptional changes within the timescale of minutes to hours.

Recent advances in computational analysis enable identification of specific gene regulatory networks (GRNs, also known as regulon of a transcription factor (TF)) underlying stable cell states by linking *cis*-regulatory sequences to single-cell gene expression. Since the regulon is scored as a whole, instead of using the expression of individual genes, this approach is more robust against experimental dropouts or stochastic variation of gene expression due to transcriptional bursting. We reasoned that applying such an approach to newly-transcribed RNAs derived from scNT-Seq may allow for analysis of regulons underlying dynamic cell states induced by external stimuli such as neuronal activity. By applying single-cell regulatory network inference and clustering (SCENIC)[18] to statistically-corrected newly-transcribed and pre-existing single-cell transcriptomes, we identified 79 regulons with significant *cis*-regulatory motif enrichment showing significant changes in response to at least one activity-pattern (**Fig. 2d**, **Supplementary Fig. 5**). Many early-response ARGs encode TFs that regulate a subsequent wave of late-response gene expression. SCENIC analysis of newly-transcribed RNAs revealed neuronal activity-dependent increase in TF regulon activity of well-established IEG TFs (e.g., Fos, Jun, and Egr family of TFs) as well as constitutively expressed TFs such as Srf and Mef2, both of which are activated by multiple calcium-dependent signaling pathways and undergo post-translational modifications (e.g., phosphorylation) in response to neuronal activity (**Fig. 2d**). Thus, these TFs represent a group of TFs that are the main mediator of activity-dependent transcription. This group also included several TFs that have not been previously implicated in neuronal activation, including Maff and Hspa5, both of which exhibited higher expression and regulon activity upon activation (**Fig. 2d**). In addition, both Rfx3, a ciliogenic TF[19], and the neural cell fate regulator Pou3f1[20] were identified as potential novel regulators induced by neuronal activity. Furthermore, we observed regulon activity of cell-type-specific TFs (Neurod1/2 for Ex, Sox2/3 for Ex-NP and Dlx1/2 for Inh), which are associated with stable cell-type identities and detected in both newly-transcribed and pre-existing transcriptomes (**Fig. 2d**). Collectively, these data suggest that scNT-Seq allows for the analysis of cell-type-specific TF regulons that underly the dynamic cellular responses of both new and old transcriptomes to acute and sustained stimuli.

scNT-Seq enables newly-synthesized transcripts based RNA velocity analysis

A fundamental question in gene regulation is how transcriptional states in single cells change over time in response to both acute (minutes) and sustained (hours to days) external stimuli. Recent work showed that the time derivative of the gene expression state, termed “RNA velocity,” can be estimated by distinguishing between unspliced (intronic reads) and spliced (exonic reads) mRNAs in scRNA-Seq datasets[21]. The RNA velocity measurement can predict the future state of individual cells on a timescale of hours. Because ultra-short metabolic labeling (5 minutes of 4sU pulse) can be used to identify transient RNAs[22] and scNT-Seq enables direct measurements of new and old transcripts from the same cell, we

next sought to augment the RNA velocity model to predict rapid cell state changes by replacing unspliced/spliced read levels with UMI-based newly-transcribed and total RNA levels. We focused on the excitatory neurons as this population of neurons robustly responds to neuronal activation. To compare unspliced/spliced ratio-based RNA velocity (termed “splicing RNA velocity”) and newly-transcribed RNA fraction-based velocity (termed “new RNA velocity”) directly, we projected and aligned both velocity fields onto the same UMAP plot (**Fig. 3a**). While splicing RNA velocity showed a steady directional flow (arrows) from 60 to 120 min (left in **Fig. 3a**), new RNA velocity revealed two distinct phases of velocity flows (early phase: 15 to 30 min; late phase: 60 to 120 min) (right in **Fig. 3a**). Thus, our analyses are consistent with recent reports that these two types of velocity measurements convey different but complementary information on the future state of a cell[8, 21].

New RNA velocity is grounded in direct measurements of newly-transcribed RNAs, and this approach promises to improve quantitative analysis of the dynamics of cell states, particularly for the analysis of transient and dynamic responses to rapid external stimulation. Indeed, new RNA velocity (area under the receiver operating characteristic curve (AUC) value = 82.5%) outperformed splicing RNA velocity (AUC=51.5%) analysis in predicting the activity-induced primary response genes (PRGs, 137 genes) that were identified in targeted bulk RNA-seq analysis of primary cortical neurons[16] (**Fig. 3b**). These results show that new RNA velocity is more reliable in revealing the rapid transcriptional changes following acute neuronal activation.

To further explore the molecular basis underlying the two distinct phases of new RNA velocity flows, we identified the TF regulons that are significantly associated with early- (n=24; induced at 15 or 30 min) or late-response (n=73; induced at 60 or 120 min) genes (**Fig. 3c**, **Supplementary Fig. 4c**). We further projected aggregated newly-transcribed RNA levels of early- and late-response genes onto respective RNA velocity fields to visualize the relationship between distinct activity-patterns and ARG induction (**Fig. 3d**). Next, we projected TF regulon activity onto new RNA velocity field, revealing distinct activity patterns of TFs primarily associated with either early- (Jun) or late- (Mef2d) response genes in dynamic cell state changes (**Fig. 3e**). Together, these results indicate that integrative analysis of new RNA velocity and TF regulon activity can be a powerful approach to reveal molecular insights into dynamic cellular processes.

scNT-Seq captures immediate transcription changes during the pluripotent-to-totipotent stem cell state transition.

Determining temporal RNA dynamics such as immediate transcriptional events within rare, transient cell populations is critical to understanding cell state transition but has remained a significant challenge for stem cell research. Recent advances in multiplex single-molecule imaging using sequential fluorescence *in situ* hybridization (seqFISH) and intron probes allows for the detection of nascent transcription sites at the single-cell level[23]; however, such imaging approaches are currently time-consuming and have not been used to study rare stem cell populations at scale. Using mESC cultures as a model, we sought to

determine whether scNT-Seq could be used to directly investigate immediate transcription changes and cell state transition within rare stem cell populations.

Cultured mESCs are derived from the inner cell mass of pre-implantation blastocysts and exhibit a high level of transcriptional heterogeneity[24]. Interestingly, cells resembling 2-cell-stage embryos (2C-like cells), which are in a totipotent state and can differentiate into both embryonic and extraembryonic tissues, arise spontaneously in mESC cultures[25]. Nearly all mESCs reversibly cycle between pluripotent and totipotent 2C-like state in culture. However, the 2C-like cells are rare in mESC cultures (<1% in standard culture conditions)[25], making it challenging to study these totipotent cells with traditional methods. Recent studies using scRNA-Seq revealed changes in total RNAs during the transition to rare 2C-like state[26, 27] and identified an intermediate state during this low-frequency transition process, supporting a stepwise model of transcriptional reprogramming of the pluripotent to 2C-like state transition[26, 27]. However, these studies cannot distinguish newly-transcribed from pre-existing RNAs at each state. Thus, a more detailed understanding of direct gene regulatory changes in response to state transition is lacking.

To capture newly-transcribed RNAs and characterize GRNs in different states, mESCs were metabolically labeled with 4sU for 4 hours and were subjected to scNT-Seq analysis (**Fig. 4a**). After quality filtering, we obtained 4,633 single-cell transcriptomes from two biological replicates (**Supplementary Fig. 6a, b**). Unbiased clustering not only readily separated contaminating mouse feeder cells (*Col1a2*/*Thbs1*+) from mESCs, but also identified three principal states within mESCs (**Supplementary Fig. 6a-c**). The majority of mESCs (98.3%) belong to the pluripotent state as they are positive for pluripotency-genes such as *Sox2* and *Zfp42* (also known as Rex1) but are negative for 2C-embryo-specific transcripts such as the *Zscan4* gene family (**Fig. 4b**). Approximately 0.7% of cells express high levels of 2C-like state markers, including *Zscan4* genes, *Zfp352* and *Usp17lc*, but low levels of *Sox2*, suggesting that this rare population contains predominantly totipotent 2C-like cells. These results are consistent with previous FACS analysis of 2C-reporter lines[25, 28], indicating that scNT-Seq accurately captures this rare cell state. Interestingly, we also identified an intermediate state (~1.0%) which expresses low levels of *Zscan4* genes, but high levels of *Nanog* and *Tbx3* (**Supplementary Fig. 6c, d**). Unlike previous studies which overexpressed a master regulator, Dux, to induce 2C-like state[27], we did not observe a significant decrease in *Sox2* expression in the intermediate state (**Supplementary Fig. 6c**), highlighting the potential difference between the Dux-induced 2C-transitions and the naturally occurring pluripotent-to-totipotent transition process.

Comparative analysis of new and old transcriptomes revealed many state-specific genes (e.g., *Sox2* and *Zscan4d*) are associated with a higher proportion of new transcripts (consistent with their regulatory roles), whereas house-keeping genes, such as *Gapdh*, are of slower turnover rate (**Supplemental Fig. 7a, b**). For many state-specific genes (e.g., *Zfp42* and *Zscan4d*), their new RNA levels exhibited a more pronounced difference between states than the change of pre-existing RNAs (**Fig. 4b, c**). Gene Ontology (GO) enrichment analysis further revealed that state-specific new and old transcripts were enriched for different GO terms (**Fig. 4d**). For instance, many pluripotent-state-specific new RNAs (but not old RNAs)

are functionally related to post-transcriptional regulatory processes, including translational initiation, mRNA processing, and RNA splicing. Furthermore, 2C-state-specific new transcripts are preferentially enriched for deubiquitinating enzyme-related genes such as *Usp17lc/d/e* (**Fig. 4b, d**). These results underscore that newly-synthesized RNAs are more robust than steady-state transcripts to uncover genes related to state-specific biological processes within heterogeneous stem cell populations.

DNA-binding proteins, such as TFs and epigenetic regulators, show rapid gene expression changes during pluripotent-to-2C transition[27, 29]. Our analysis of immediate transcription changes during the pluripotent-to-2C transition identified 26 genes encoding DNA-binding proteins that showed a significant difference at new RNA levels between states (**Fig. 4e**). Consistent with previous scRNA-Seq analysis of sorted cell populations[29], 13 out of 14 commonly detected genes (*Tet1*, *Sox2*, *Nanog*, *Rex1*/*Zfp42*, *Sp110*, *Zfp352,* and *Zscan4* family) showed similar patterns in our scNT-Seq analysis of new RNAs (**Fig. 4e**). To further investigate TF/epigenetic regulator activity during the pluripotent-to-2C transition, we applied SCENIC analysis to both new and old transcriptomes. Among all regulons uncovered in mESCs (**Supplementary Fig. 8a**), we identified 25 TF/epigenetic regulators showing differential new regulon activity between states (**Fig. 4f** and **Supplementary Fig. 8b**). The activity of several positive regulators of the cell cycle, such as the E2F family of TFs, decreased in 2C-like states, which is consistent with previous observations that 2C-like cells are associated with longer cell cycles[30]. Interesting, new RNA levels and regulon activity of Nanog peak in intermediate state compared to pluripotent and totipotent 2C-like states, suggesting that this TF may play an unrecognized regulatory role in intermediate state. In addition, several epigenetic regulators, including Phf8, Hdac2, Ezh2, and Tet1, were associated with a decrease in regulon activity (**Fig. 4f**), suggesting a role of these enzymes in promoting the pluripotent state. Notably, we did not identify the regulon activity of Zscan4 in 2C-like cells, potentially due to the lack of Zscan4 motif information in the SCENIC TF database. Together, scNT-Seq directly captures state-specific newly-transcribed RNAs in rare totipotent cells in heterogeneous mESC cultures and allows analysis of dynamics of new regulon activity during state transitions.

TET-dependent regulation of the stepwise pluripotent-to-2C transition

The TET family of proteins (Tet1-3) are DNA dioxygenases that mediate active DNA demethylation at many *cis*-regulatory elements such as promoters and distal enhancers[31, 32], thereby regulating gene expression. During pluripotent-to-2C transition, both new RNA level and regulon activity of *Tet1* rapidly decreased (**Fig. 4c, f**). The new transcript level of *Tet2* also decreased in both intermediate and 2C-like states, while *Tet3* was undetected in all three states (**Supplementary Fig. 7c**). Interestingly, genetic inactivation of all TET proteins in mESCs resulted in a substantially higher proportion of 2C-like cells, and elevated expression level of 2C-specific transcripts[28].

To better understand the role of TET enzymes in state transition, we generated mESCs deficient for all three Tet proteins (*Tet1/2/3* triple knockout, *Tet*-TKO) via CRISPR/Cas9 genome editing using previously-tested sgRNAs[33] and analyzed isogenic WT and *Tet*-TKO mESCs (J1 strain) by scNT-Seq after 4 hours of metabolic labeling. The genotype of *Tet*-TKO cells was confirmed by both Sanger sequencing and reads from scNT-Seq (**Supplementary Fig. 9a**). Clustering and new RNA velocity analyses of combined WT and *Tet*-TKO mESCs revealed two discrete phases of new RNA velocity flows (arrows). The velocity flow within the pluripotent state is potentially driven by cell-cycle progression (**Supplementary Fig. 9b, c**), whereas new RNA velocity field of intermediate/2C-like states exhibited a strong directional flow from intermediate to 2C-like state (**Fig. 5a, b**). These observations are consistent with a stepwise model of the state-transition process and suggest that the pluripotent-to-intermediate state transition is a rate-limiting step.

WT and *Tet*-TKO cells were separately clustered within the pluripotent state and *Tet*-TKO (but not WT) cells were found immediately next to the intermediate/2C-like states (**Fig. 5a, b**), suggesting that *Tet*-TKO cells are transcriptionally reprogrammed to adopt a more poised state to transition to intermediate/2C-like states. Compared to WT cells, *Tet*-TKO cells exhibited a 2.2-fold or 3.6-fold increase in intermediate or 2C-like states (**Fig. 5c**), respectively, indicating that Tet enzymes negatively regulate the transition from pluripotent to intermediate/2C-like state in WT cells. Because pluripotent and 2C-like cells are globally distinct in levels of both major histone modifications[25] and DNA methylation[26], we next examined global levels of newly-synthesized transcripts in each state. Interestingly, aggregating new transcriptomes within specific states revealed that a totipotent 2C-like state is associated with markedly lower levels of global RNA biogenesis compared to pluripotent and intermediate states in WT cells (**Fig. 5d**). In contrast, *Tet*-TKO cells already exhibited a substantially lower level of global transcription than WT cells in pluripotent state (**Fig. 5d**), suggesting that Tet proteins may act as an epigenetic barrier for the pluripotent-to-2C transition by maintaining a pluripotent state-specific RNA biogenesis profile.

Pulse-chase scNT-Seq enables transcriptome-wide measurement of mRNA stability in rare 2C-like cells.

Pulse-chase assays combined with bulk SLAM-Seq have been used to measure mRNA stability in mESCs[5], but rare intermediate and 2C-like cells have not been studied. Given the enhanced sensitivity of scNT-Seq compared to bulk assays, combining a pulse-chase strategy with scNT-Seq may enable transcriptome-wide measurement of mRNA stability in rare cell populations. To test this, we metabolically labeled mESC cultures with 4sU for 24 hours (pulse), followed by a washout and chase using medium containing a higher concentration of uridine. Cells from multiple chase time-points were harvested and cryo-preserved first, and all samples were then re-hydrated and simultaneously analyzed by scNT-Seq analysis. After computing the proportion of labeled transcripts for each gene at every time-point relative to 0 h (right after pulse), we calculated the half-life ($t_{1/2}$) of mRNAs by fitting a single-exponential decay

model in each cell state (**Fig. 6a**). In total, 20,190 cells were profiled by scNT-Seq from 7 time-points (**Fig. 6b**). After 24 hours of metabolic labeling, we observed a substantial accumulation of T-to-C substitution (**Fig. 6c**), which is consistent with bulk assay observations[5]. The T-to-C substitution rate decreased over time and returned to the baseline level 24 hours after chase (**Fig. 6c**). Inspection of two genes, *Sox2* and *Topa2a*, revealed that total RNA levels did not change across the 24-hour time course, whereas the level of metabolically labeled transcripts decreased over time (**Supplementary Fig. 10a**), confirming that this strategy can accurately measure mRNA decay in a transcript-specific manner.

Based on expression of marker genes, clustering analysis readily separated mESCs into three states (n=7, pluripotent: 97.4%+/-0.78%, intermediate: 1.5%+/-0.48%, and 2C-like: 1.1%+/-0.38%) (**Fig. 6b** and **Supplementary Fig 10b**), suggesting that this pulse-chase approach did not significantly alter the state transition. By filtering out genes expressed in less than 5% of cells, we were able to determine the half-life of 2,310 genes in pluripotent state, and the RNA half-life determined by pulse-chase scNT-Seq is highly concordant with previous observation derived from bulk SLAM-Seq assays[5] (R^2=0.81, **Fig. 6d**). RNA stability can also be estimated by assuming simple exponential kinetics, which is calculated by measuring the ratio of labeled and unlabeled transcripts after metabolic labeling for a specific time[34]. However, RNA half-life estimated from one timepoint labeling (4sU, 4 hour) experiment is substantially less correlated with measurements derived from bulk assays (R^2=0.49, **Fig. 6d**). These results suggest that a pulse-chase strategy may more reliably measure RNA decay rate. Furthermore, the top 10% most stable and unstable transcripts were enriched for similar GO terms that are uncovered by bulk SLAM-Seq assays[5] (**Supplementary Fig. 10c**). Further analysis of cells in intermediate and 2C-like states revealed the half-life of 1,743 and 821 transcripts in these rare cell states, respectively. Next, we analyzed commonly detected transcripts between cell states to reveal state-specific regulation of mRNA stability (**Supplementary Fig. 11**). Thus, scNT-seq enables transcriptome-wide measurement of RNA stability in rare stem cell populations in heterogeneous mESC cultures.

Because dynamic changes in RNA levels are regulated by the interplay of RNA biogenesis and degradation, we sought to investigate the relationship between mRNA stability and gene expression during the pluripotent-to-2C transition (**Supplementary Fig. 12**). Consistent with previous findings[2], RNA stability and total mRNA levels are not highly correlated, suggesting mRNA stability is not the major contributor to total RNA level during stem cell state transition. Nevertheless, we identified a group of genes showing coordinated changes of RNA stability and gene expression level between pluripotent and 2C-like states, suggesting mRNA stability may play a role in regulating total mRNA level for a subset of genes (**Fig. 6e**). Further analysis showed that six RNA-binding protein (RBP) binding motifs were enriched in 3'UTR of 15 genes showing coordinated changes (**Fig. 6f**). Among these RBPs, the Pabpc family proteins (Pabpc1 and Pabpc4) are known regulators of mRNA stability and translation efficiency by binding poly-A tails of mRNAs[35]. Collectively, these results indicate that scNT-Seq can serve as a

powerful approach to investigate RNA stability and post-transcriptional regulatory mechanisms in rare cells within a heterogeneous population of cells.

DISCUSSION

By combining the widely accessible Drop-Seq platform and TimeLapse chemistry, scNT-Seq provides a novel strategy for high-throughput analysis of new and old transcriptomes from the same cell. A major advantage of scNT-Seq is its ability to accurately estimate newly-transcribed RNA levels using UMI-based statistical correction, which not only reveals acute transcriptional changes that are not apparent in standard scRNA-seq, but also allows for the analysis of dynamic TF regulatory networks and single-cell transcriptional trajectory. Our scNT-Seq is conceptually similar to sci-fate[36], a recently developed method that integrates SLAM-Seq chemistry with single-cell combinatorial indexing RNA-Seq. While scNT-Seq directly captures whole-cell total RNAs on barcoded beads in nanoliter droplets, sci-fate requires paraformaldehyde fixation and permeabilization of cells for *in situ* chemical conversion and reverse transcription, which may pose a challenge to unbiasedly recover both nuclear (enriched for newly-transcribed) and cytoplasmic (enriched for pre-existing) RNAs for accurately estimating new RNA fractions.

Because the purine analog, 6-thioguanine (6tG), enables G-to-A conversions by TimeLapse chemistry in bulk RNA-Seq[37], dual labeling of 4sU and 6tG followed by scNT-Seq may allow two independent recordings of newly-synthesized transcriptome at single-cell levels. With new computational approaches such as Dynamo[38], which can take advantage of controlled metabolic labeling to predict past and future cell states over an extended time period, high-throughput single-cell transcriptomic analysis of newly-transcribed and pre-existing RNAs can open new lines of inquiry regarding cell-type-specific RNA regulatory mechanisms and provide a broadly applicable strategy to investigate dynamic biological systems.

ACKNOWLEDGEMENTS

We are grateful to all members of the Wu lab for helpful discussion. This work was supported by the Penn Epigenetics Institute, the National Human Genome Research Institute (NHGRI) grants R00-HG007982 and R01-HG010646, National Heart Lung and Blood Institute (NHLBI) grant DP2-HL142044, and National Cancer Institute (NCI) grant U2C-CA233285 (to H.W.). The work of P.C. is partially supported by Stand-Up-To-Cancer Convergence 2.0.

AUTHOR CONTRIBUTIONS

H.W., Q.Q., and P.H. conceived of and developed the scNT-Seq approach. Q.Q. conducted most of the experiments. P.H. generated *Tet*1/2/3 TKO mESC line and performed most of the computational analysis. K.G. and P.C. contributed to the statistical modeling of T-to-C substitution. Q.Q., P.H., and H.W. analyzed the results and wrote the manuscript, with contributions from all of the authors. H.W. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing interest.

Online Methods

Experimental procedures

Mouse embryonic stem cell cultures and 4sU labeling

*Tet*1/2/3 triple-knock out (*Tet*-TKO) J1 mESCs were generated by CRISPR/Cas9 genome editing as previously described[33] and the genotype was confirmed by Sanger sequencing and scRNA-Seq. Wild-type (WT) and *Tet*-TKO J1 mESCs (ATCC, SCRC-1010) were initially cultured in presence of Mitomycin C inactivated mouse embryonic fibroblasts on 0.1% gelatin-coated (Millipore, ES-006-B) 6-well plates in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco, 11965084) supplemented with 15% fetal bovine serum (Gibco, 16000044), 0.1 mM nonessential amino acid (Gibco, 11140050), 1 mM sodium pyruvate (Gibco, 11360070), 2 mM L-glutamine (Gibco, 25030081), 50 μM 2-mercaptoethanol (Gibco, 31350010), 1 μM MEK inhibitor PD0325901 (Axon Med Chem, Axon 2128) and 3 μM GSK3 inhibitor CHIR99021 (Axon Med Chem, Axon 2128), and 1,000 U/mL LIF (Gemini Bio-Products, 400-495-7).

4-thiouridine (4sU) (Alfa Aesar, J60679) were dissolved in DMSO to make 1 M stock. Before 4sU labeling experiments, WT and *Tet*-TKO mESCs were passaged and cultured in feeder-free conditions (0.1% gelatin-coated plates) for 48 hrs. For 4sU labeling in WT and Tet TKO mESC cultures, the medium was replaced with fresh mESC medium supplemented with 100 μM 4sU. After 4 h, mESCs were rinsed once with PBS and dissociated with TrypLE-Express (Gibco, 12605010) for 5 min at 37°C. After neutralizing with culture medium, cells were pelleted at 1,000 rpm for 3 min. After cell counting with the Countess II system, the single cell suspension was diluted to 100 cells/uL with DPBS containing 0.01% BSA for scNT-Seq analysis.

Human K562 cell cultures and species mixing experiments

Human K562 cells (ATCC, CCL-243) were cultured in RPMI media supplemented with 10% FBS (Sigma, F6178). For species mixing experiments, the mESC or K562 media was replaced with media supplemented with 4sU (100 μM). After 4 h, the mESCs and K562 cells were rinsed with PBS and harvested for scNT-Seq analysis.

Mouse primary neuronal culture and stimulation

Mouse cortices were dissected from embryonic day 16 (E16) C57BL/6 embryos of mixed sex (Charles River). Cortical neurons were dissociated with papain (Worthington) and plated on 6-well plates (at a density of ~600,000 cells/well) coated with poly-ornithine (30mg/mL, Sigma, P2533). Mouse cortical neuronal cultures were maintained in neurobasal media (Gibco, 21103049) supplemented with B27 (Gibco, 17504044), 2 mM GlutaMAX (Gibco, 35050061), and 1X Penicillin/streptomycin (Gibco, 15140122).

After 4 days *in vitro* culture, primary cortical cultures were stimulated with a final concentration of 55mM potassium chloride (KCl) for various time (0, 15, 30, 60, and 120 minutes). Before neuronal activation, the fresh media supplemented with 4sU (200 μM) was added. After 4sU labeling (2 h), the cells were washed with PBS once and were digested with 0.05% Trypsin-EDTA (Gibco, 25300054) for 20 min at 37°C, then replaced the buffer with 1 mL of DPBS and dissociated cells with a cell-scraper. After cell counting with Countess II system, the single cell suspension was diluted to 100 cells/μL with DPBS containing 0.01% BSA for scNT-Seq analysis.

Cell fixation, cryopreservation and rehydration for scNT-Seq

The cell fixation was performed as previously described[39]. Cultured mESCs in 6-well plates were digested with TrypLE-Express and harvested as aforementioned. After washing once with DPBS, the cells were resuspended with 0.4 mL of DPBS containing 0.01% BSA. Split the cells to two 1.5 mL LoBind tubes (Eppendorf) and add 0.8 mL methanol dropwise at final concentration of 80% methanol in DPBS. After mixing and incubating the cell suspension on ice for 1 hour, store the fixed cells in LoBind tubes at -80°C

freezer for up to one month. For rehydration, cells were kept on ice after moved from –80 °C and kept in the cold throughout the procedure. After the cells were spun-down at 1,000 g for 5 min at 4°C, Methanol-PBS solution was removed and cells were resuspended in 1 mL 0.01% BSA in DPBS supplemented with 0.5% RNase-inhibitor (Lucigen, 30281-2). After cell counting with the Countess II system, the single cell suspension was diluted to 100 cells/µL and immediately used for scNT-Seq analysis.

Pulse-chase experiment for RNA stability analysis

Remove the medium from plates and add mESC medium supplemented with 200 µM 4sU[5]. The fresh medium with 4sU were changed every 4 h to enhance 4sU incorporation and mESCs were labeled for 24 hrs. After 4sU-labeling, the 4sU-containing medium was removed and cells were washed once with DPBS. Then mESC medium containing 10 mM Uridine (Sigma, U6381) was added to the culture before cells were harvested at different time-points ranging from 0.5 hour to 24 hours. Cells were fixed with methanol as aforementioned and store at -80°C for future use.

scNT-Seq library preparation and sequencing

Drop-Seq was performed as previously described with minor modifications[10]. Specifically, the single cell suspension was diluted to a concentration of 100 cells/µL in DPBS containing 0.01% BSA. Approximately 1.5 mL of diluted single cell suspension was loaded for each scNT-Seq run. The single-nucleus suspension was then co-encapsulated with barcoded beads (ChemGenes) using an Aquapel-coated PDMS microfluidic device (uFluidix) connected to syringe pumps (KD Scientific) via polyethylene tubing with an inner diameter of 0.38 mm (Scientific Commodities). Barcoded beads were resuspended in lysis buffer (200 mM Tris-HCl pH8.0, 20 mM EDTA, 6% Ficoll PM-400 (GE Healthcare/Fisher Scientific), 0.2% Sarkosyl (Sigma-Aldrich), and 50 mM DTT (Fermentas; freshly made on the day of run) at a concentration of 120 beads/µL. The flow rates for cells and beads were set to 3,200 µL/hour, while QX200 droplet generation oil (Bio-rad) was run at 12,500 µL/h. A typical run lasts ~20 min.

Droplet breakage with Perfluoro-1-octanol (Sigma-Aldrich). After droplet breakage, the beads were treated with TimeLapse chemistry to convert 4sU to cytidine-analog[6]. Briefly, 50,000-100,000 beads were washed once with 450 µL washing buffer (1 mM EDTA, 100 mM sodium acetate (pH 5.2)), then the beads were resuspended with a mixture of TFEA (600 mM), EDTA (1 mM) and sodium acetate (pH 5.2, 100 mM) in water. A solution of 192 mM $NaIO_4$ was then added (final concentration: 10 mM) and incubated at 45°C for 1 hour with rotation. The beads were washed once with 1 mL TE, then incubated in 0.5 mL 1 X Reducing Buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA) at 37°C for 30 min with rotation, followed by washing once with 0.3 mL 2X RT-buffer.

Reverse transcription and exonuclease I treatment were performed, as previously described, with minor modifications[10]. Specifically, up to 120,000 beads, 200 µL of reverse transcription (RT) mix (1x Maxima RT buffer (ThermoFisher), 4% Ficoll PM-400, 1 mM dNTPs (Clontech), 1 U/µL RNase inhibitor, 2.5 µM Template Switch Oligo (TSO: AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG)[10], and 10 U/ µL Maxima H Minus Reverse Transcriptase (ThermoFisher)) were added. The RT reaction was incubated at room temperature for 30 minutes, followed by incubation at 42°C for 120 minutes. To determine an optimal number of PCR cycles for amplification of cDNA, an aliquot of 6,000 beads (corresponding to ~100 nuclei) was amplified by PCR in a volume of 50 µL (25 µL of 2x KAPA HiFi hotstart readymix (KAPA biosystems), 0.4 µL of 100 µM TSO-PCR primer (AAGCAGTGGTATCAACGCAGAGT[10], 24.6 µL of nuclease-free water) with the following thermal cycling parameter (95°C for 3 min; 4 cycles of 98°C for 20 sec, 65°C for 45 sec, 72°C for 3 min; 9 cycles of 98°C for 20 sec, 67°C for 45 sec, 72°C for 3 min; 72°C for 5 min, hold at 4°C). After two rounds of purification with 0.6x SPRISelect beads (Beckman Coulter), amplified cDNA was eluted with 10 µL of water. 10% of amplified cDNA was used to perform real-time PCR analysis (1 µL of purified cDNA, 0.2 µL of 25 µM TSO-PCR primer, 5 µL of 2x KAPA FAST qPCR readymix, and 3.8 µL of water) to determine the additional number of PCR cycles needed for optimal cDNA amplification (Applied Biosystems QuantStudio 7 Flex). We then prepared PCR reactions per total number of barcoded beads collected for each scNT-Seq run, adding 6,000 beads per PCR tube, and ran the aforementioned program to enrich the cDNA for 4 + 10 to 12 cycles. We then tagmented cDNA using the Nextera XT DNA sample preparation kit (Illumina, cat# FC-131-1096), starting with 550 pg of cDNA

pooled in equal amounts, from all PCR reactions for a given run. Following cDNA tagmentation, we further amplified the library with 12 enrichment cycles using the Illumina Nextera XT i7 primers along with the P5-TSO hybrid primer (AATGATACGGCGACCACCGAGATCTACACGCCTGTCCGCGGA AGCAGTGGTATCAACGCAGAGT*A*C)[10]. After quality control analysis using a Bioanalyzer (Agilent), libraries were sequenced on an Illumina NextSeq 500 instrument using the 75- or 150-cycle High Output v2 or v2.5 Kit (Illumina). We loaded the library at 2.0 pM and added Custom Read1 Primer (GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC) at 0.3 µM to position 7 of the reagent cartridge. Paired-end sequencing were performed on Illumina NextSeq 500 sequencer as described previously[11]. The sequencing configuration was 20 bp (Read1), 8 bp (Index1), and 60 or 130 bp (Read2).

SLAM-Seq reaction on barcoded Drop-Seq beads

After droplet breakage, the beads were washed once with $NaPO_4$ buffer with 30% DMSO (50 mM, pH 8.0), then incubate beads in 500 µL reaction-mix containing 10 mM IAA for either 15 min at 50°C (standard SLAM-Seq) or 1 hour at 45°C (modified condition)[5]. Stop reaction by adding 10 µL 1 M DTT (final concentration: 20 mM).

Bioinformatic analysis

Read mapping and quantification of labeled and unlabeled transcripts

Paired-end sequencing reads of scNT-Seq were processed as previous described[11] with some modifications. Briefly, each mRNA read (read2) was tagged with the cell barcode (bases 1 to 12 of read 1) and unique molecular identifier (UMI, bases 13 to 20 of read 1), trimmed of sequencing adaptors and poly-A sequences, and aligned using STAR v 2.5.2a to the mouse (mm10, Gencode release vM13), human genome (GRCh38, Gencode release v23), or a concatenation of the mouse and human (for the species mixing experiment) reference genome assembly. Both exonic and intronic reads mapped to predicted strands of annotated genes were retained for the downstream analysis. To qualify the labeled and unlabeled transcripts, uniquely mapped reads with mapping score > 10 were grouped by UMI barcodes in every cell and were used to determine the T > C substitution using sam2tsv[40], T > C substitutions with a base quality of Phred score > 27 were retained. For each experiment, locus with background T > C substitutions (detected in the sample without TFEA/$NaIO_4$ treatment) was determined and was excluded for T > C substitution identification. After background correction of T > C substitution, a UMI was defined as labeled if there is a T > C substitution in any one of the reads belongs to that particular UMI. To this end, every UMI will be assigned to labeled or unlabeled based on the existence of T > C substitution (**Figs 1d and 5d**). For each transcript, the total number of uniquely labeled and unlabeled UMI sequences were counted and finally were assembled into a matrix using gene name as rows and cell barcode as columns.

Cell type clustering

For mouse cortical neurons and RNA-decay experiment (**Figs 2b** and **6b**), the raw digital expression matrices of new and old UMI counts were added up and loaded into the R package Seurat[41] (v 2.3.4). For normalization, UMI counts for all cells were scaled by library size (total UMI counts), multiplied by 10,000 and transformed to log space. Only genes detected in >10 cells were retained. Cell with a relatively high percentage of UMIs mapped to mitochondrial genes were discarded (QC metrics in **Supplementary Table 1**). Moreover, cells with lower or higher detected genes were discarded (QC metrics in **Supplementary Table 1**). As a result, 20,547 cells of mouse cortical culture and 20,190 cells of mESC (in pulse-chase assay) were retained, respectively. The highly variable genes (HVGs) were identified using the function *FindVariableGenes* with the parameters: *x.low.cutoff* = .05,*y.cutoff* = .5 in Seurat, resulting in 2,290 HVGs of primary cortical culture sample and 2,165 HVGs of mESC sample. The expression level of highly variable genes in the cells was scaled and centered along each gene and was conducted to principal component analysis (PCA). The most significant 30 PCs were selected and used for 2-dimension reduction by uniform manifold approximation and projection[42] (UMAP), implemented by the

Seurat software with the default parameters. Clusters were identified using the function *FindCluster* in Seurat with the resolution parameter set to 1. To obtain high level of cell type classification, we merge the adjacent clusters in UMAP which highly expressed excitatory markers (*Neurod2* and *Neurod6*) and define it as "Ex" cluster in mouse cortical neurons, while several close clusters highly expressed *Sox2* were combined to "pluripotent" cluster in mESC Pulse-chase experiment. Cell type specific markers were identified using function *FindMarkers* in Seurat with wilcoxon rank sum test with default parameters.

Cell clustering will be affected both by genotype[28] and by cell type. To enable directly comparative analyses within cell types between WT and *Tet*1/2/3 triple-knock out (*Tet*-TKO) mESCs, we used Seurat 3 (v. 3.0.0.9000) which was demonstrated as an effective strategy to perform joint analyses[43] (**Supplementary Fig. 6a**). The raw digital expression matrices of new and old UMI counts were added up and loaded into the Seurat 3. For normalization, UMI counts for all cells were scaled by library size (total UMI counts), multiplied by 10,000 and transformed to log space. Only genes detected in >10 cells were retained. Cell with a relatively high percentage of UMIs mapped to mitochondrial genes (>=0.05) were discarded. Moreover, cells with fewer than 500 or more than 5,000 detected genes were discarded, resulting in 4,633 WT cells and 2,319 *Tet*-TKO cells. Top 2,000 HVGs were identified using the function *FindVariableFeatures* with "vst" method. Canonical correlation analysis (CCA) was used to identify common sources of variation between WT and *Tet*-TKO cells. The first 20 dimensions of the CCA was chosen to integrate the two datasets. After integration, the expression level of HVGs in the cells was scaled and centered along each gene and was conducted to PCA analysis. The 20 most significant PCs were selected and used for 2-dimension reduction by UMAP. Clusters were identified using the function *FindCluster* in Seurat with the resolution parameter set to 3. As above mentioned, adjacent clusters highly expressed *Sox2* were combined to "pluripotent" cluster.

Estimation of the fraction of newly-synthesized transcripts

We implemented a statistical modeling strategy as previously described with some modifications for UMI-based scNT-Seq analysis[6]. For each experiment, a binomial mixture model, was used to approximate the number of T-to-C substitutions y_i for each gene transcript i:

$$f(y_i|\theta, p, q) = \theta \,\mathrm{Binom}(y_i; p, n_i) + (1-\theta)\,\mathrm{Binom}(y_i; q, n_i)$$

In this expression, θ is the fraction of new transcripts in each experiment, p and q are the probabilities of a T-to-C substitution at each nucleotide for new and old transcripts, respectively, and n_i is the number of uridine nucleotides observed in the transcript i. A consensus sequence for each transcript is built by pooling reads with the same UMI barcode and taking the most frequent variant at each position. 10,000 UMIs were randomly sampled and the global substitution probabilities p and q were estimated based on the above mixture model. The model was fit by maximizing the likelihood function using the Nelder-Mead algorithm. The optimization was repeated ten times with random initialization values for θ, p, and q in the range [0,1], keeping the best fit with $\theta \in [0,1]$.

For mouse cortical neuronal culture datasets, besides Ex and Inh1, we respectively combined 4 inhibitory neuronal clusters (Inh 2-4 and Inh-NP) and 3 non-neuronal clusters (Ex-NP1, Ex-NP2 and RG) to obtain enough UMI for global parameters estimation at each time-point. By doing this, Ex, Inh, and 2 combined clusters were subjected to statistical modelling. For mESC data sets, we assumed that *Tet*-TKO will not affect 4sU utilization efficiency and thus combined WT and TKO data sets to estimate unified global parameters, *p* and *q*, for 3 cell states (pluripotent, intermediate and 2C-like). To this end, 20 sets of *p* and *q* (5 time-points X 4 combined clusters) were determined in cortical neuronal data sets and 3 sets of *p* and *q* (pluripotent, intermediate and 2C-like clusters) were estimated for mESC data sets. Once these global parameters were determined, they were used to estimate the fraction of new transcripts.

1. Computing aggregate new transcripts for cell clusters

For Figs 2C and S3d, we aggregate all the UMIs belongs to the same cell-type and estimate the fraction of new transcripts θ for each gene with more than 100 UMIs in that cell type at each time-point. The likelihood function for the mixture model above was maximized using the Brent algorithm with the constraint $\theta \in [10^{-8}, 1]$. The 95% confidence interval was calculated from the Hessian matrix, and θ

estimates for genes with a confidence interval greater than 0.2 were thrown out. The new transcript (N) was then estimated:

$$N = \theta(L + U)$$

Where θ is the fraction of newly-transcribed RNAs for a gene in a cell type, L is labeled transcripts of a gene, U is unlabeled transcripts of a gene. The pre-existing transcript was calculated by: $(1 - \theta)(L + U)$.

2. Computing new transcript for each individual cell

In theory, for every cell the fraction of new transcripts should be calculated for each gene, which requires adequate gene coverage of each individual cells as bulk RNA-seq analysis. However, for data sets generated by high-throughput droplet-based scRNA-Seq methods and shallow sequencing, it is not feasible to obtain sufficient coverage for every gene in individual cells. Moreover, modeling every gene in thousands of cells were computationally intensive. To obtain the single-cell level estimation of new transcripts in each cell, we introduced detection rate α to estimate the new transcript for each cell. Since 4sU incorporation is random but each cell may vary in 4sU incorporation and most genes have the highly similar detection rate α (**Supplementary Fig. 3d**), we used the strategy above to estimate the fraction of new transcripts for individual cell. The detection rate α was then computed by diving all the labeled transcripts of a cell by estimated new transcript of that cell. We discarded the cells without-range values ($\alpha > 1$). We computed the detection rate α for 88.3% (18,133/20,547) of mouse cortical cells, whereas 95.1% (6,609/6,952) of mESC cells were retained. The mean of detection rates α were 60% and 66% in cortex and mESC, respectively. For each individual gene, the new transcript was computed according to the formula:

$$N = \min\left(\frac{L}{\alpha}, (L + U)\right)$$

Where α is the T-to-C detection rate of a cell, L is the number of labeled transcripts of a gene in that cell, U is the number of unlabeled transcripts of a gene. The number of pre-existing transcripts was calculated by: $L + U - N$. The computed new and old transcripts were used for all downstream single-cell-based analysis, including stacked UMAP (**right panel of Fig. 2b** and **Supplementary Fig. 4a**), SCENIC (**Figs. 2d, 3c, 4f, Supplementary Fig. 5** and **Supplementary Fig. 8a**) and new RNA velocity analysis (**Figs. 3a** and **5b**).

Gene ontology (GO) enrichment analysis

GO enrichment analysis was performed as previously described[12]. To identify functional categories associated with defined gene lists, the GO annotations were downloaded from the Ensembl database. An enrichment analysis was performed via a hypergeometric test. The *P* value was calculated using the following formula:

$$\mathrm{P} = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the total number of background genes, n is the total number of selected genes, M is the number of genes annotated to a certain GO term, and i is the number of selected genes annotated to a certain GO term. *P* value was corrected by function *p.adjust* with false discovery rate (FDR) correction in R. GO terms with FDR<0.05 were considered enriched. All statistical calculations were performed in R.

For enrichment analysis of stable/unstable mRNAs (**Supplementary Fig.10c**), genes were ranked by the RNA half-life. Top 10% genes with longest half-lives were defined as stable genes, whereas top 10% genes with shortest half-life were considered as unstable. Then the stable and unstable gene sets were subjected into GO enrichment analysis. For **Fig. 4d**, genes showed >1.5-fold changes between pluripotent and 2C-like state were selected and subjected into GO enrichment analysis.

Identification of differentially expressed genes (DEGs)

Differential gene expression analysis of new transcripts between different time-points of neuronal activation (15, 30, 60 and 120 min) and control (0 min) was performed using the function *FindMarkers* in Seurat, using a Wilcoxon rank sum test. New transcripts with a fold-change of more than 1.5 and an adjust *P*-value less than 0.05/4 were considered to be differentially expressed (**Supplementary Fig. 4b**). Neuronal induction genes were defined if a new transcript was significantly increased in at least one time-point with KCl stimulation in any one of cell-types (**Fig. 2c and Supplementary Fig. 4b**). For comparison among 3 cell states within WT mESCs (**Fig. 4e)**, genes encoding DNA binding proteins with a fold-change of new transcript expression level more than 0.25 (log-scale) and an adjust *P*-value less than 0.2 were listed.

Estimation of RNA half-life

For each gene, we separately aggerate labeled and unlabeled UMI counts in each cell state (**Fig. 6** and **Supplementary Fig.11**). Then the fraction of labeled transcripts was calculated with summed labeled UMI counts divided by total UMI counts (labeled + unlabeled). The fraction of labeled transcripts was normalized to chase-onset (0 min). R function *nls* was used to perform curve fitting with the parameters setting: "*y ~ I(a*exp(-b*x)+c)*", "*start=list(a=1, b=1, c=1)*" and "*na.action=na.exclude*". We kept the fit with the goodness of $R^2 > 0.7$.

RNA velocity analysis

For standard RNA velocity analysis (splicing RNA velocity), we started with the bam files which were generated by the Drop-Seq computational analysis pipeline. The reads were demultiplexed using dropEst[44] (version 0.8.5) pipeline, using "-m -V -b -f -L eiEIBA" options to annotate bam files. The genome annotations (mm10, Gencode release vM13) were used to count spliced and unspliced molecules for each experiment. The python package scVelo[21, 45, 46] (https://github.com/theislab/scvelo, version 0.1.19) were employed to perform RNA velocity analysis. Default parameter settings were used, unless stated otherwise. After loading spliced and unspliced molecules to scVelo, for analysis of excitatory neuron data sets **(left panel of Fig 3a**), genes with less than 5 counts of spliced or unspliced molecules were filtered out. Spliced and unspliced counts of top 1,000 highly variable genes were KNN-imputed in a PCA reduced space with 10 components, using 15 neighbors and 'distances' mode. For new RNA velocity analysis (using the ratio of new over total transcripts), all the parameters are the same as splicing RNA velocity analysis except that we loaded new transcripts (as unspliced counts) and total transcript (as spliced counts) into scVelo (**right panel of Fig 3a**). For new RNA velocity analysis of mESCs (**Fig 5a**), top 3,000 highly variable genes were KNN-imputed in with 10 PCs, using 15 neighbors and 'distances' mode as well. The gene-specific velocities are obtained by fitting a ratio between new and total mRNA by function '*scv.tl.velocity*'. Finally, function '*scv.pl.velocity_embedding_stream*' was used to project the velocities onto UMAP.

To directly compare splicing and new RNA velocity, we asked which of them could best predict whether well-known activity induced genes upregulated with different durations of KCl stimulation in excitatory neurons. The list of 137 primary response genes was previously defined by bulk RNA-Seq[16]. We calculated the average induction fold of genes (expressed in >1% cells) at 30 min of KCl stimulation compared to 0 min (**Fig 3b**). For splicing RNA velocity, we used unspliced counts to calculate the induction fold while estimated new transcript were used in new RNA velocity analysis. Finally, *auc* function in R package pROC was conducted to calculate the area under receiver operating characteristic curve (AUC), which reflects the accuracy of predicting the known PRGs by gene induction fold computed in different method. The plots were generated by function *plot.roc* in pROC package[47].

Gene regulatory network (GRN) analysis by SCENIC

To assess the regulatory activity of transcript factors associated with different cell states or cell-types, we used SCENIC[18] (version 1.1.2.2) to perform gene regulatory network (GRN or regulon) analysis. Regulatory modules are identified by inferring co-expression between TFs and genes containing TF

binding motif in their promoters. We separate expression matrix into two parts based on the expression level of new and old transcripts, then combined them as inputs to SCENIC analysis, which enabled us to identify specific regulatory modules associated with either new or old transcriptomes from the same cell. Two gene-motif rankings, 10kb around the TSS and 500 bp upstream, were loaded from RcisTarget databases (mm9). Gene detected in > 1% of all the cells and listed in the gene-motif ranking databases were retained. To this end, 8,744 genes in cortical culture data and 9,388 mESC genes were subjected into the downstream analysis. Then GRNBoost, which was implemented in pySCENIC, was used to infer the co-expression modules and quantify the weight between TFs and target genes. Targets genes that did not show a positive correlation (> 0.03) in each TF-module were filtered out. SCENIC found 4,944 and 5,406 TF-modules in mouse cortical neuronal culture and mESC data sets, respectively. A *cis*-regulatory motif analysis on each of the TF-modules with RcisTarget revealed 277 and 325 regulons in cortical culture and mESC data, respectively. The top 1 percentile of the number of detected genes per cell was used to calculate the enrichment of each regulon in each cell. For **Figs. 2d, 4f, and Supplementary Fig. 8**, we computed the mean AUC of all cells belongs to defined groups, then scaled the mean AUC by function *scale* in R. R package pheatmap[48] was used to draw the heatmap.

For **Fig. 4f and Supplementary Fig. 8**, AUC values of TFs were obtained and then subjected to Wilcoxon rank sum test to access significance of the difference of TF activity. TFs with a fold-change of mean AUC values more than 1.5 and an adjust *P*-value (Bonferroni corrected) less than 0.05 were considered differentially regulated.

RNA binding protein motif enrichment analysis

For **Fig. 6f**, 3' UTR sequences of 552 genes were retrieved from Ensembl Biomart[49]. All the 3' UTR sequences of a gene were obtained to explore all possible binding site. The position weight matrix (PWM) of 188 mouse RNA binding protein motifs were downloaded from CISBP-RNA database[50] (http://cisbp-rna.ccbr.utoronto.ca/). FIMO[51] (version 5.0.5) motif scanning software in MEME Suite was used to search for the RNA binding motifs in the 3' UTR sequences. 3' UTR sequences with *P*- values smaller than the default threshold (0.0001) were considered to contain a binding motif. As a result, 533 genes were identified to have RNA binding motif in their 3' UTR. Hypergeometric test was performed in R by function *phyper*, which was used to access significance of RNA binding motif enrichment in defined gene sets. The pairs of RNA binding protein and targets were visualized in Cytoscape.

Data availability

All sequencing data associated with this study will be available through the Gene Expression Omnibus (GEO) database.

Code availability

The analysis source code underlying the final version of the paper will be available on GitHub repository (https://gibhub.com/wulabupenn/scNT-seq).

REFERENCES

1. Rabani, M. et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698-1710 (2014).
2. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**, 436-442 (2011).
3. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331-338 (2017).
4. Wissink, E.M., Vihervaara, A., Tippens, N.D. & Lis, J.T. Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* **20**, 705-723 (2019).
5. Herzog, V.A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nature methods* **14**, 1198-1204 (2017).
6. Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C. & Simon, M.D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* **15**, 221-225 (2018).
7. Riml, C. et al. Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. *Angew Chem Int Ed Engl* **56**, 13479-13483 (2017).
8. Erhard, F. et al. scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571**, 419-423 (2019).
9. Hendriks, G.J. et al. NASC-seq monitors RNA synthesis in single cells. *Nature communications* **10**, 3138 (2019).
10. Macosko, E.Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
11. Hu, P. et al. Dissecting Cell-Type Composition and Activity-Dependent Transcriptional State in Mammalian Brains by Massively Parallel Single-Nucleus RNA-Seq. *Molecular cell* **68**, 1006-1015 e1007 (2017).
12. Hu, P. et al. Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. *Genes Dev* **32**, 1344-1357 (2018).
13. Lake, B.B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
14. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature methods* **14**, 955-958 (2017).
15. Yap, E.L. & Greenberg, M.E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* **100**, 330-348 (2018).
16. Tyssowski, K.M. et al. Different Neuronal Activity Patterns Induce Different Gene Expression Programs. *Neuron* **98**, 530-546 e511 (2018).
17. Duffy, E.E., Schofield, J.A. & Simon, M.D. Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. *Wiley interdisciplinary reviews. RNA* **10**, e1513 (2019).
18. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).
19. Laclef, C. et al. The role of primary cilia in corpus callosum formation is mediated by production of the Gli3 repressor. *Hum Mol Genet* **24**, 4997-5014 (2015).
20. Machado, C.B. et al. Reconstruction of phrenic neuron identity in embryonic stem cell-derived motor neurons. *Development* **141**, 784-794 (2014).
21. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
22. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225-1228 (2016).

804 23. Shah, S. et al. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH.
805 *Cell* **174**, 363-376 e316 (2018).
806 24. Kolodziejczyk, A.A. et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular
807 Transcriptional Variation. *Cell stem cell* **17**, 471-485 (2015).
808 25. Macfarlan, T.S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus
809 activity. *Nature* **487**, 57-63 (2012).
810 26. Eckersley-Maslin, M.A. et al. MERVL/Zscan4 Network Activation Results in Transient Genome-
811 wide DNA Demethylation of mESCs. *Cell reports* **17**, 179-192 (2016).
812 27. Fu, X., Wu, X., Djekidel, M.N. & Zhang, Y. Myc and Dnmt1 impede the pluripotent to totipotent
813 state transition in embryonic stem cells. *Nature cell biology* **21**, 835-844 (2019).
814 28. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and
815 telomere elongation. *Genes Dev* **28**, 2103-2119 (2014).
816 29. Rodriguez-Terrones, D. et al. A molecular roadmap for the emergence of early-embryonic-like
817 cells in culture. *Nature genetics* **50**, 106-119 (2018).
818 30. Nakai-Futatsugi, Y. & Niwa, H. Zscan4 Is Activated after Telomere Shortening in Mouse
819 Embryonic Stem Cells. *Stem cell reports* **6**, 483-495 (2016).
820 31. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions.
821 *Cell* **156**, 45-68 (2014).
822 32. Pastor, W.A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA
823 demethylation and transcription. *Nature reviews. Molecular cell biology* **14**, 341-356 (2013).
824 33. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-
825 mediated genome engineering. *Cell* **153**, 910-918 (2013).
826 34. Russo, J., Heck, A.M., Wilusz, J. & Wilusz, C.J. Metabolic labeling and recovery of nascent RNA to
827 accurately quantify mRNA stability. *Methods* **120**, 39-48 (2017).
828 35. Yi, H. et al. PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and
829 Block Precocious Decay. *Molecular cell* **70**, 1081-1088 e1085 (2018).
830 36. Cao, J., Zhou, W., Steemers, F., Trapnell, C. & Shendure, J. Characterizing the temporal dynamics
831 of gene expression in single cells with sci-fate. *bioRxiv*, 666081 (2019).
832 37. Kiefer, L., Schofield, J.A. & Simon, M.D. Expanding the Nucleoside Recoding Toolkit: Revealing
833 RNA Population Dynamics with 6-Thioguanosine. *Journal of the American Chemical Society* **140**,
834 14567-14570 (2018).
835 38. Qiu, X. et al. Mapping Vector Field of Single Cells. *bioRxiv*, 696724 (2019).
836 39. Alles, J. et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC*
837 *biology* **15**, 44 (2017).
838 40. Lindenbaum, P. Jvarkit: java utilities for bioinformatics. (2015).
839 41. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic
840 data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411
841 (2018).
842 42. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for
843 dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).
844 43. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* (2019).
845 44. Petukhov, V. et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-
846 based single-cell RNA-seq experiments. *Genome Biology* **19**, 78 (2018).
847 45. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis.
848 *Genome Biology* **19**, 15 (2018).
849 46. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell
850 states through dynamical modeling. *bioRxiv*, 820936 (2019).

47. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
48. Kolde, R. & Kolde, M.R. Package 'pheatmap'. *R Package* **1** (2015).
49. Kinsella, R.J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011** (2011).
50. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172 (2013).
51. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).

FIGURE LEGENDS

Figure 1. Development and validation of scNT-Seq. (**a**) Overview of single-cell Nascent transcript Tagging sequencing (scNT-Seq). (**b**) Multi-species mixing experiment measures scNT-Seq specificity. Mouse ESCs and human K562 cells were mixed at 1:1 ratio after 4sU labeling (100 μM, 4 hrs). Scatterplot shows the number of transcripts (UMIs) mapped to mouse (y-axis) or human (x-axis) genome for each cell (dot). Cells with mostly mouse transcripts are labeled as mouse (red), while cells with mostly human transcripts are labeled as human (blue). Cells with a relatively high percentage of both mouse and human transcripts are labeled as mixed (green). (**c**) Transcriptome-wide nucleotide substitutions rates demonstrate specific T-to-C conversion in 4sU labeled K562 cells after TEFA/$NaIO_4$-treatment compared to untreated control. (**d**) Box plots showing percentage of T-to-C containing transcripts (UMIs) in individual K562 cells. The box plots display the median (center line) and interquartile range (IQR, from the 25th to 75th percentile), the whiskers represent 1.5 times the interquartile range, and the circles represent outliers. (**e**) All UMIs of the *ACTG1* gene are shown for one K562 cell with or one without TFEA/$NaIO_4$-treatment. Grey circles stand for T without T-to-C substitution, while crosses ("X"s) stand for sites of T-to-C substitution in at least one read. The color of "X" indicates the read coverage in a UMI containing T-to-C substitution at a particular site. All reads from one exemplary transcript (2nd UMI from TFEA/$NaIO_4$-treated sample) are further shown in the lower panel.

Figure 2. scNT-Seq captures cell-type-specific RNA dynamics and TF activity in response to distinct neuronal activity patterns. (**a**) Schematics of applying scNT-Seq analysis to study neuronal activation of primary mouse cortical cultures. (**b**) Uniform manifold approximation and projection (UMAP) visualization of 20,547 cells from primary mouse cortical cultures. Relative percentage of each cell-type is listed on the left. Randomly sampled 1,000 excitatory neurons from 5 time points (200 cells from each time point) are sub-clustered based on newly-synthesized (top) or pre-existing transcripts (bottom). Cells from each time point are color-coded and projected into the stacked UMAP on the right panel. The new and old transcriptomes of same cell were connected by a black line. Ex, excitatory neurons; Inh, inhibitory neurons; NP, neural progenitors; RG, radial glia. (**c**) Cell-type-specific responses of activity-regulated genes (ARGs) upon neuronal activation. The mean new and old RNA expression levels (transcripts per 10k, TP10k) were shown in line plot. (**d**) Heat map showing cell-type-specific TF regulon activity (AUC value quantified by SCENIC) in response to distinct activity durations. Red color highlights known regulators of ARGs, while blue color highlights potential novel regulators induced by neuronal activity.

Figure 3. scNT-Seq reveals neuronal activity-induced new RNA velocity. (**a**) RNA velocity analysis of excitatory neurons using ratios of unspliced-to-spliced transcripts (splicing RNA velocity, left panel) or new-to-total transcripts (new RNA velocity, right panel) as inputs. The cells are color-codeed by different neuronal activity durations. The arrow indicates the projection from the observed state to extrapolated future state. (**b**) Comparison between splicing and new RNA velocity. Spliced counts or new RNA counts (30 min KCl stimulation) were used to predict induction of 137 known primary response genes[16]. The area under the curve (AUC) values of predictions from these two methods are shown. (**c**) Enrichment analysis of TF regulons for early- or late-response genes. Bar plots showing the significance of enriched TF regulons by hypergeometric test. TF regulons that are enriched for both groups, only early- or only late-response genes are highlighted by three boxes (left to right). The cut off of significance was indicated by red dash line (adjusted *P*-value < 0.001). (**d**) UMAP plots showing clustered Ex neurons from five time-points (left), expression level of early-response (middle), and late-response genes (right). (**e**) UMAP plots showing the regulon activity of two representative TFs regulating early- (Jun) or late-response genes (Mef2d), respectively. The UMAP plots in Fig. 3d-e are the same as in Fig. 3a.

Figure 4. scNT-Seq captures both new and old transcripts in different states of mESCs. (**a**) Schematics of the scNT-Seq analysis of mESCs. (**b**) MA-plot showing differential gene expression of new and old transcripts between pluripotent and 2C-like states. Dashed line denotes 1.5-fold differences. (**c**) The line plots showing new and old transcript levels of *Tet1* and *Zscan4d* in pluripotent, intermediate, and 2C-like state. (**d**) Heat map showing enriched GO terms of state-specific genes. Significance of enrichment (FDR) is scaled by colors. (**e**) Heat map showing new transcript levels of state-specific genes

935 encoding DNA-binding proteins. (**f**) Heat map showing the regulon activity of state-specific TFs or
936 epigenetic regulators. The color scales depict the regulon activity.

937 **Figure 5. scNT-Seq reveals TET-dependent regulation of the plutipotent-to-2C transition**. (*a*) New
938 RNA velocity analysis of combined wild-type (WT) and *Tet*-TKO mESCs. The arrow indicates the
939 projection from the observed state to extrapolated future state. (**b**) Projection of cell state and the new
940 RNA level of the *Zscan4a* gene on the same UMAP plot as in Fig. 5a. (**c**) Relative composition of three
941 stem cell states in WT (4,633 cells) and *Tet*-TKO (2,319 cells) mESCs. (**d**) The fraction of labeled
942 transcripts in WT and *Tet*-TKO mESCs across three states. The box plots display the median (center line)
943 and interquartile range (IQR, from the 25th to 75th percentile), the whiskers represent 1.5 times the
944 interquartile range, and the circles represent outliers.

945 **Figure 6. Pulse-chase scNT-Seq reveals cell state-specific mRNA decay in mESCs.** (**a**) Schematics
946 of pulse-chase scNT-Seq experiment. (**b**) UMAP plots of 20,190 mESCs profiled in the pulse-chase
947 scNT-Seq analysis. Cells are colored by three states (left) or by 7 time-points (right). Cell numbers of
948 each state across different time-points are also shown. (**c**) Line plots showing temporal changes of
949 nucleotide substitution rates. (**d**) Scatterplots showing the correlation of RNA half-life measurements
950 between this study (top: one timepoint inference analysis; bottom: pulse-chase analysis) and bulk SLAM-
951 Seq[5]. (**e**) Heat map (left) and bar plot (right) showing concordant change of RNA half-life (left) and gene
952 expression (right) between pluripotent and 2C-like states, respectively. (**f**) Network of RNA-binding
953 proteins (pink quadrate) and their binding targets (green circle). Enriched RNA-binding proteins were
954 inferred by motif sequenced enriched in 3'UTR of target genes.

Figure 1

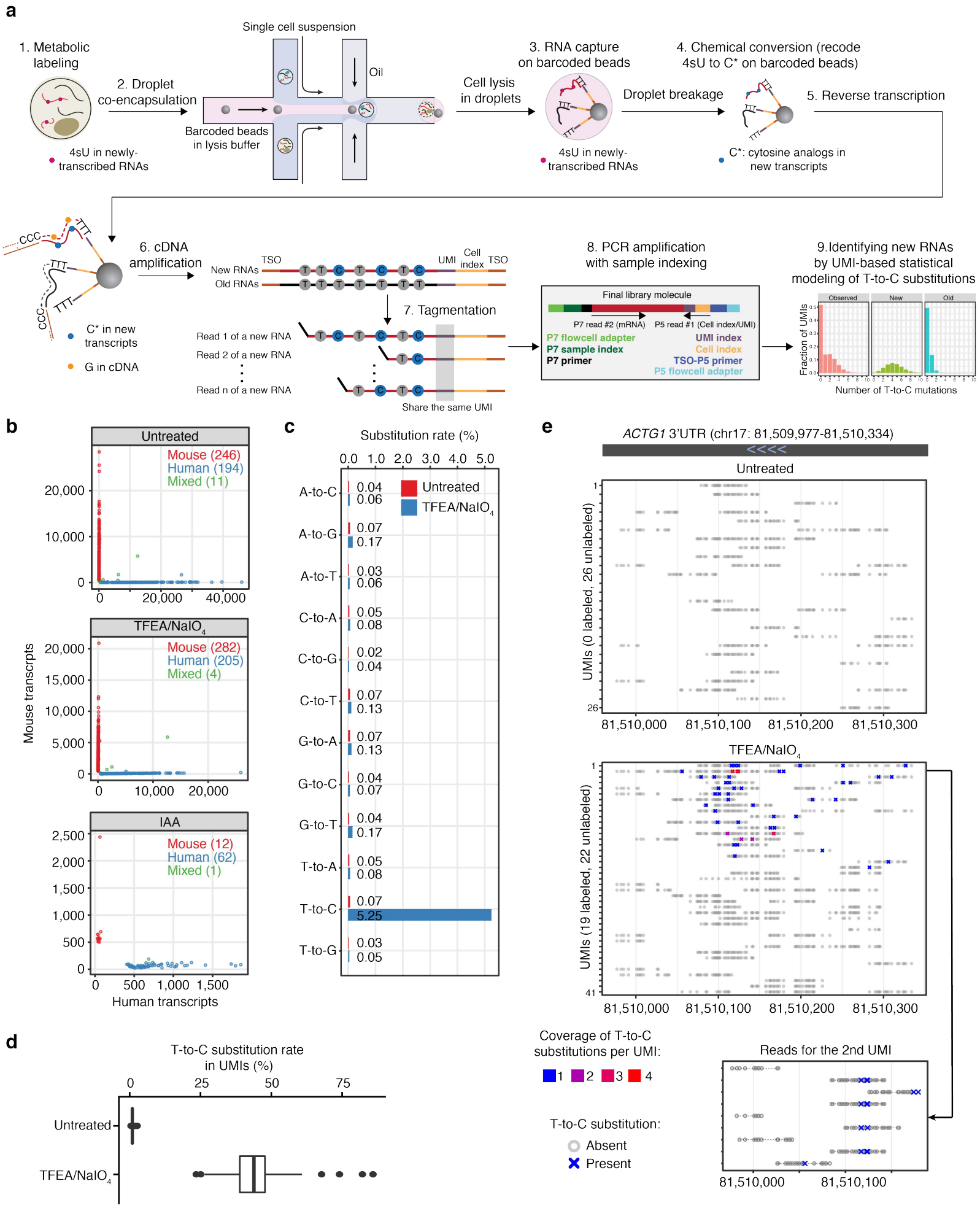

a
1. Metabolic labeling
4sU in newly-transcribed RNAs
2. Droplet co-encapsulation
Single cell suspension
Oil
Barcoded beads in lysis buffer
Cell lysis in droplets
3. RNA capture on barcoded beads
4sU in newly-transcribed RNAs
Droplet breakage
4. Chemical conversion (recode 4sU to C* on barcoded beads)
C*: cytosine analogs in new transcripts
5. Reverse transcription
CCC
C* in new transcripts
G in cDNA
6. cDNA amplification
TSO
New RNAs
Old RNAs
UMI
Cell index
TSO
7. Tagmentation
Read 1 of a new RNA
Read 2 of a new RNA
Read n of a new RNA
Share the same UMI
8. PCR amplification with sample indexing
Final library molecule
P7 read #2 (mRNA)
P5 read #1 (Cell index/UMI)
P7 flowcell adapter
P7 sample index
P7 primer
UMI index
Cell index
TSO-P5 primer
P5 flowcell adapter
9.Identifying new RNAs by UMI-based statistical modeling of T-to-C substitutions
Observed
New
Old
Fraction of UMIs
Number of T-to-C mutations
b
Untreated
Mouse (246)
Human (194)
Mixed (11)
TFEA/NaIO4
Mouse (282)
Human (205)
Mixed (4)
IAA
Mouse (12)
Human (62)
Mixed (1)
Mouse transcripts
Human transcripts
c
Substitution rate (%)
Untreated
TFEA/NaIO4
A-to-C 0.04 0.06
A-to-G 0.07 0.17
A-to-T 0.03 0.06
C-to-A 0.05 0.08
C-to-G 0.02 0.04
C-to-T 0.07 0.13
G-to-A 0.07 0.13
G-to-C 0.04 0.07
G-to-T 0.04 0.17
T-to-A 0.05 0.08
T-to-C 0.07 5.25
T-to-G 0.03 0.05
e
ACTG1 3'UTR (chr17: 81,509,977-81,510,334)
Untreated
UMIs (0 labeled, 26 unlabeled)
TFEA/NaIO4
UMIs (19 labeled, 22 unlabeled)
81,510,000
81,510,100
81,510,200
81,510,300
Coverage of T-to-C substitutions per UMI:
1 2 3 4
T-to-C substitution:
Absent
Present
Reads for the 2nd UMI
d
T-to-C substitution rate in UMIs (%)
Untreated
TFEA/NaIO4

Figure 2

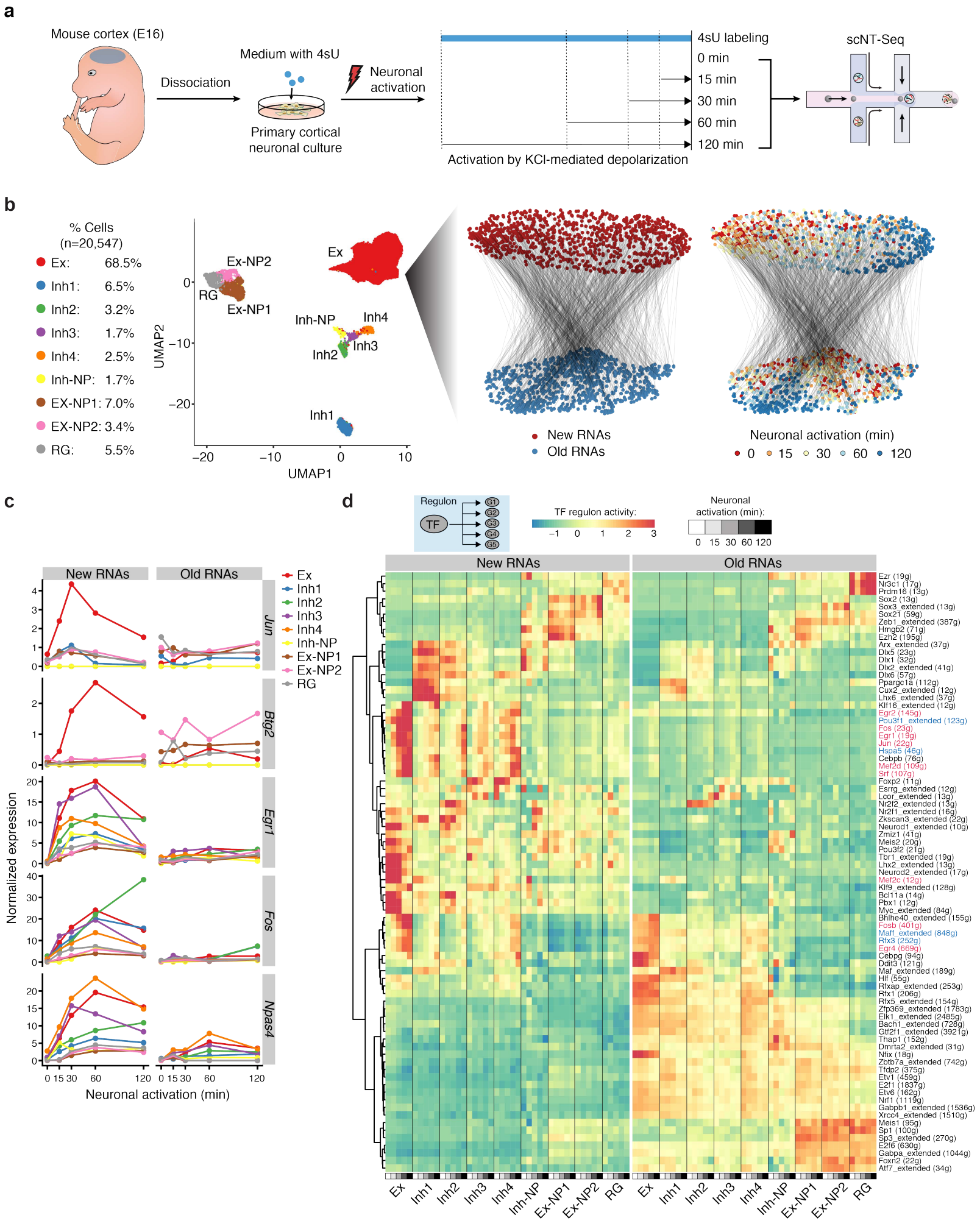

a
Mouse cortex (E16)
Dissociation
Medium with 4sU
Primary cortical neuronal culture
Neuronal activation
4sU labeling
0 min
15 min
30 min
60 min
120 min
Activation by KCl-mediated depolarization
scNT-Seq
b
% Cells (n=20,547)
Ex: 68.5%
Inh1: 6.5%
Inh2: 3.2%
Inh3: 1.7%
Inh4: 2.5%
Inh-NP: 1.7%
EX-NP1: 7.0%
EX-NP2: 3.4%
RG: 5.5%
UMAP2
UMAP1
New RNAs
Old RNAs
Neuronal activation (min)
0 15 30 60 120
c
New RNAs
Old RNAs
Ex
Inh1
Inh2
Inh3
Inh4
Inh-NP
Ex-NP1
Ex-NP2
RG
Jun
Btg2
Egr1
Fos
Npas4
Normalized expression
Neuronal activation (min)
d
Regulon
TF
G1
G2
G3
G4
G5
TF regulon activity:
Neuronal activation (min):
New RNAs
Old RNAs
Ezr (19g)
Nr3c1 (17g)
Prdm16 (13g)
Sox2 (13g)
Sox3_extended (13g)
Sox21 (59g)
Zeb1_extended (387g)
Hmgb2 (71g)
Ezh2 (195g)
Arx_extended (37g)
Dlx5 (23g)
Dlx1 (32g)
Dlx2_extended (41g)
Dlx6 (57g)
Ppargc1a (112g)
Cux2_extended (12g)
Lhx6_extended (37g)
Klf16_extended (12g)
Egr2 (145g)
Pou3f1_extended (123g)
Fos (23g)
Egr1 (19g)
Jun (22g)
Hspa5 (46g)
Cebpb (76g)
Mef2d (109g)
Srf (107g)
Foxp2 (11g)
Esrrg_extended (12g)
Lcor_extended (13g)
Nr2f2_extended (13g)
Nr2f1_extended (16g)
Zkscan3_extended (22g)
Neurod1_extended (10g)
Zmiz1 (41g)
Meis2 (20g)
Pou3f2 (21g)
Tbr1_extended (19g)
Lhx2_extended (13g)
Neurod2_extended (17g)
Mef2c (12g)
Klf9_extended (128g)
Bcl11a (14g)
Pbx1 (12g)
Myc_extended (84g)
Bhlhe40_extended (155g)
Fosb (401g)
Maff_extended (848g)
Rfx3 (252g)
Egr4 (669g)
Cebpg (94g)
Ddit3 (121g)
Maf_extended (189g)
Hlf (55g)
Rfxap_extended (253g)
Rfx1 (206g)
Rfx5_extended (154g)
Zfp369_extended (1783g)
Elk1_extended (2485g)
Bach1_extended (728g)
Gtf2f1_extended (392!g)
Thap1 (152g)
Dmrta2_extended (31g)
Nfix (18g)
Zbtb7a_extended (742g)
Tfdp2 (375g)
Etv1 (459g)
E2f1 (1837g)
Etv6 (162g)
Nrf1 (1119g)
Gabpb1_extended (1536g)
Xrcc4_extended (1510g)
Meis1 (95g)
Sp1 (100g)
Sp3_extended (270g)
E2f6 (630g)
Gabpa_extended (1044g)
Foxn2 (22g)
Atf7_extended (34g)
Ex
Inh1
Inh2
Inh3
Inh4
Inh-NP
Ex-NP1
Ex-NP2
RG

Figure 3

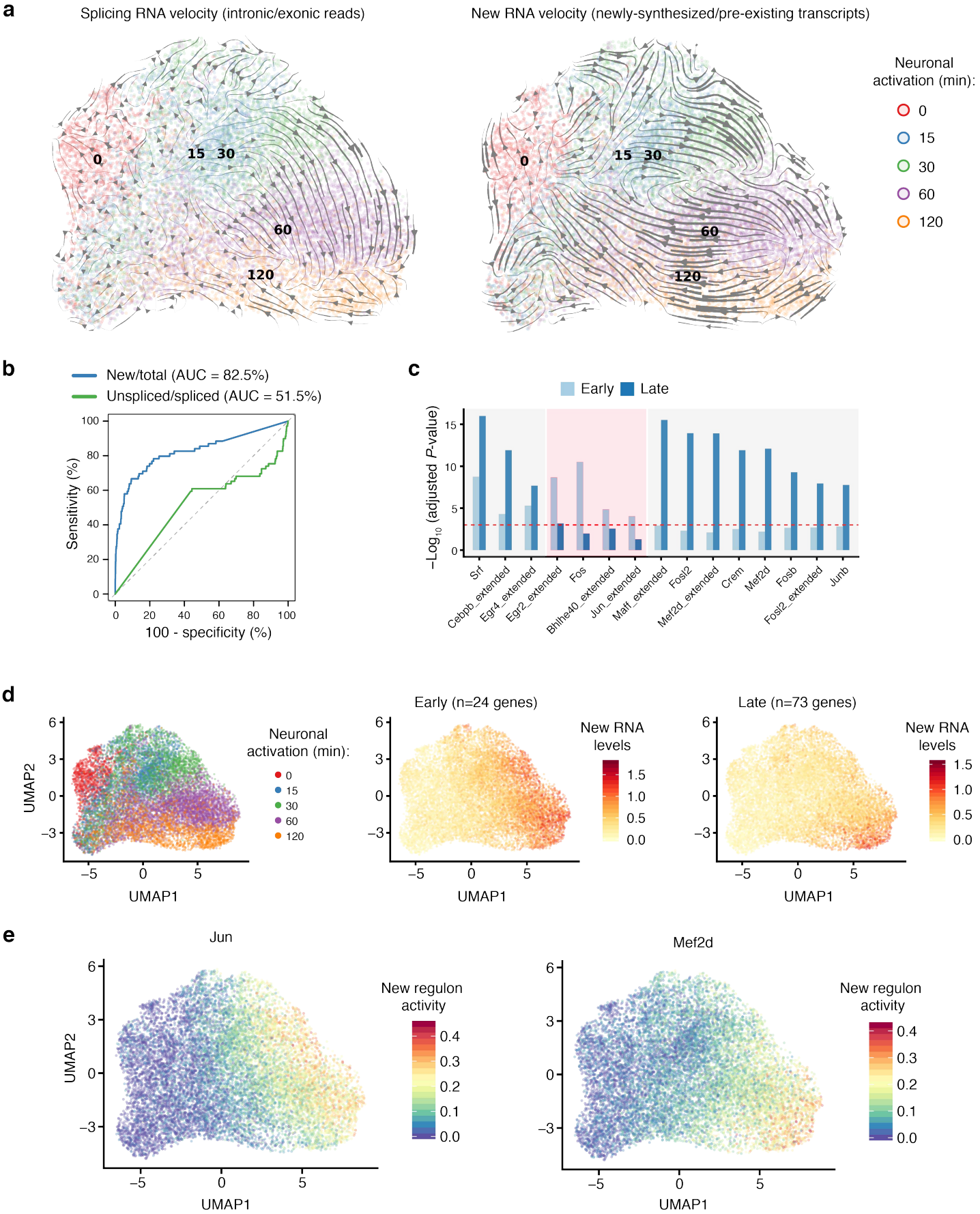

a
Splicing RNA velocity (intronic/exonic reads)
New RNA velocity (newly-synthesized/pre-existing transcripts)
0
15
30
60
120
Neuronal activation (min):
0
15
30
60
120
b
New/total (AUC = 82.5%)
Unspliced/spliced (AUC = 51.5%)
Sensitivity (%)
100 - specificity (%)
c
Early
Late
−Log10 (adjusted P-value)
Srf
Cebpb_extended
Egr4_extended
Egr2_extended
Fos
Bhlhe40_extended
Jun_extended
Maff_extended
Fosl2
Mef2d_extended
Crem
Mef2d
Fosb
Fosl2_extended
Junb
d
UMAP2
UMAP1
Neuronal activation (min):
0
15
30
60
120
Early (n=24 genes)
New RNA levels
Late (n=73 genes)
New RNA levels
e
Jun
Mef2d
New regulon activity
New regulon activity

Figure 4

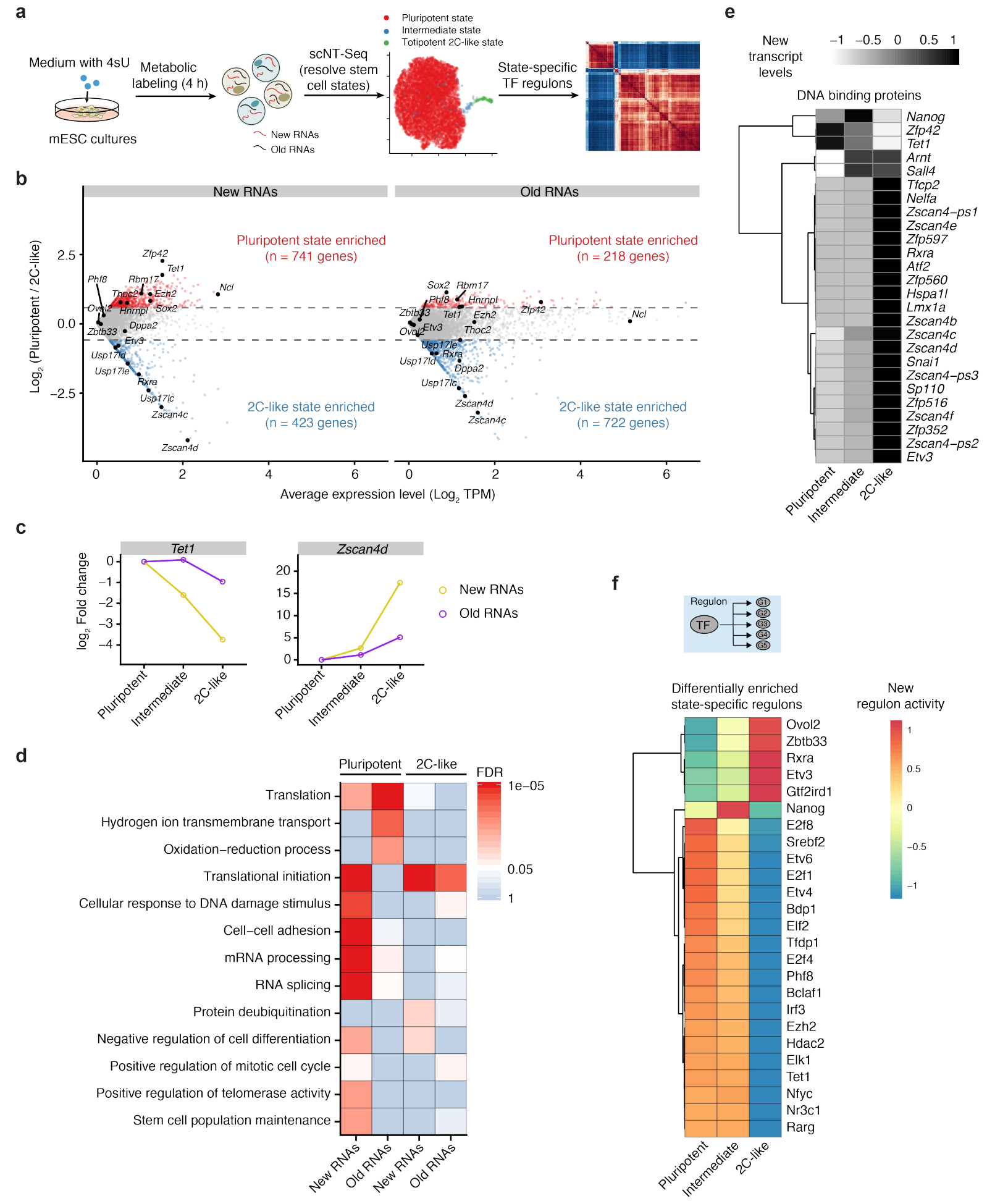

a
Medium with 4sU
mESC cultures
Metabolic labeling (4 h)
New RNAs
Old RNAs
scNT-Seq (resolve stem cell states)
Pluripotent state
Intermediate state
Totipotent 2C-like state
State-specific TF regulons
b
New RNAs
Old RNAs
Pluripotent state enriched (n = 741 genes)
Pluripotent state enriched (n = 218 genes)
2C-like state enriched (n = 423 genes)
2C-like state enriched (n = 722 genes)
Log2 (Pluripotent / 2C-like)
Average expression level (Log2 TPM)
Zfp42
Tet1
Phf8
Rbm17
Thoc2
Ezh2
Ncl
Ovol2
Hnrnpl
Sox2
Dppa2
Zbtb33
Etv3
Usp17ld
Usp17le
Rxra
Usp17lc
Zscan4c
Zscan4d
c
Tet1
Zscan4d
log2 Fold change
New RNAs
Old RNAs
Pluripotent
Intermediate
2C-like
d
Pluripotent
2C-like
FDR
1e−05
0.05
1
Translation
Hydrogen ion transmembrane transport
Oxidation−reduction process
Translational initiation
Cellular response to DNA damage stimulus
Cell−cell adhesion
mRNA processing
RNA splicing
Protein deubiquitination
Negative regulation of cell differentiation
Positive regulation of mitotic cell cycle
Positive regulation of telomerase activity
Stem cell population maintenance
New RNAs
Old RNAs
e
New transcript levels
−1 −0.5 0 0.5 1
DNA binding proteins
Nanog
Zfp42
Tet1
Arnt
Sall4
Tfcp2
Nelfa
Zscan4−ps1
Zscan4e
Zfp597
Rxra
Atf2
Zfp560
Hspa1l
Lmx1a
Zscan4b
Zscan4c
Zscan4d
Snai1
Zscan4−ps3
Sp110
Zfp516
Zscan4f
Zfp352
Zscan4−ps2
Etv3
Pluripotent
Intermediate
2C-like
f
Regulon
TF
G1
G2
G3
G4
G5
Differentially enriched state-specific regulons
New regulon activity
Ovol2
Zbtb33
Rxra
Etv3
Gtf2ird1
Nanog
E2f8
Srebf2
Etv6
E2f1
Etv4
Bdp1
Elf2
Tfdp1
E2f4
Phf8
Bclaf1
Irf3
Ezh2
Hdac2
Elk1
Tet1
Nfyc
Nr3c1
Rarg
Pluripotent
Intermediate
2C-like

Figure 5

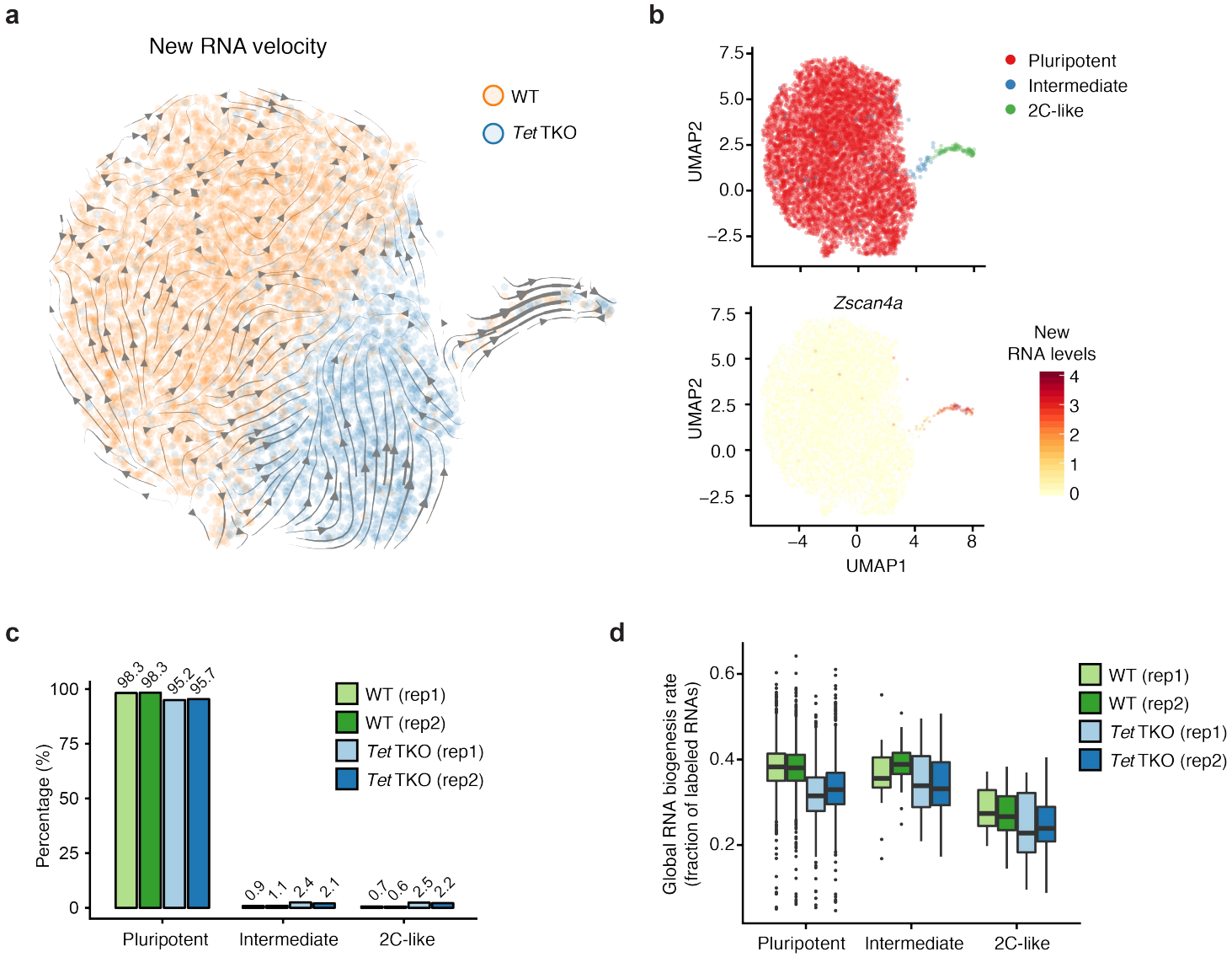
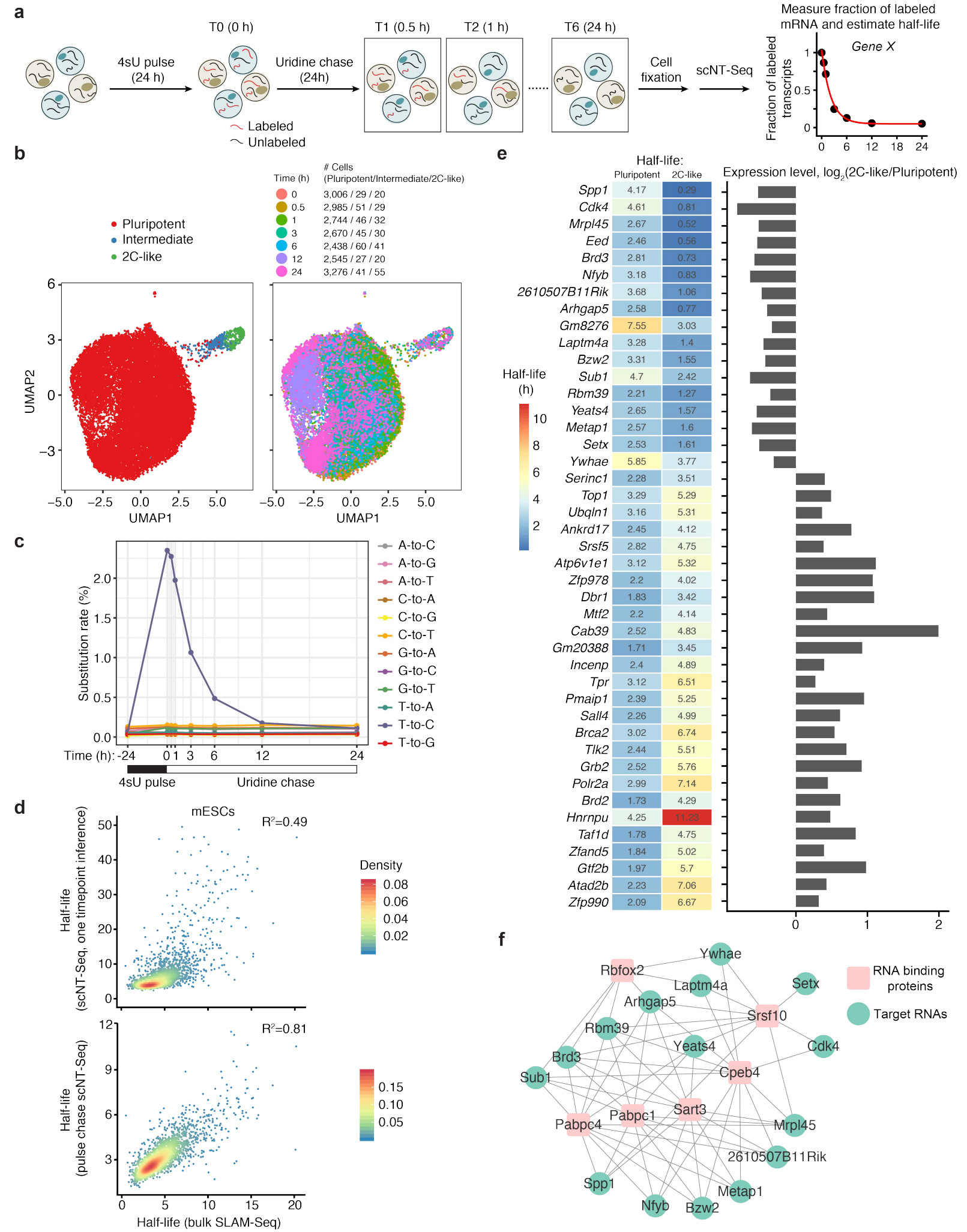
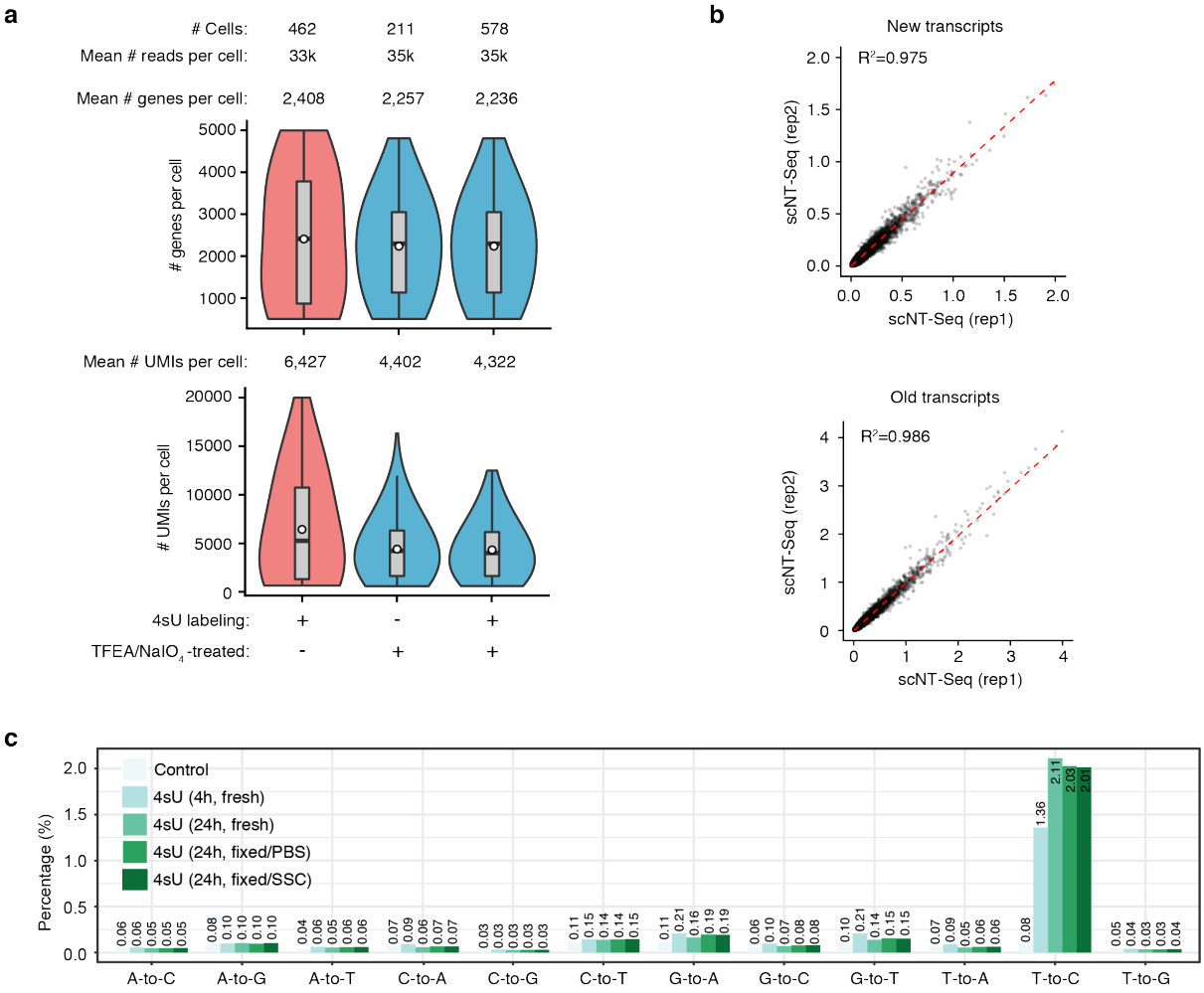

a
New RNA velocity
WT
Tet TKO
b
Pluripotent
Intermediate
2C-like
UMAP2
UMAP1
Zscan4a
New RNA levels
c
Percentage (%)
98.3
98.3
95.2
95.7
0.9
1.1
2.4
2.1
0.7
0.6
2.5
2.2
WT (rep1)
WT (rep2)
Tet TKO (rep1)
Tet TKO (rep2)
Pluripotent
Intermediate
2C-like
d
Global RNA biogenesis rate (fraction of labeled RNAs)
WT (rep1)
WT (rep2)
Tet TKO (rep1)
Tet TKO (rep2)
Pluripotent
Intermediate
2C-like

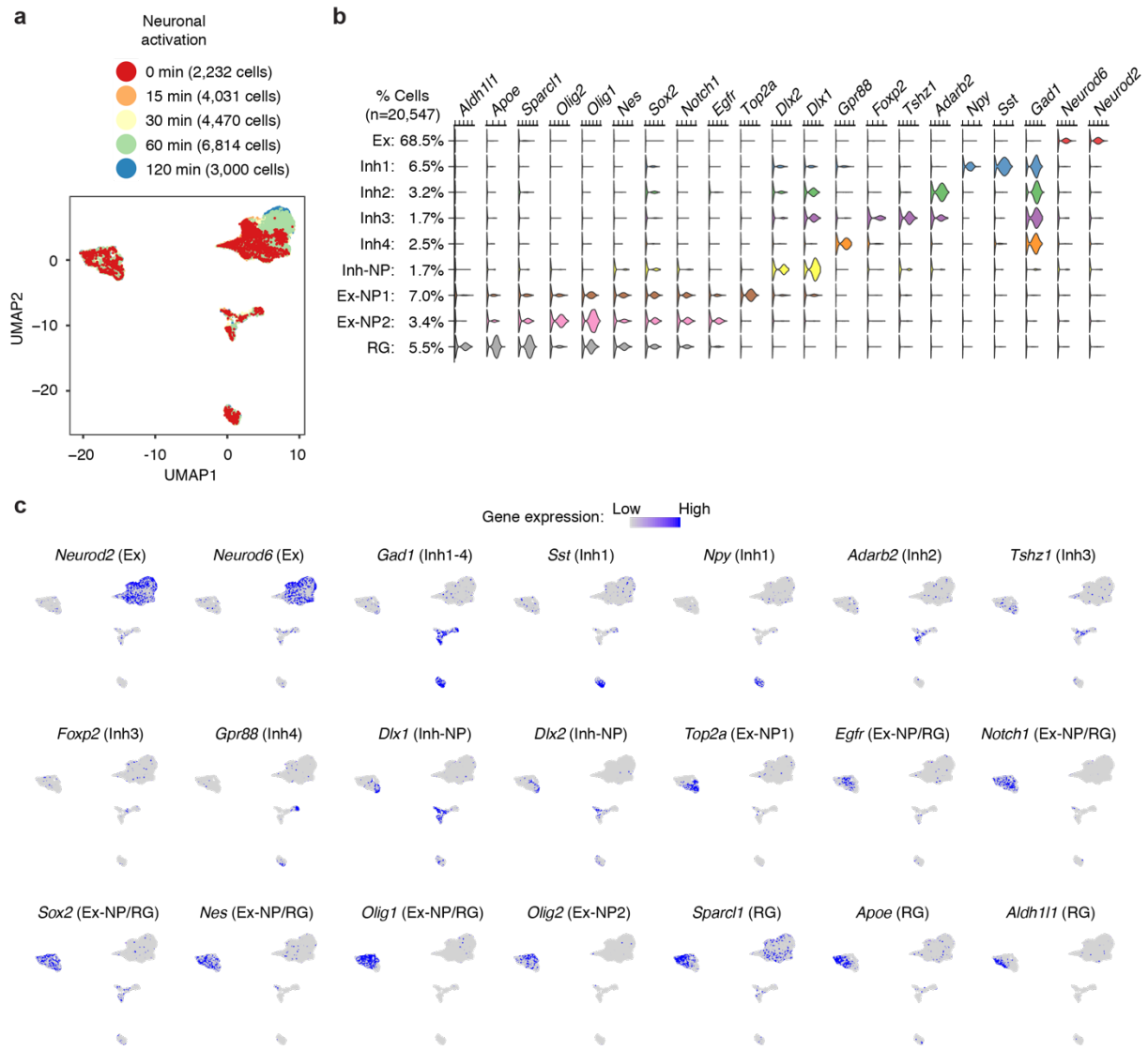
Figure 6


a
4sU pulse (24 h)
T0 (0 h)
Uridine chase (24h)
T1 (0.5 h)
T2 (1 h)
T6 (24 h)
Cell fixation
scNT-Seq
Labeled
Unlabeled
Measure fraction of labeled mRNA and estimate half-life
Gene X
Fraction of labeled transcripts
b
Pluripotent
Intermediate
2C-like
Time (h)
Cells (Pluripotent/Intermediate/2C-like)
0 3,006 / 29 / 20
0.5 2,985 / 51 / 29
1 2,744 / 46 / 32
3 2,670 / 45 / 30
6 2,438 / 60 / 41
12 2,545 / 27 / 20
24 3,276 / 41 / 55
UMAP1
UMAP2
c
Substitution rate (%)
Time (h): -24 0 1 3 6 12 24
4sU pulse
Uridine chase
A-to-C
A-to-G
A-to-T
C-to-A
C-to-G
C-to-T
G-to-A
G-to-C
G-to-T
T-to-A
T-to-C
T-to-G
d
mESCs
R²=0.49
R²=0.81
Half-life (scNT-Seq, one timepoint inference)
Half-life (pulse chase scNT-Seq)
Half-life (bulk SLAM-Seq)
Density
e
Half-life:
Pluripotent
2C-like
Expression level, log2(2C-like/Pluripotent)
Half-life (h)
Spp1 4.17 0.29
Cdk4 4.61 0.81
Mrpl45 2.67 0.52
Eed 2.46 0.56
Brd3 2.81 0.73
Nfyb 3.18 0.83
2610507B11Rik 3.68 1.06
Arhgap5 2.58 0.77
Gm8276 7.55 3.03
Laptm4a 3.28 1.4
Bzw2 3.31 1.55
Sub1 4.7 2.42
Rbm39 2.21 1.27
Yeats4 2.65 1.57
Metap1 2.57 1.6
Setx 2.53 1.61
Ywhae 5.85 3.77
Serinc1 2.28 3.51
Top1 3.29 5.29
Ubqln1 3.16 5.31
Ankrd17 2.45 4.12
Srsf5 2.82 4.75
Atp6v1e1 3.12 5.32
Zfp978 2.2 4.02
Dbr1 1.83 3.42
Mtf2 2.2 4.14
Cab39 2.52 4.83
Gm20388 1.71 3.45
Incenp 2.4 4.89
Tpr 3.12 6.51
Pmaip1 2.39 5.25
Sall4 2.26 4.99
Brca2 3.02 6.74
Tlk2 2.44 5.51
Grb2 2.52 5.76
Polr2a 2.99 7.14
Brd2 1.73 4.29
Hnrnpu 4.25 11.23
Taf1d 1.78 4.75
Zfand5 1.84 5.02
Gtf2b 1.97 5.7
Atad2b 2.23 7.06
Zfp990 2.09 6.67
f
Ywhae
Rbfox2
Laptm4a
Setx
Arhgap5
Srsf10
Rbm39
Yeats4
Cdk4
Brd3
Sub1
Cpeb4
Sart3
Pabpc1
Pabpc4
Mrpl45
2610507B11Rik
Spp1
Metap1
Nfyb
Bzw2
RNA binding proteins
Target RNAs



Supplementary Figure 1: Quality control of scNT-Seq in mESCs.

a. Violin plot comparing the number of genes and number of UMIs detected in individual cells among different treatment conditions. Cell number and sequencing depth were shown on the top.
b. Scatter plot showing reproducibility between two replicates.
c. Transcriptome-wide nucleotide substitution rates reveal the effect of labeling time (100 μM 4sU/4hr or 200 μM 4sU/24hr) and methanol fixation (two rehydration buffers, PBS versus SSC) on T-to-C conversion rate. The control sample (100 μM 4sU/4hr) was not treated with TFEA/$NaIO_4$.

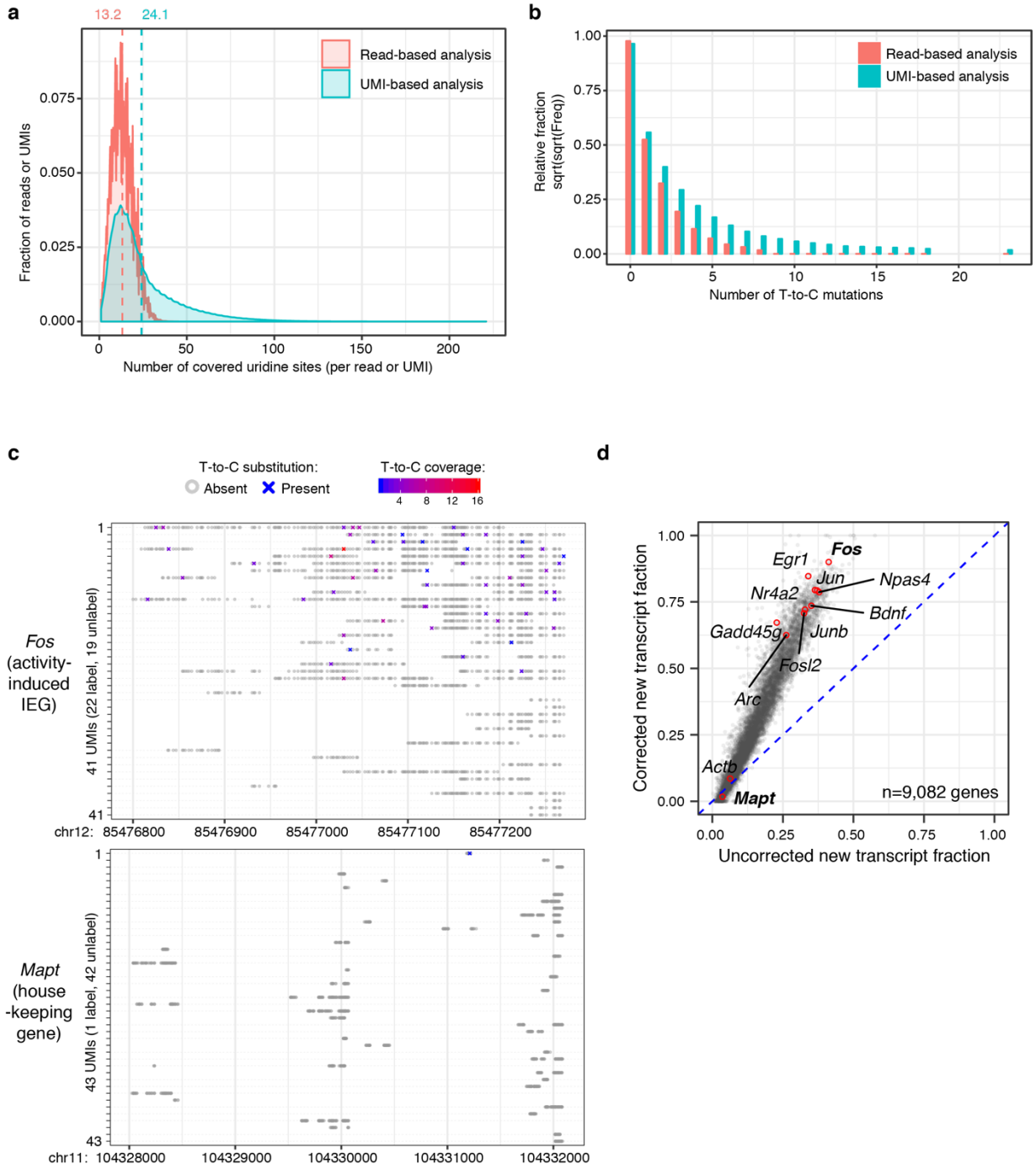


Supplementary Figure 2: scNT-Seq identifies different cell types in primary mouse cortical cultures.

a. UMAP visualization of 20,547 cells from mouse cortical cultures (same UMAP plot as Fig. 2b). The cells are colored by different durations of neuronal activation.

b. Violin-plot showing gene expression levels of representative marker genes in different cell-types.

c. Marker gene expression is scaled by colors in the same UMAP as (a). Ex, excitatory neurons; Inh, inhibitory neurons; NP, neural progenitors; RG, radial glial progenitor cells.



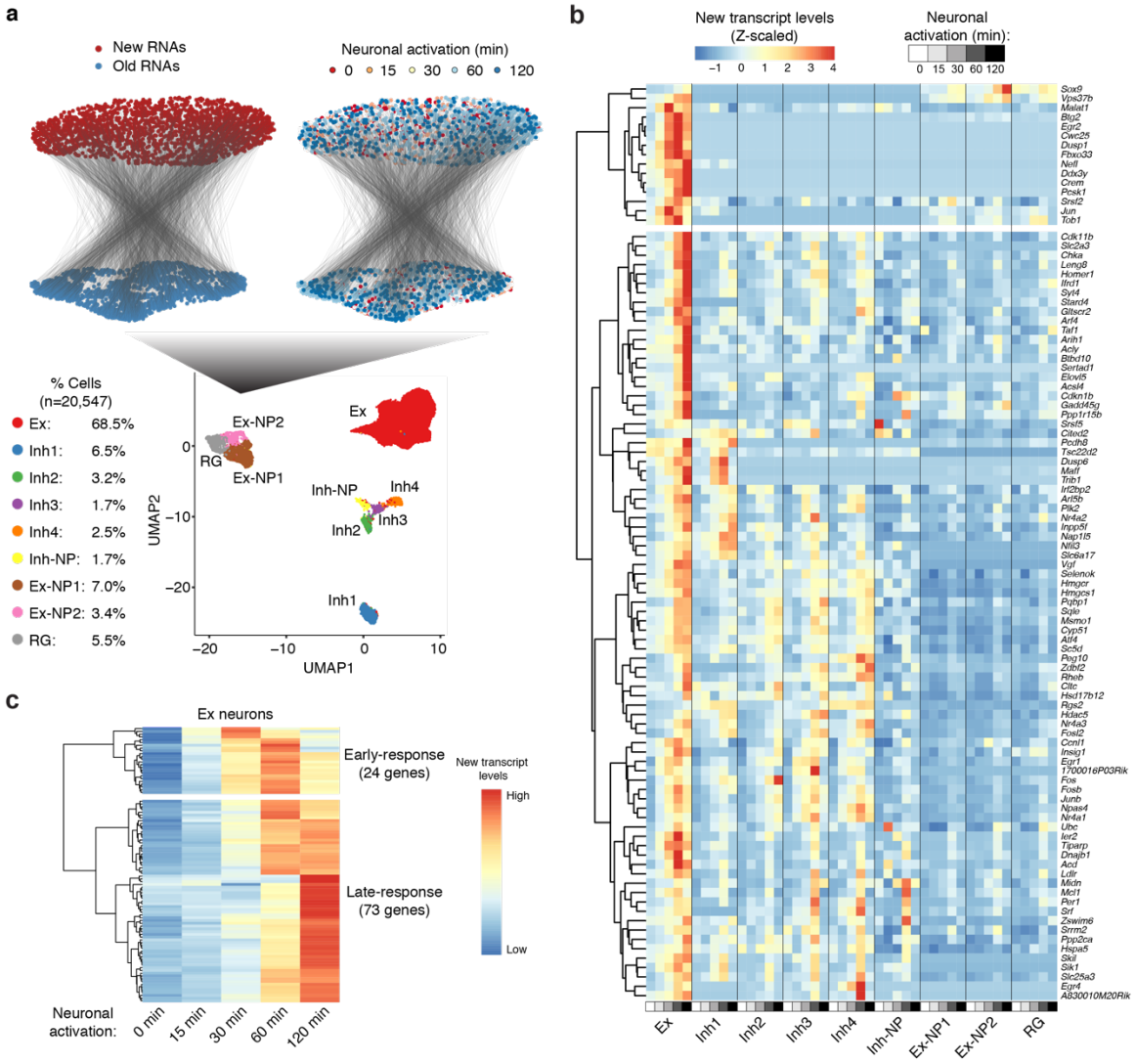
Supplementary Figure 3: UMI-based statistical correction of fractions of newly-transcribed RNAs.

a. Density plot of the number of covered uridine sites per read (60 bp) or UMI. Data set of excitatory neurons with 60 min KCl stimulation were shown.

b. Bar plot of the number of T-to-C substitutions per read (60 bp) or UMI. Data set of excitatory neurons with 60 min KCl stimulation were shown.

c. All UMIs of the *Fos* and *Mapt* genes from one excitatory neuron with 60 min KCl stimulation were shown. Grey circles stand for T without T-to-C conversion, while crosses ("X"s) denote sites of T-to-C conversion in at least one read.

d. Comparison of new RNA levels of each gene in excitatory neurons (with 60 min KCl stimulation) before and after statistical correction. Ten neuron activity-dependent genes and two house-keeping genes are highlighted with red dots.

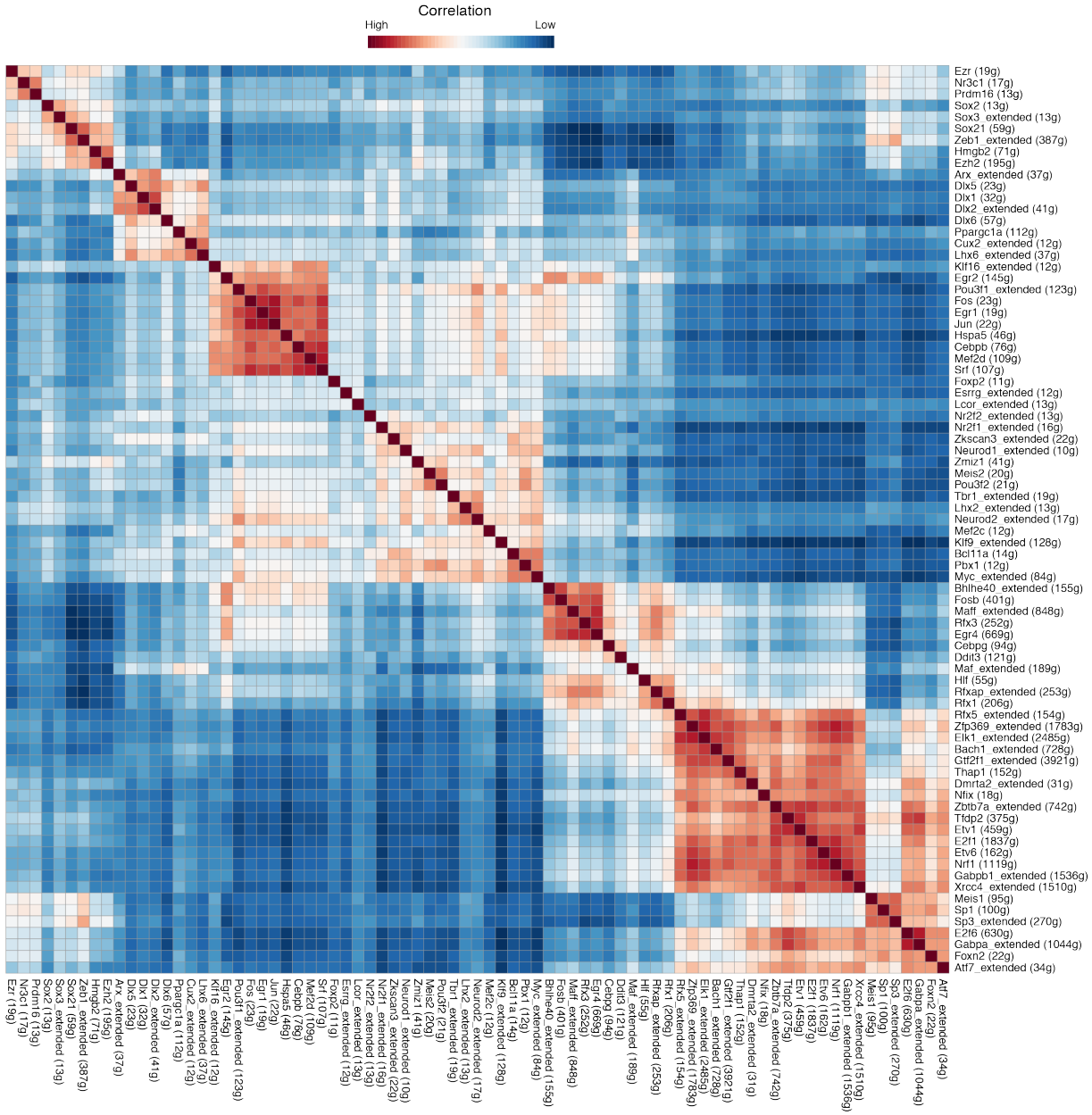


Supplementary Figure 4: Different cell types show different neuron activity-dependent gene expression programs.

a. UMAP visualization of the expression profiles of 20,547 cells from primary mouse cortical cultures with or without KCl stimulation (bottom panel, same UMAP plot as Figure 2b). Cells from non-neuronal cell clusters (RG/Ex-NP) are sub-clustered based on newly-transcribed or pre-existing mRNAs (top). The newly-transcribed and pre-existing mRNAs of same cell were connected by black line. Ex, excitatory neurons; Inh, inhibitory neurons; NP, neural progenitors; RG, radial glial progenitor cells.

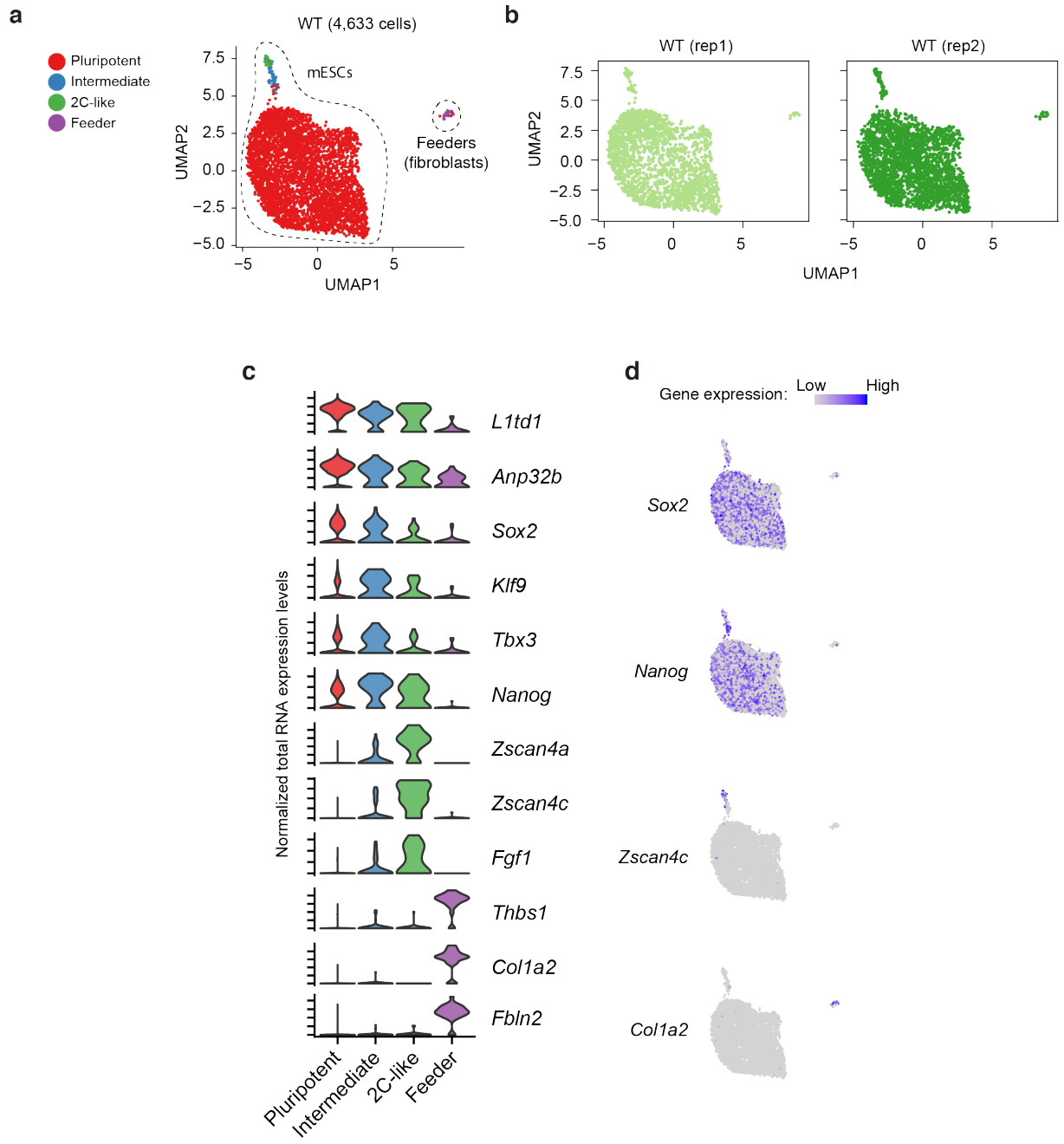
b. Heat map showing new transcript levels of genes with significantly increased expression level after KCl stimulation in at least one cell type.

c. Heat map showing new transcript levels of early- and late-response genes in excitatory neurons with different durations of KCl stimulation. 97 significantly induced genes were clustered into 2 groups (early- and late-response).



Supplementary Figure 5: Heat map of activity-regulated TF regulons in mouse cortical cultures.

Heat map showing correlated transcription factor activity (quantified by SCENIC) during neuronal activity-dependent gene expression.



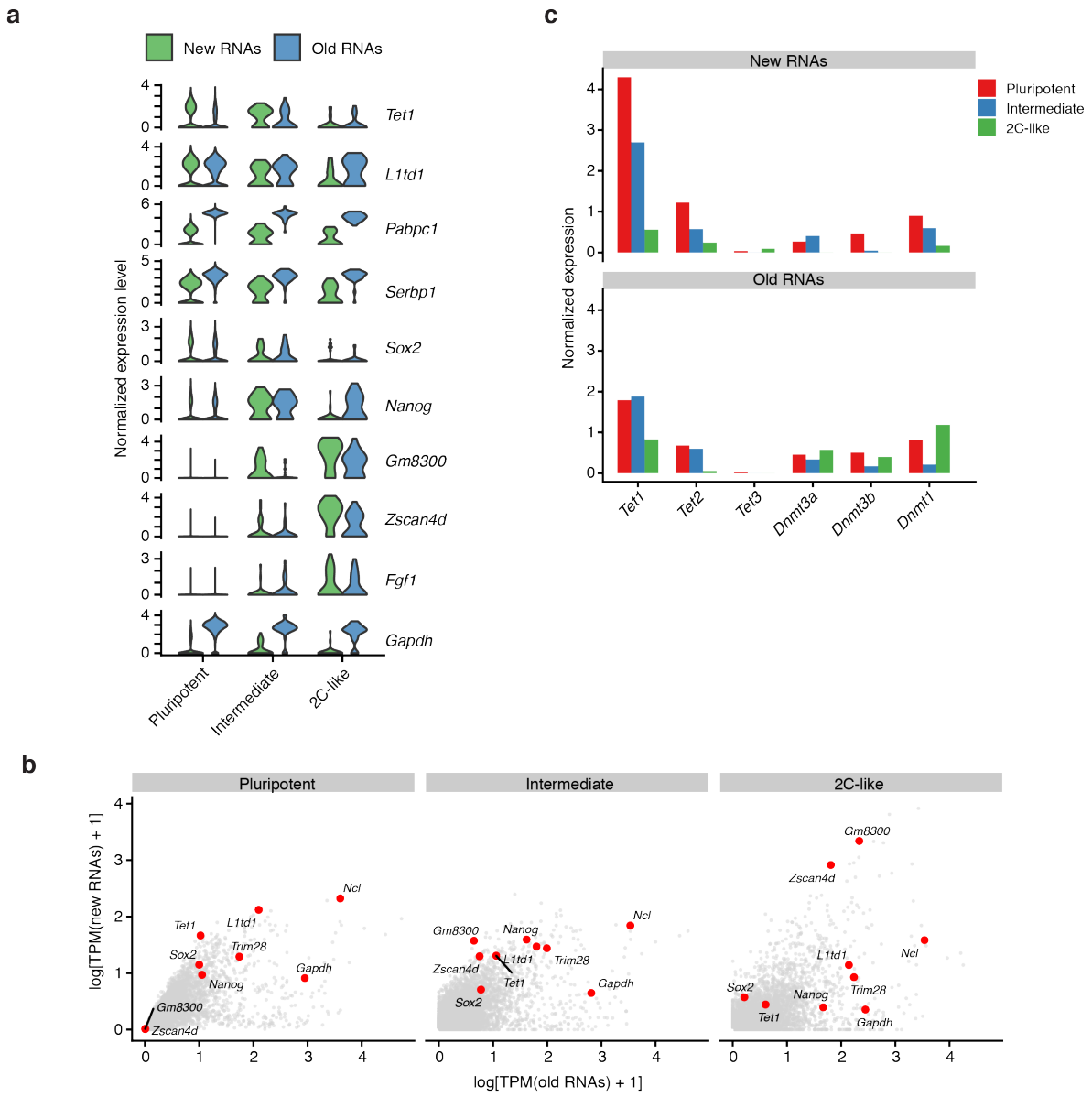
Supplementary Figure 6: scNT-Seq reveals different stem cell states in mESC cultures.

a. UMAP visualization of 4,633 WT mESCs from two replicates. The cells are colored by different cell-types or cell-states. Feeders (purple dots) are contaminated mouse embryonic fibroblasts in cell culture.

b. UMAP plots were colored by replicate.

c. Violin plot showing gene expression level of marker genes in different stem cell states or cell-types.

d. Feature-plots showing marker gene expression is scaled by colors in the same UMAP as (a).

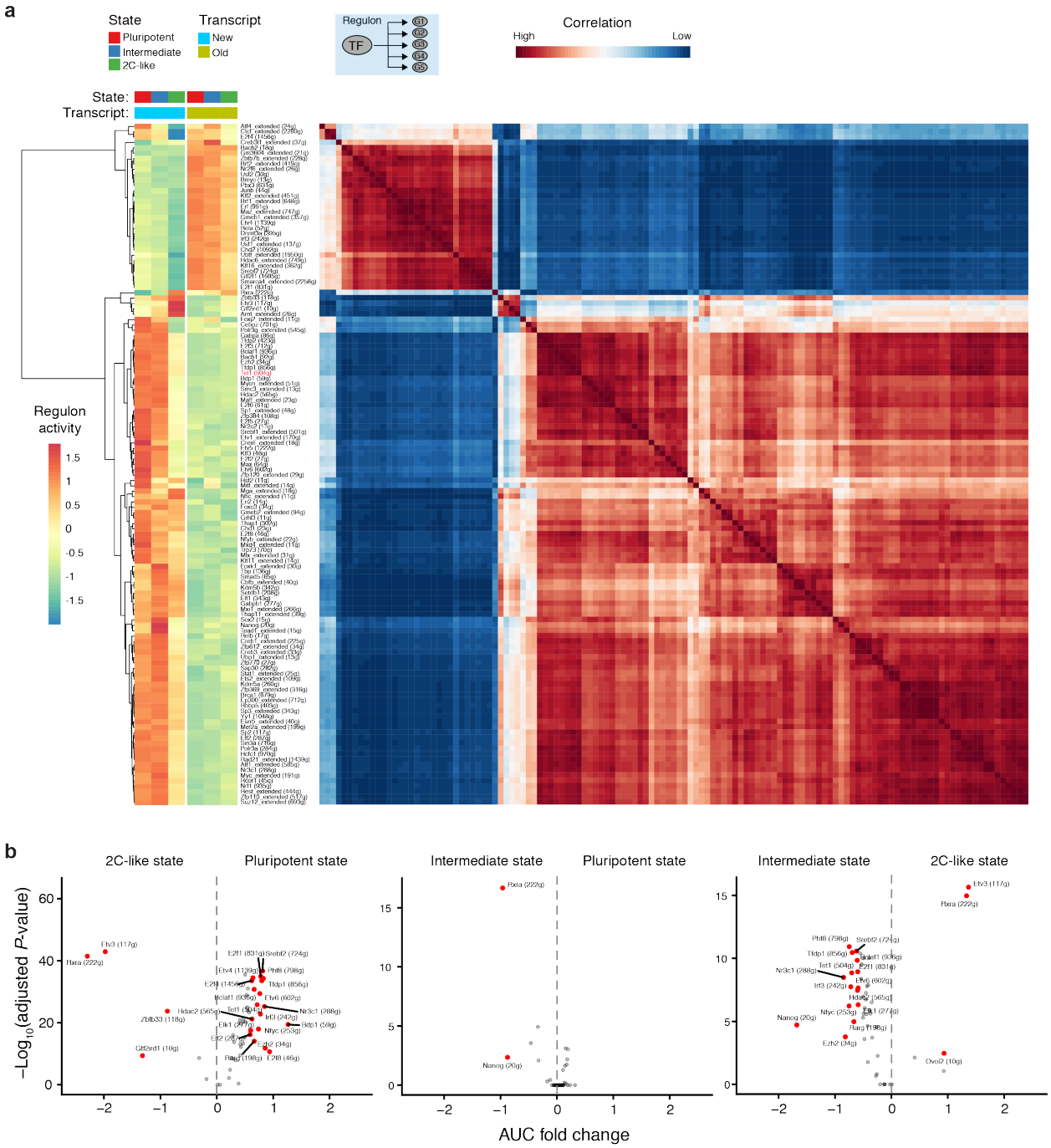


Supplementary Figure 7: scNT-Seq reveals newly-transcribed and pre-existing RNA level of representative genes in mESCs.

a. Violin plots showing expression levels of newly-transcribed and pre-existing transcripts of selected genes in three cell-states of mESCs.

b. Scatterplots showing newly-transcribed and pre-existing RNA levels of representative genes, including pluripotent genes (*Sox2, Nanog*), 2C-like state specific genes (*Zscan4d, Gm8300*), and house-keeping gene (*Gapdh*).

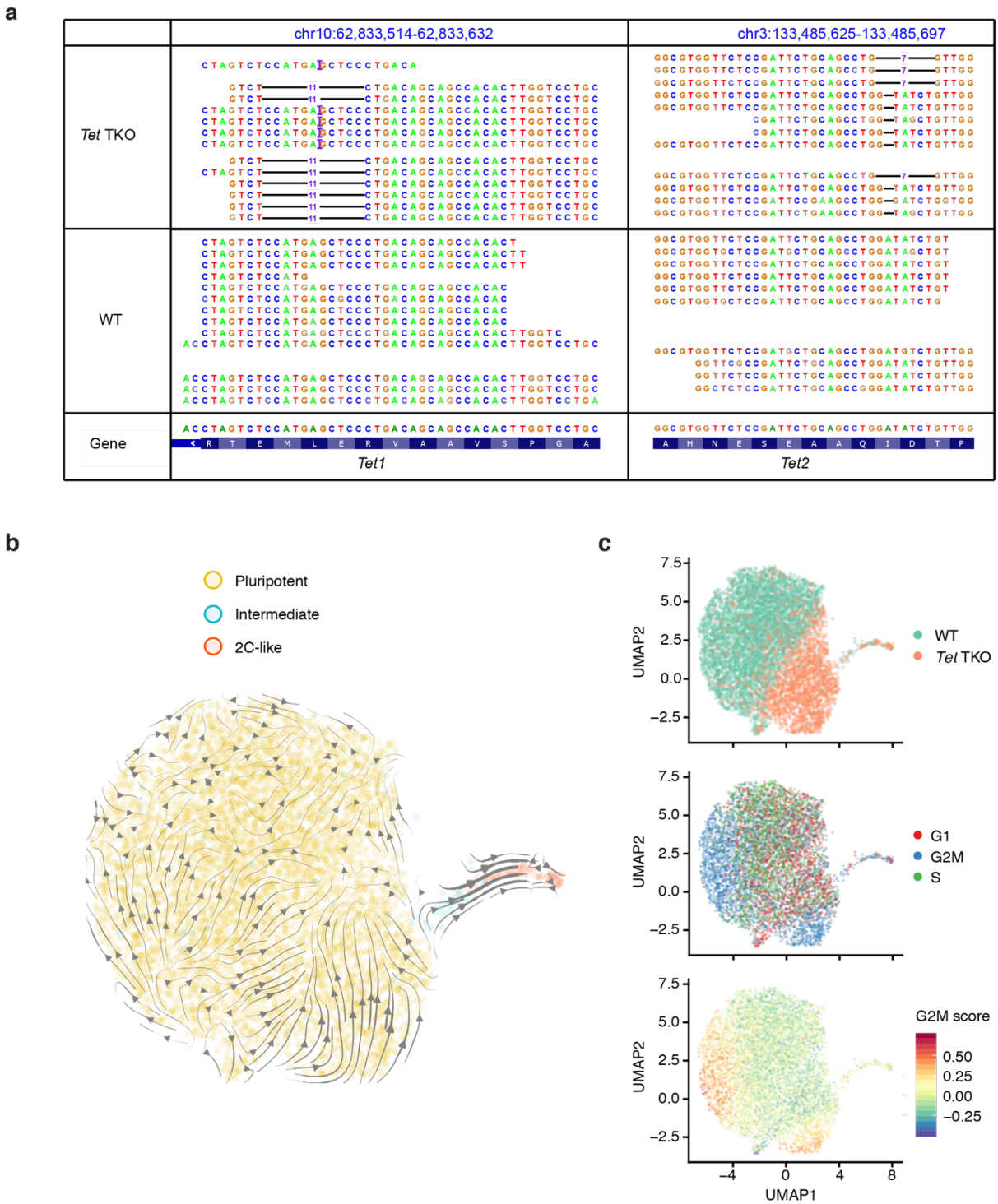
c. Newly-transcribed and pre-existing RNA levels of major DNA methylation regulators in three stem cell states.



Supplementary Figure 8: scNT-Seq reveals TF activity changes during pluripotent-to-2C transition in mESCs.

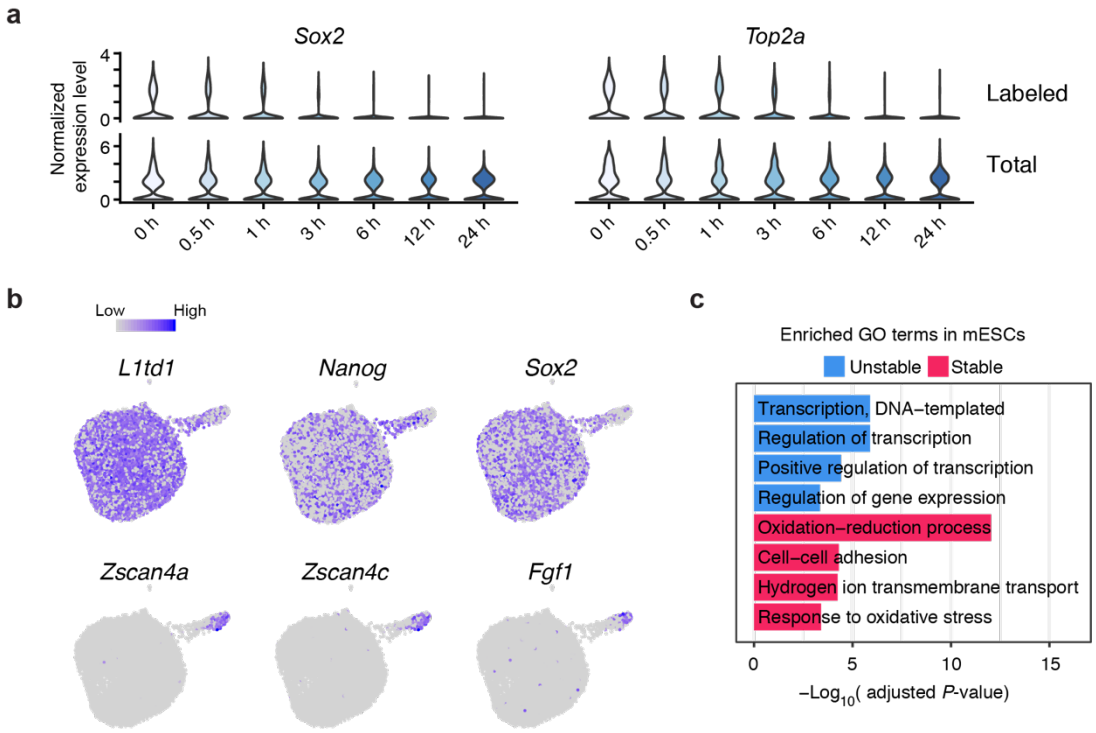
a. Heat maps showing TF regulon activity (left) and the similarity of transcription factors activities in three stem cell states (right).

b. Volcano-plot showing state-specific TF regulon activity. TFs with a fold-change of mean AUC values more than 1.5 and adjust *P*-value less than 0.05 were highlighted in red.



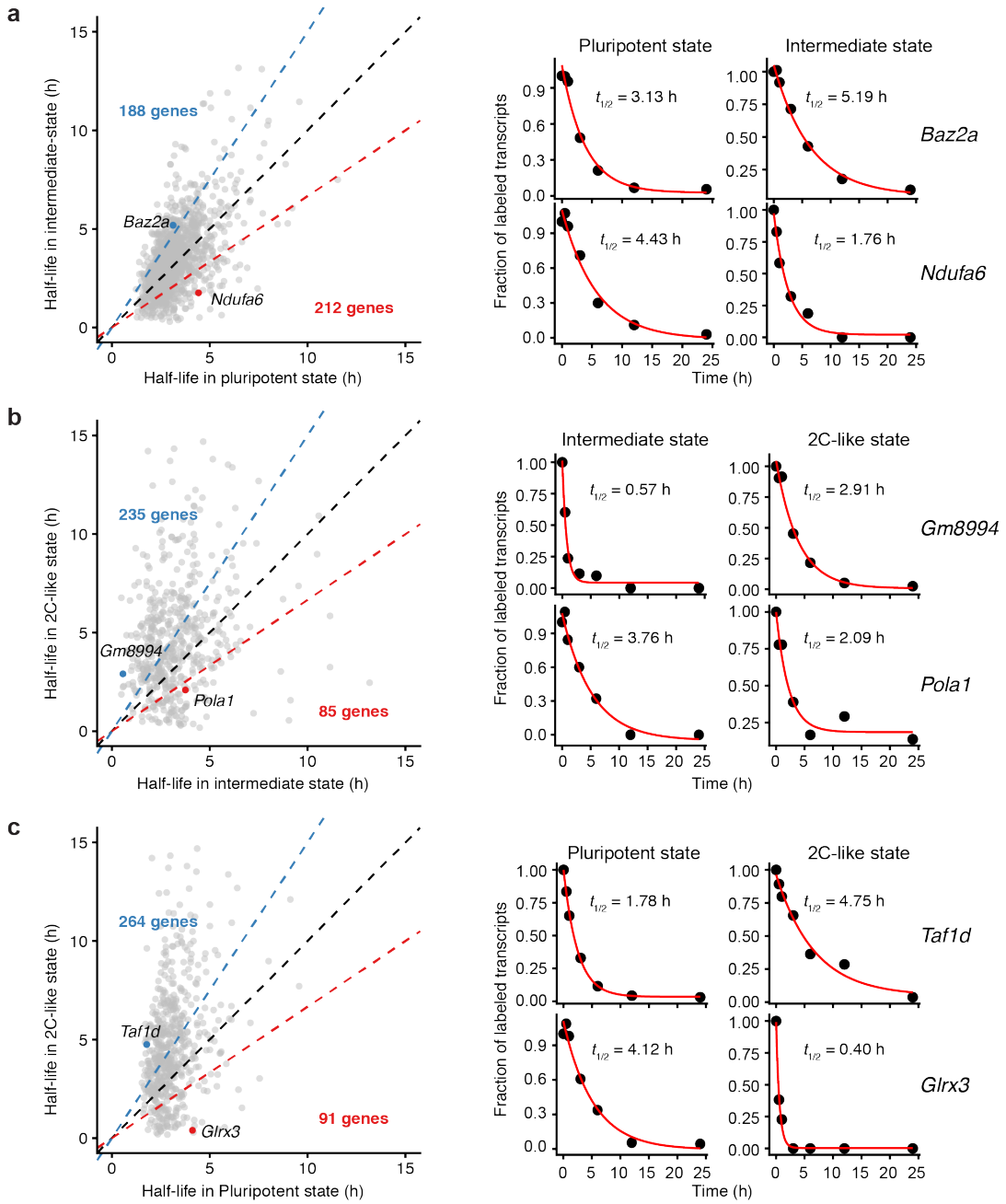
Supplementary Figure 9: scNT-Seq analysis of TET-dependent regulation of the plutipotent-to-2C transition in mESCs.

a. Validation of genotypes of the *Tet1* (-11bp/+1bp) and *Tet2* (-7bp/-1bp) genes in *Tet*-TKO cells by aligning scNT-Seq reads to the CRISPR/Cas9 genome editing sites.
b. Projection of cell states onto the same UMAP plot as Fig. 5a.
c. Projection of genotypes and the cell-cycle states onto the same UMAP plot as Fig. 5a.



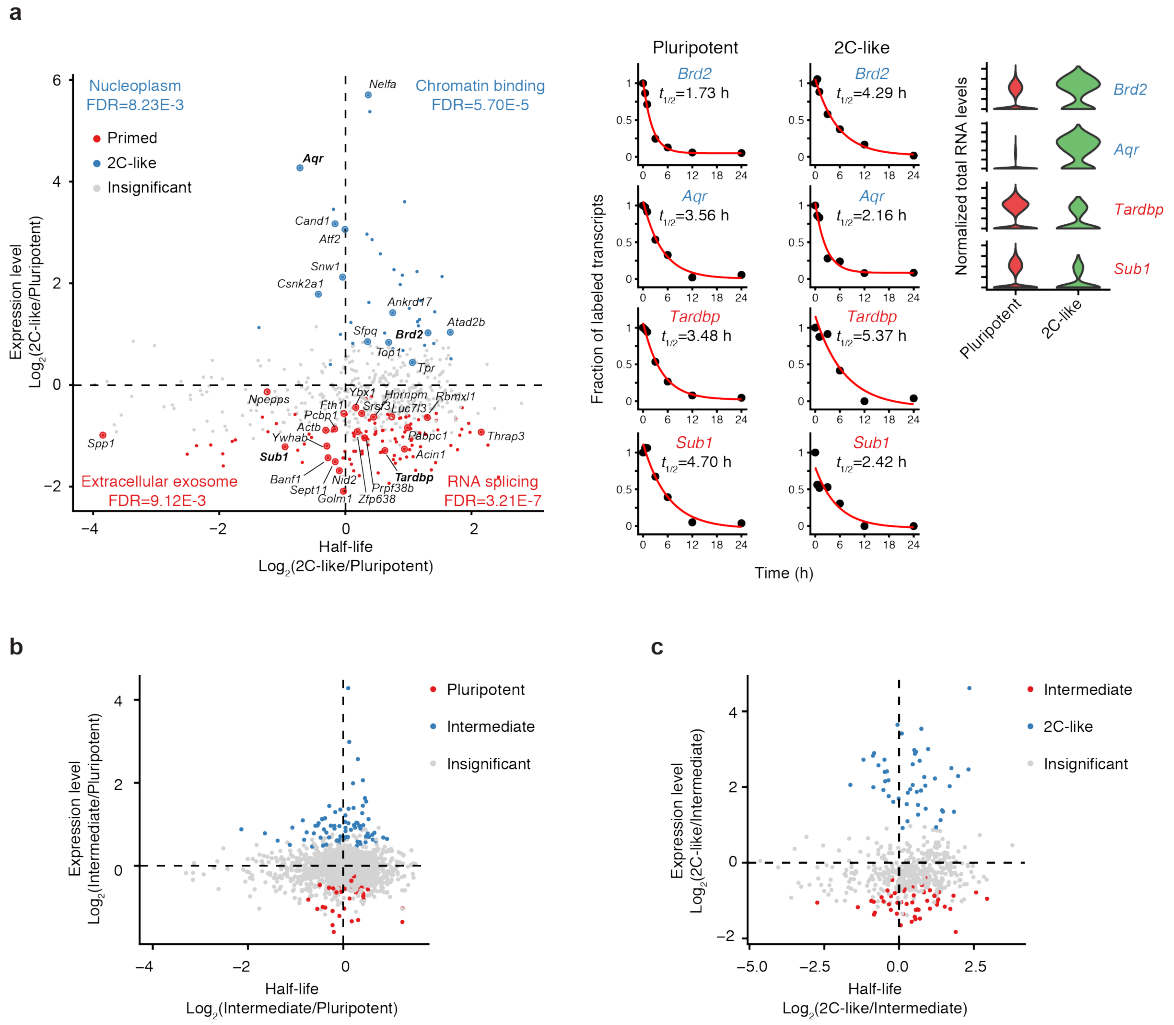
Supplementary Figure 10: Pulse-chase scNT-Seq reveals transcript-specific mRNA decay of mESCs.

a. Violin plots showing levels of labeled and total transcripts of the *Sox2* and *Top2a* genes during pulse-chase assay.
b. Marker gene expression is scaled by colors in the same UMAP plot as Fig. 6b.
c. Enrichment analysis of GO terms within stable (top 10% genes with longest half-lives) and unstable genes (top 10% genes with shortest half-life) in pluripotent state mESCs.



Supplementary Figure 11: Pulse-chase scNT-Seq reveals cell state-specific mRNA decay in mESCs.

Scatterplots comparing the RNA half-life of commonly detected transcripts between two stem cell states (**a**: pluripotent vs. intermediate; **b**: intermediate vs. 2C; **c**: pluripotent vs. 2C) and number of genes showing >1.5-fold change in RNA half-life (indicated by blue and red dashed line) between two states are shown (left panels). Representative genes showing state-specific RNA half-life are shown in the right panels.



Supplementary Figure 12: Characterization of mRNA stability and gene expression levels in different states of mESCs.

a. Scatterplot showing the correlation between gene expression levels (total transcripts) and RNA half-life in pluripotent and 2C-like states. Representative GO terms enriched in genes in each quadrant, as well as genes belong to these GO terms, are highlighted. The mRNA decay kinetics and expression level of selected 4 genes were also shown in the middle panel. Representative state-specific genes with different RNA half-life are shown in the right panel. 2C-like state enriched genes are highlighted in blue, while genes preferentially expressed in the pluripotent state are highlighted in red.

b-c. Scatterplot showing the correlation between gene expression levels (total transcripts) and RNA half-life in different states (b: Pluripotent vs. intermediate; c: intermediate vs. 2C-like). Genes that are differentially expressed in one stem cell state are highlighted in red or blue.