**The landscape of alternative polyadenylation in single cells of the developing mouse embryo**

Vikram Agarwal1,2,\*, Sereno Lopez-Darwin1,\*, Jay Shendure1,3,4,♱

1Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

2Calico Life Sciences, South San Francisco, CA 94080, USA

3Howard Hughes Medical Institute, Seattle, WA 98195, USA

4Brotman Baty Institute for Precision Medicine, University of Washington, Seattle, WA 98195, USA

\*These authors contributed equally to this work

♱Correspondence to Jay Shendure ([shendure@uw.edu](mailto:shendure@uw.edu))

**ABSTRACT**

3′ untranslated regions (3′ UTRs) post-transcriptionally regulate mRNA stability, localization, and translational activity. While 3′-UTR isoforms have been globally quantified in a limited set of cell types using bulk measurements, their spatiotemporal properties among diverse cells types remain poorly characterized. In this study, we examined a dataset of >2 million cells, spanning 5 stages of mouse embryonic development, to quantify changes in average 3′ UTR lengths for each gene and cell. Although we observe a general 3′ UTR shortening in hematopoietic lineages and lengthening in neuronal cell types, 3′ UTR length globally increases as a function of developmental time in all cell types. By measuring 3′-UTR isoforms in an expansive single cell dataset, our work generates a transcriptome-wide and organism-wide map of the dynamic landscape of alternative polyadenylation events during early mammalian development.

**INTRODUCTION**

During transcriptional elongation, the cleavage and polyadenylation machinery governs the specification of the 3′ terminal end of an mRNA [(Di Giammartino et al., 2011)](https://paperpile.com/c/Z1pIEY/0NI5). This regulated process can generate a diversity of 3′-UTR isoforms for any given gene, dramatically changing the 3′-UTR lengths of the resulting mature transcripts [(Elkon et al., 2013)](https://paperpile.com/c/Z1pIEY/TJqr). This phenomenon, known as alternative polyadenylation (APA), has been observed in over 70% of mammalian genes [(Derti et al., 2012; Tian and Manley, 2013)](https://paperpile.com/c/Z1pIEY/KqnY+JU5c). Alternative 3′-UTR isoforms bind to different sets of microRNAs and RNA-binding proteins, which collectively modulate a multitude of post-transcriptional gene regulatory mechanisms [(Tian and Manley, 2017)](https://paperpile.com/c/Z1pIEY/Oa0f). These include changes in a mRNA localization [(Berkovits and Mayr, 2015)](https://paperpile.com/c/Z1pIEY/a422), degradation rate [(Agarwal et al., 2015; Lianoglou et al., 2013; Nam et al., 2014)](https://paperpile.com/c/Z1pIEY/jGHJ+9xZl+WH6q), and translational efficiency [(Mayr, 2016)](https://paperpile.com/c/Z1pIEY/hbqQ). The differential abundance of a variety of nuclear factors, which recognize sequence motifs and RNA secondary structures, serves to regulate APA in a cell-type-specific manner [(Di Giammartino et al., 2011)](https://paperpile.com/c/Z1pIEY/0NI5). Abnormal regulation of the cleavage and polyadenylation machinery has also been associated with hyperproliferative or disease states such as cancer [(Fu et al., 2011; Mayr and Bartel, 2009; Sandberg et al., 2008)](https://paperpile.com/c/Z1pIEY/CKSk+rHGZ+KLoA).

Existing techniques to directly measure APA in the transcriptome largely rely upon the isolation of bulk tissue, resulting in an average readout of the landscape of 3′ ends in a heterogeneous population of cells. The earliest transcriptome-wide approaches relied upon Expressed Sequence Tag [(Tian et al., 2005)](https://paperpile.com/c/Z1pIEY/LKrB) and Serial Analysis of Gene Expression techniques [(Ji et al., 2009)](https://paperpile.com/c/Z1pIEY/hlhO) to observe APA, but were limited by low sequencing throughput and enzymatic biases. Subsequent methodologies were successfully applied to annotate the 3′ ends of mRNA globally in worms [(Jan et al., 2011; Mangone et al., 2010)](https://paperpile.com/c/Z1pIEY/8tST+X4Nk), flies [(Agarwal et al., 2018; Sanfilippo et al., 2017; Smibert et al., 2012)](https://paperpile.com/c/Z1pIEY/5dW2+owq0+LGUY), and mammals [(Derti et al., 2012; Hoque et al., 2013; Nam et al., 2014; Spies et al., 2013)](https://paperpile.com/c/Z1pIEY/KqnY+hQEq+TcIC+9xZl). These studies have previously demonstrated general lengthening of mammalian 3′ UTRs in neuronal cell types [(Miura et al., 2013)](https://paperpile.com/c/Z1pIEY/06Ex), at different developmental stages.

In contrast to bulk methods, a number of single-cell RNA sequencing protocols enrich and sequence mRNA 3′ ends [(Cao et al., 2017; Hashimshony et al., 2016; Jaitin et al., 2014; Macosko et al., 2015; Semrau et al., 2017; Zheng et al., 2017)](https://paperpile.com/c/Z1pIEY/2vgE+ggp3+TRmF+2fMt+yD6S+G1Sv). By capturing intermediate cell states, these technologies present an unprecedented opportunity to observe APA events during the process of cellular differentiation. They also enable the decomposition of complex tissues into individual cell types and subtypes. Recent methods have greatly expanded the number of cells measured during mouse development in a number of cell types simultaneously [(Cao et al., 2019; Pijuan-Sala et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k+RW09). [(Shulman and Elkon, 2019)](https://paperpile.com/c/Z1pIEY/fWbH). In this study, we examined APA using the most expansive single-cell RNA sequencing dataset published to date, encompassing over 2 million cells across five stages of embryonic mouse development [(Cao et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k).

**RESULTS**

**Enriching genuine polyA sites**

We implemented a simple heuristic strategy to enrich for reads mapping to genuine polyA sites in the genome, inspired by methods developed for bulk analyses [(Derti et al., 2012; Jan et al., 2011)](https://paperpile.com/c/Z1pIEY/8tST+KqnY) but adapted to accommodate biases prevalent in single cell sequencing data arising from nuclear RNA extraction.

**Making a 3′ UTR Dataset**

This was analyzed in conjunction with a set of files which contained sci-RNA-seq read data for each mRNA transcript within each cell. Using the original transcriptional data, we built a 3′ UTR length dataset based on read loci and gene UTR annotations.

We used our established transcript read set [(Cao et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k) together with a reference dataset which contains the longest known 3′ UTR regions for each gene in the mm10 genome [(Agarwal et al., 2015)](https://paperpile.com/c/Z1pIEY/WH6q). We then overlapped the sci-RNA-seq3 reads to identify which 3′ UTRs each read overlapped (if any). This eliminated about 90% of the data, the remaining of which we annotated with cell and gene origin. Then, using the read positions and strand information, we measured the distance from each read to the beginning of the 3′ UTR which it overlapped, taking into account other upstream 3′ UTR exons of the same gene. With this data, we constructed a full cell-by-gene matrix of each gene's 3′ UTR length in each cell. For any gene that had multiple reads of different 3′ UTR lengths in a single cell, we averaged the lengths. We used this matrix to find the median 3′ UTR length of each gene across the entire dataset, then re-centered each coordinate to this median length. This gave us a matrix of the "3′ UTR deviation" of each gene in each cell: the number of base pairs that the 3′ UTR of each gene was shortened or lengthened in that cell on average.

**UMAP Trajectory Analysis of Mouse 3′ UTR Lengths**

(If including t-SNE analysis in final version, we can add many statements here about t-SNE and UMAP analysis done by Jun, as the two are linked according to Jun’s paper). Our original annotated dataset has coordinates of UMAP clustering of our data based on cell transcriptomes. Due to the developmental window and size of the dataset, Monocle3 was developed and implemented to analyze the dataset with greater efficacy [(Trapnell et al., 2014)](https://paperpile.com/c/Z1pIEY/A7fV)[(Cao et al., 2019; Qiu et al., 2017a, 2017b)](https://paperpile.com/c/Z1pIEY/u7c2+dfwO+nx7k). Of the original 2,000,000 cells, 1,524,792 million cells with UMI counts greater than 400 were used for UMAP clustering. After initial clustering into 10 overarching trajectories, doublet-annotated cells (12% of the dataset) were eliminated, and the resulting groups were iteratively re-clustered into 56 distinct sub-trajectories [(Cao et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k).

We developed our 3′ UTR length dataset indexed to both cells and genes, with each cell-gene combination (of around 2,000,000 and 22,000 genes) being assigned a “3′ UTR length deviation” coordinate based on that cell’s relative mean 3′ UTR lengths compared to the median 3′UTR lengths of each gene. By taking the mean of each 3′ UTR length deviation, we collapsed the data into a single vector of 3′UTR deviations, one for each cell. These values were calculated and normalized taking gene expression levels and unexpressed genes into account, so each value is an accurate representation of the mean 3′ UTR length deviation overall for each cell’s set of genes.

**Lengthening of 3′ UTRs During Mouse Organogenesis**

Though initial UMAP analysis provided an overarching image of 3′ UTR lengthening over cell development, the data was on a cell-by-cell basis, and overlaid on existing UMAP coordinates, causing it to be noisy and difficult to analyze more deeply. To get a more precise view of 3′ UTR length development in our sets of mouse cells, we grouped them by various parameters (one always being embryonic age, from 9.5 to 13.5 embryonic days) and plotted mean cell 3′ UTR deviation over these parameters.

Each cell was organized into two categories: embryonic age and either t-SNE or UMAP clustering (we repeated the process for sub-clusters and sub-trajectories, though the data for these was more sparse due to the fact that many of these developmental categories only existed at certain embryonic ages in our dataset). For each intersection of these categories, we averaged the collapsed cell data (as was used in the UMAP trajectory analysis) for all cells in the category, resulting in one value representing the average 3′ UTR length deviation for that intersection. We then plotted these values as a cluster-by-age heatmap to visualize 3′ UTR lengthening over cell development for each cluster.

For both the t-SNE and UMAP clusters, we observed overall 3′ UTR lengthening across the entire data set during our timeframe. This lengthening was especially pronounced from days 9.5 to 10.5, and days 11.5 to 12.5.

Though the lengthening trend was maintained throughout the entire dataset, certain genes had inconsistent lengthening patterns. Investigating a subset of 1000 genes selected for data count and consistent expression, we observed several developmental 3′ UTR lengthening patterns. Each pattern resulted in 3′ UTR lengthening during organogenesis, but the timeframe of most pronounced lengthening and general trend over all four developmental days differed. Many genes would also undergo 3′ UTR shortening over one day despite an overall lengthening over the other three. When we visualized individual genes using the UCSC Genome Browser [(Kent et al., 2002)](https://paperpile.com/c/Z1pIEY/oJS9), their individual reads showed a clear distribution of dynamic PAS usage over organogenesis that matched their ‘patterns’ in our heatmap.

**Lengthening of 3′ UTRs Over Developmental/UMAP Pseudotime**

Due to the size of the cellular dataset, separating each embryo into five distinct age categories could lead to some inaccuracy. Additionally, the original dataset grouped cells by ‘developmental trajectories,’ which relied on a continuous UMAP pseudotime scale based on gene expression levels. Using the pseudotime coordinates and original developmental trajectory annotations, we analyzed our 3′ UTR data based on this continuous scale. We used sigmoidal normalization as well as crude binning to separate the pseudotime points into distinct categories to allow for visualization alongside individual gene data. Sigmoidal normalization proved to have some missing data bins, however, due to the distance between clusters of pseudotime coordinates. Because of this, we also used a flat binning which binned pseudotime points to their percentile rather than on a normalized scale.

Pseudotime coordinates were provided on both UMAP trajectory and UMAP sub-trajectory scales, and we analyzed each. Sub-trajectory scales had less distinct lengthening and more missing data (though some sub-trajectories were quite robust), but the overall data trends still followed the progressive lengthening trend observed in our embryonic age analysis. Additionally, groups of genes with distinct identifiable lengthening and shortening patterns were observed in every sub-trajectory group, although these groups were also variable per sub-trajectory. This was mirrored in t-SNE cluster analysis of our embryonic age data, which revealed distinct sets of 3′ UTR lengthening patterns at a single-gene resolution in different tissue types.

At a trajectory scale, different patterns emerged once again, but the overall 3′ UTR lengthening of the dataset was more pronounced across pattern sets. This trend was not as consistent over pseudotime as it is over age categories, and there was no significant lengthening between the earliest and latest pseudotime bins, although there was over the course of the full pseudotime scale.

**DISCUSSION**

During mouse embryo development, the 3′ UTRs of cells lengthen as a whole, with some tissue types ending up with longer 3′ UTRs on average [(Ji et al., 2009; Miura et al., 2013)](https://paperpile.com/c/Z1pIEY/06Ex+hlhO). Using scRNA-seq, we are able to sequence the entire transcriptome of millions of mouse cell [(Briggs et al., 2018; Farrell et al., 2018; Wagner et al., 2018)](https://paperpile.com/c/Z1pIEY/JUBj+vKuB+V9z2), which gives us a low-level view of these developmental processes at work. With the foundation of a high-resolution, extensive sci-RNAseq dataset, our analysis has revealed features of 3′ UTR development that were unseen in previous analyses. Though trends such as 3′ UTR lengthening over cell development and variable length trends in different tissues, the contribution of individual genes and subsets of cells hasn’t been seen or analyzed. With this insight, we can also analyze how the developmental landscape differs between these tissues in the same fashion.

Over the embryonic days (E) 9.5 to 13.5, mouse embryos develop nearly all major organ systems and cell count increases hundred-fold to more than ten million. Because this stage is the focus of developmental gene study, the scRNA-seq dataset is over this timeframe [(Cao et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k). Initial analysis on this dataset revealed transcriptomic differences between these cells, and annotated them based on t-SNE and UMAP clustering of this variation. Strong marker genes were used for initial t-SNE clustering, but subsetted clustering used marker gene sets as well as differential expression. UMAP trajectory and sub-trajectory analysis used exclusively transcriptomic differences. We used the resulting cluster and trajectory placements of each cell to construct a pseudotime frame for each gene, to visualize genes’ 3′ UTR variation during organogenesis.

Because of the variability of the size of our distinct datasets (some trajectories had much more robust data than others), sparsity of data for certain analyses was a concern despite the size of the overall dataset. Though our overall trends matched the trends of each subset as we split our data further into each one, emergent features were still present in some subsets. To verify the consistency of the smaller subsets, we downsampled larger subsets of the same resolution to the sizes of the smaller ones, and observed the same data trends as we had before downsampling. This indicates that our trends were visible at low enough resolution that even with some of the less robust subsets, visible trends were still indicators of cellular features. However, on the basis of individual genes, these lower resolution datasets were less useful as some of our analyzed genes had no data points for certain cell subsets.

Actual data trends observed are generally consistent with known biology [(Miura et al., 2013)](https://paperpile.com/c/Z1pIEY/06Ex), [(Ji et al., 2009)](https://paperpile.com/c/Z1pIEY/hlhO), but new trends have been revealed and existing trends expanded upon. Certain tissue types which were previously only found to have less ‘distal polyadenylation site (PAS) usage’ we have observed to have significant 3′ UTR shortening after differentiation. Though our data is linked in this way to previous literature, previous analyses have also used a binary format of 3′ UTR length analysis, taking into account only ‘long’ and ‘short’ 3′ UTRs, depending on usage of a ‘distal PAS’ or a ‘proximal PAS’, respectively. This binary was used to give each tissue type single length values based on the percentage of genes that preferred distal PAS [(Miura et al., 2013)](https://paperpile.com/c/Z1pIEY/06Ex); [(Ji et al., 2009)](https://paperpile.com/c/Z1pIEY/hlhO). However, due to many genes’ usage of more than two PAS, this binary analysis is inaccurate and low-resolution [(Ozsolak et al., 2010)](https://paperpile.com/c/Z1pIEY/qw3V). Not only does it not take into account medial PAS, it also doesn’t take into account the length difference between each PAS, which also contributes to length trend data. Additionally, because of the inconsistency in lengthening patterns of individual genes in our dataset, the established 3′ UTR variability in cell types and developmental stages is shown to not be a ubiquitous phenomenon across the transcriptome, but rather resulting from certain subsets of genes which have distinct developmental patterns. Because APA has been shown to have distinct regulatory mechanisms for small sets of genes [(Tian and Manley, 2017)](https://paperpile.com/c/Z1pIEY/Oa0f) and specific tissues [(Nimura et al., 2016)](https://paperpile.com/c/Z1pIEY/MdPU), this implies that different regulatory mechanisms change activity in distinct ways over cell development and tissue differentiation.

As more specific mechanisms of APA are discovered [(Di Giammartino et al., 2011; Kasowitz et al., 2018; Martin et al., 2012; Mayr, 2016; Nimura et al., 2016; Tian and Manley, 2017; Zhu et al., 2018)](https://paperpile.com/c/Z1pIEY/Oa0f+F2ni+hbqQ+MdPU+mfcF+K3kA+0NI5), our collection of high-resolution development patterns will give more insight into the control and context of these mechanisms. Additionally, as more APA regulation of our analyzed genes are demonstrated, we may be able to more distinctly categorize our gene groupings beyond pure length data trends.

**METHODS**

**Stringent filtering for polyA-site-mapping reads**

Sci-RNA-seq3 reads had been previously mapped after clipping all 3′ Adenosines [(Cao et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k). Reads, which were single-ended and 56nt in length, were filtered to retain the subset terminating in three or more 3′ Adenosines. Counting the number of 3′ Adenosines, we extracted the same number of nucleotides downstream of the mapped reads from the mouse genome (build mm10), up to a length of 10 base pairs. Reads harboring 50% or more As in this region were thrown away in order to heavily enrich for sequences harboring untemplated polyA tracks, characteristic of the product of a polyadenylate polymerase. These filters arose from the observation that a short (~5-6nt) polyA stretch was sufficient to nucleate a poly(dT) primer binding event, given that the majority of the corresponding genomic regions did not possess characteristic hallmarks of polyA cleavage sites such as the AATAAA signal and downstream polyT stretch.

**3′ UTR annotation set and selection of 3′-UTR-mapping reads**

We acquired a set of mouse 3′-UTR annotations for protein-coding genes in which each unique stop codon was associated with a representative transcript with the longest annotated 3′ UTR as observed in a diversity of gene annotation databases [(Agarwal et al., 2015)](https://paperpile.com/c/Z1pIEY/WH6q). The choose\_all\_genes\_for\_TargetScan.pl Perl script in the TargetScanTools Github [(Agarwal et al., 2018)](https://paperpile.com/c/Z1pIEY/LGUY) was used to integrate these databases, with the longest 3′-UTR annotations being further augmented using bulk 3P-seq, 3′-seq, and RNA-seq data from mouse muscle, heart, liver, lung, kidney, brain, testes, and white adipose tissues as well as NIH 3T3 and mESC cell lines [(Agarwal et al., 2015; Derti et al., 2012; Miura et al., 2013; Nam et al., 2014)](https://paperpile.com/c/Z1pIEY/9xZl+KqnY+06Ex+WH6q).

The filtered polyA-site-mapping reads were intersected with this set of 3′-UTR annotations using bedtools intersect (-wa -wb -s) [(Quinlan and Hall, 2010)](https://paperpile.com/c/Z1pIEY/ZKTn). Counting the total number of reads mapping to each transcript, the transcript with the greatest number of reads (or a random top-ranked one in the case of a tie) was carried forward to represent each gene (using <http://www.targetscan.org/mmu_72/mmu_72_data_download/Gene_info.txt.zip> to associate transcripts to gene IDs). This procedure avoided the bias of genes with many transcript isoforms being overrepresented in downstream results, or being counted redundantly in cases in which multiple different 3′ UTRs overlapped the same genomic coordinates. Using the remaining set of transcripts and overlapping reads, we computed the 3′-UTR length associated with a read as the number of exonic nucleotides between the 3′ terminal end of the read and the stop codon. These 3′-UTR lengths were used to compute a gene by cell matrix, in which each entry corresponded to the average 3′-UTR lengths among the reads mapping to the gene and a given cell. Finally, for each gene, the median among all cells, or the cluster median, was used to center the non-zero entries in the matrix to evaluate relative differences in 3′-UTR lengths between clusters or subclusters, respectively, based upon the t-SNE clusters attributed to each cell [(Cao et al., 2019)](https://paperpile.com/c/Z1pIEY/nx7k).

**Visualizing Data Using Python Heatmaps**

Heatmaps were generated using the matplotlib and seaborn packages for python [(Waskom et al. 2014; Caswell et al. 2019)](https://paperpile.com/c/Z1pIEY/7V07+YCvA). Each heatmap used data collapsed into distinct developmental categories. For the first set of heatmaps, we collapsed our data matrices based on objective cell ages, separating each cell into one of five ages. For each age, each gene's 3′ UTR deviation was averaged over all nonzero values for the 3′ UTR deviation in each cell of that age. We did not take empty data coordinates into account for any of this normalization, to avoid skewing the means and medians. The heatmap was then plotted using these collapsed bins alongside each gene. For the second set of heatmaps, we instead used analyzed each cell based on its pseudotime coordinate. Pseudotime coordinates are based on the entire dataset and prone to outliers, so first we grouped each pseudotime coordinate based on which UMAP sub-trajectory the cell belonged to. Next, we performed a sigmoid normalization over each pseudotime coordinate in each sub-trajectory. We binned these pseudotime coordinates by tenths on a regular 0 to 1 scale. We then collapsed all genes across each of these bins within each cell in each unique sub-trajectory. Each heatmap was plotted using these collapsed bins alongside each gene expressed within cells belonging to that sub-trajectory, resulting in one heatmap per sub-trajectory.

**FIGURES**

**Figure 1. A next-generation massively parallel reporter assay (MPRA) strategy to measure the transcriptional activity of >10,000-100,000 enhancers simultaneously.**

**SUPPLEMENTARY FIGURES**

**Supplemental Figure 1. Histograms.** Shown.

**SUPPLEMENTAL TABLES**

**Supplemental Table 1.** Genomic.

**AUTHOR CONTRIBUTIONS**

V.A. conceived of the study and designed analyses, and S.L-D. performed the computational analyses. V.A. and S.L-D. generated tables and figures, and all authors wrote the paper.

**ACKNOWLEDGEMENTS**

We thank Junyue Cao for providing early access to single cell data, and members of the Shendure laboratory for critical commentary.

**REFERENCES**

[Agarwal V, Bell GW, Nam J-W, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**. doi:](http://paperpile.com/b/Z1pIEY/WH6q)[10.7554/eLife.05005](http://dx.doi.org/10.7554/eLife.05005)

[Agarwal V, Subtelny AO, Thiru P, Ulitsky I, Bartel DP. 2018. Predicting microRNA targeting efficacy in Drosophila. *Genome Biology*. doi:](http://paperpile.com/b/Z1pIEY/LGUY)[10.1186/s13059-018-1504-3](http://dx.doi.org/10.1186/s13059-018-1504-3)

[Berkovits BD, Mayr C. 2015. Alternative 3’ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**:363–367.](http://paperpile.com/b/Z1pIEY/a422)

[Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM. 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**:eaar5780.](http://paperpile.com/b/Z1pIEY/vKuB)

[Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. doi:](http://paperpile.com/b/Z1pIEY/G1Sv)[10.1126/science.aam8940](http://dx.doi.org/10.1126/science.aam8940)

[Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, Trapnell C, Shendure J. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**:496–502.](http://paperpile.com/b/Z1pIEY/nx7k)

[Caswell TA, Droettboom M, Hunter J, Lee A, Firing E, Stansby D, Klymak J, de Andrade ES, Nielsen JH, Varoquaux N, Hoffmann T, Root B, Elson P, May R, Dale D, Lee J-J, Seppänen JK, McDougall D, Straw A, Hobson P, Gohlke C, Yu TS, Ma E, Vincent AF, Silvester S, Moad C, Katins J, Kniazev N, Ariza F, Ernest E. 2019. matplotlib/matplotlib: REL: v3.1.1. doi:](http://paperpile.com/b/Z1pIEY/YCvA)[10.5281/zenodo.3264781](http://dx.doi.org/10.5281/zenodo.3264781)

[Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**:1173–1183.](http://paperpile.com/b/Z1pIEY/KqnY)

[Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**:853–866.](http://paperpile.com/b/Z1pIEY/0NI5)

[Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**:496–506.](http://paperpile.com/b/Z1pIEY/TJqr)

[Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**. doi:](http://paperpile.com/b/Z1pIEY/V9z2)[10.1126/science.aar3131](http://dx.doi.org/10.1126/science.aar3131)

[Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, Xu A. 2011. Differential genome-wide profiling of tandem 3’ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* **21**:741–747.](http://paperpile.com/b/Z1pIEY/KLoA)

[Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A, Yanai I. 2016. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* **17**:77.](http://paperpile.com/b/Z1pIEY/ggp3)

[Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3’ region extraction and deep sequencing. *Nat Methods* **10**:133–139.](http://paperpile.com/b/Z1pIEY/TcIC)

[Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I. 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**:776–779.](http://paperpile.com/b/Z1pIEY/2fMt)

[Jan CH, Friedman RC, Graham Ruby J, Bartel DP. 2011. Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. *Nature*. doi:](http://paperpile.com/b/Z1pIEY/8tST)[10.1038/nature09616](http://dx.doi.org/10.1038/nature09616)

[Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* **106**:7028–7033.](http://paperpile.com/b/Z1pIEY/hlhO)

[Kasowitz SD, Ma J, Anderson SJ, Leu NA, Xu Y, Gregory BD, Schultz RM, Wang PJ. 2018. Nuclear m6A reader YTHDC1 regulates alternative polyadenylation and splicing during mouse oocyte development. *PLoS Genet* **14**:e1007412.](http://paperpile.com/b/Z1pIEY/mfcF)

[Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**:996–1006.](http://paperpile.com/b/Z1pIEY/oJS9)

[Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. 2013. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* **27**:2380–2396.](http://paperpile.com/b/Z1pIEY/jGHJ)

[Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**:1202–1214.](http://paperpile.com/b/Z1pIEY/TRmF)

[Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, Mis E, Zegar C, Gutwein MR, Khivansara V, Attie O, Chen K, Salehi-Ashtiani K, Vidal M, Harkins TT, Bouffard P, Suzuki Y, Sugano S, Kohara Y, Rajewsky N, Piano F, Gunsalus KC, Kim JK. 2010. The Landscape of C. elegans 3′UTRs. *Science* **329**:432–435.](http://paperpile.com/b/Z1pIEY/X4Nk)

[Martin G, Gruber AR, Keller W, Zavolan M. 2012. Genome-wide analysis of pre-mRNA 3’ end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep* **1**:753–763.](http://paperpile.com/b/Z1pIEY/K3kA)

[Mayr C. 2016. Evolution and Biological Roles of Alternative 3’UTRs. *Trends Cell Biol* **26**:227–237.](http://paperpile.com/b/Z1pIEY/hbqQ)

[Mayr C, Bartel DP. 2009. Widespread shortening of 3′ UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**:673–684.](http://paperpile.com/b/Z1pIEY/CKSk)

[Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. 2013. Widespread and extensive lengthening of 3’ UTRs in the mammalian brain. *Genome Res* **23**:812–825.](http://paperpile.com/b/Z1pIEY/06Ex)

[Nam J-W, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, Yildirim MA, Rodriguez A, Bartel DP. 2014. Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol Cell* **53**:1031–1043.](http://paperpile.com/b/Z1pIEY/9xZl)

[Nimura K, Yamamoto M, Takeichi M, Saga K, Takaoka K, Kawamura N, Nitta H, Nagano H, Ishino S, Tanaka T, Schwartz RJ, Aburatani H, Kaneda Y. 2016. Regulation of alternative polyadenylation by Nkx2-5 and Xrn2 during mouse heart development. *Elife* **5**. doi:](http://paperpile.com/b/Z1pIEY/MdPU)[10.7554/eLife.16030](http://dx.doi.org/10.7554/eLife.16030)

[Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM. 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**:1018–1029.](http://paperpile.com/b/Z1pIEY/qw3V)

[Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, Mulas C, Ibarra-Soria X, Tyser RCV, Ho DLL, Reik W, Srinivas S, Simons BD, Nichols J, Marioni JC, Göttgens B. 2019. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*. doi:](http://paperpile.com/b/Z1pIEY/RW09)[10.1038/s41586-019-0933-9](http://dx.doi.org/10.1038/s41586-019-0933-9)

[Qiu X, Hill A, Packer J, Lin D, Ma Y-A, Trapnell C. 2017a. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* **14**:309–315.](http://paperpile.com/b/Z1pIEY/u7c2)

[Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017b. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**:979–982.](http://paperpile.com/b/Z1pIEY/dfwO)

[Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.](http://paperpile.com/b/Z1pIEY/ZKTn)

[Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating Cells Express mRNAs with Shortened 3’ Untranslated Regions and Fewer MicroRNA Target Sites. *Science* **320**:1643–1647.](http://paperpile.com/b/Z1pIEY/rHGZ)

[Sanfilippo P, Wen J, Lai EC. 2017. Landscape and evolution of tissue-specific alternative polyadenylation across Drosophila species. *Genome Biol* **18**:229.](http://paperpile.com/b/Z1pIEY/owq0)

[Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, van Oudenaarden A. 2017. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. *Nat Commun* **8**:1096.](http://paperpile.com/b/Z1pIEY/yD6S)

[Shulman ED, Elkon R. 2019. Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data. *Nucleic Acids Res*. doi:](http://paperpile.com/b/Z1pIEY/fWbH)[10.1093/nar/gkz781](http://dx.doi.org/10.1093/nar/gkz781)

[Smibert P, Miura P, Westholm JO, Shenker S, May G, Duff MO, Zhang D, Eads BD, Carlson J, Brown JB, Eisman RC, Andrews J, Kaufman T, Cherbas P, Celniker SE, Graveley BR, Lai EC. 2012. Global patterns of tissue-specific alternative polyadenylation in Drosophila. *Cell Rep* **1**:277–289.](http://paperpile.com/b/Z1pIEY/5dW2)

[Spies N, Burge CB, Bartel DP. 2013. 3’ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res* **23**:2078–2090.](http://paperpile.com/b/Z1pIEY/hQEq)

[Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* **33**:201–212.](http://paperpile.com/b/Z1pIEY/LKrB)

[Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**:18–30.](http://paperpile.com/b/Z1pIEY/Oa0f)

[Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**:312–320.](http://paperpile.com/b/Z1pIEY/JU5c)

[Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381–386.](http://paperpile.com/b/Z1pIEY/A7fV)

[Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**:981–987.](http://paperpile.com/b/Z1pIEY/JUBj)

[Waskom M, Botvinnik O, Hobson P, Cole JB, Halchenko Y, Hoyer S, Miles A, Augspurger T, Yarkoni T, Megies T, Coelho LP, Wehner D, cynddl, Ziegler E, diego, Zaytsev YV, Hoppe T, Seabold S, Cloud P, Koskinen M, Meyer K, Qalieh A, Allan D. 2014. seaborn: v0.5.0 (November 2014). doi:](http://paperpile.com/b/Z1pIEY/7V07)[10.5281/zenodo.12710](http://dx.doi.org/10.5281/zenodo.12710)

[Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**:14049.](http://paperpile.com/b/Z1pIEY/2vgE)

[Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, Engelman AN, Xie X, Hertel KJ, Shi Y. 2018. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Mol Cell* **69**:62–74.e4.](http://paperpile.com/b/Z1pIEY/F2ni)

**Backup text**

Sci-RNA-seq read data was stored in the SAM format, which was converted to BED format for alignment with annotations using bedops sam2bed. Our annotation dataset was converted to BED format using bedops gtf2bed. Annotations were stored in GTF format, also converted to BED format for alignment.

Each read, along with its overlapped gene and parent cell were stored as plain text for processing in python.

Overlap data was centered to global gene medians (or, for some analyses, to trajectory or cluster gene medians) then averaged over several axes to give a two-dimensional set of 3′ UTR lengths (generally gene-by-cell). These two-dimensional arrays were stored as sparse numpy matrices (data type int8) in HDF5 files. Cell names and gene names were stored in the same HDF5 files as their matrices as plain lists, indexed to the appropriate rows and columns of the matrices. Relevant cell information (cluster, trajectory, and age coordinates) was read from the annotation set and stored as a python dictionary object within a pickle file, keyed by cell name. In this way, any cell row of the stored numpy matrices can be attached to cell data coordinates by referencing the indexed cell name from this stored dictionary. This allowed us to create higher-dimensional matrices which attached additional data to each cell-gene 3′ UTR coordinate, or concatenate 3′ UTR coordinates amongst groups of cells.