Nucleic Acids Research, 2026, **54**, gkag368
https://doi.org/10.1093/nar/gkag368
Methods

Single-cell trajectory inference for detecting transient events in biological processes

Alexandre Hutton, Jesús Muñoz-Estrada, Jesse G. Meyer*

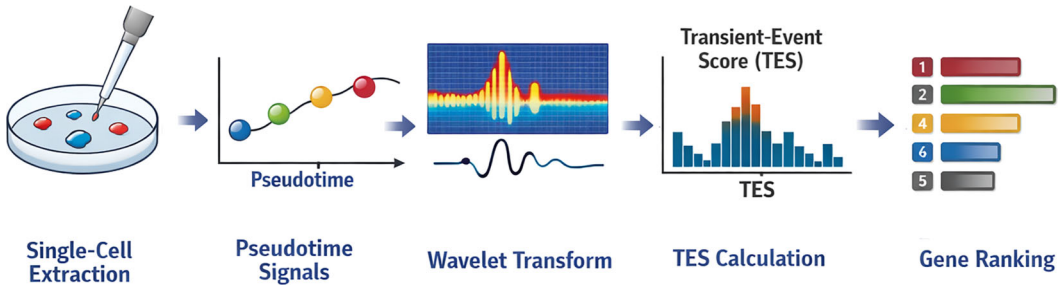
Department of Computational Biomedicine, Board of Governors Innovation Center, Cedars Sinai Medical Center, Los Angeles CA 90048, United States

*To whom correspondence should be addressed. Email: Jesse.Meyer@csmc.edu

Abstract

Transient changes in gene or protein expression often mark the key regulatory checkpoints that propel cells from one functional state to the next, yet they are easy to miss in sparse, noisy single-cell omics data. We introduce **scTransient**, which transforms single-cell expression profiles into continuous pseudotime signals and uses wavelet-based signal processing to isolate short-lived but biologically meaningful bursts of gene activity. After ordering cells with supervised pseudotime, scTransient windows expression values with supervised pseudotime, applies a continuous wavelet transform, and assigns every gene a transient-event score (TES) that rewards sharp, isolated coefficients while penalizing background fluctuations. Synthetic benchmarks demonstrate that TES robustly recovers transient events (TE) across a wide range of parameters, including cell numbers, signal-to-noise ratios, and event widths. Applying scTransient to two real datasets—induced neuron development and single-cell cell cycle—demonstrates scTransient's ability to detect TEs along pseudotime and identify proteins known to be related to the biological process under study. These include stem cell regulators in induced neuronal development and S-phase DNA replication factors in A549 cells. By extending trajectory inference from descriptive ordering to quantitative detection of fleeting regulatory programs, scTransient offers a practical route to uncover transient molecular events that drive development, differentiation, and disease.

Graphical abstract



Introduction

Omic profiles can describe the functional state of individual cells. The quantification of messenger RNA (mRNA) expression or protein levels is continuous across state transitions, supporting the idea that cell states are inherently continuous rather than discrete [1, 2]. Cells nonetheless experience changes in their omic profiles as the cells transition between states, and some of these processes require transient events (TEs) that guide the cell's molecular profile to match its intended function. We expect that, for some processes, it is possible to detect these TEs by measuring omics profiles of cells during that process. Specifically, we distinguish between long-term changes that reflect a change in a cell's function and state-related changes that are only relevant as the cell transitions from one state to another. We do not know *a priori* which genes are relevant to a particular process, and single-cell acquisition methods that enable cell-specific omics profiles are promising approaches. Methods for acquiring such profiles are generally destructive of the sample, limiting our ability to track changes within a single cell. Trajectory inference aims to relate single-cell samples to one another and establish a rough progression of states between samples in a dataset. Although trajectory inference methods are not inherently tied to time, it is possible to use it to examine transitions by connecting a trajectory to a known biological process (e.g. by looking at markers of cell cycle differentiation).

We introduce a method for detecting these TEs that are represented as temporary changes in gene expression by examining the pseudotime course of a biological process, summarized in Fig. 1. Identification is complicated by several factors: (i)

Received: May 8, 2025. **Revised:** March 14, 2026. **Accepted:** March 20, 2026

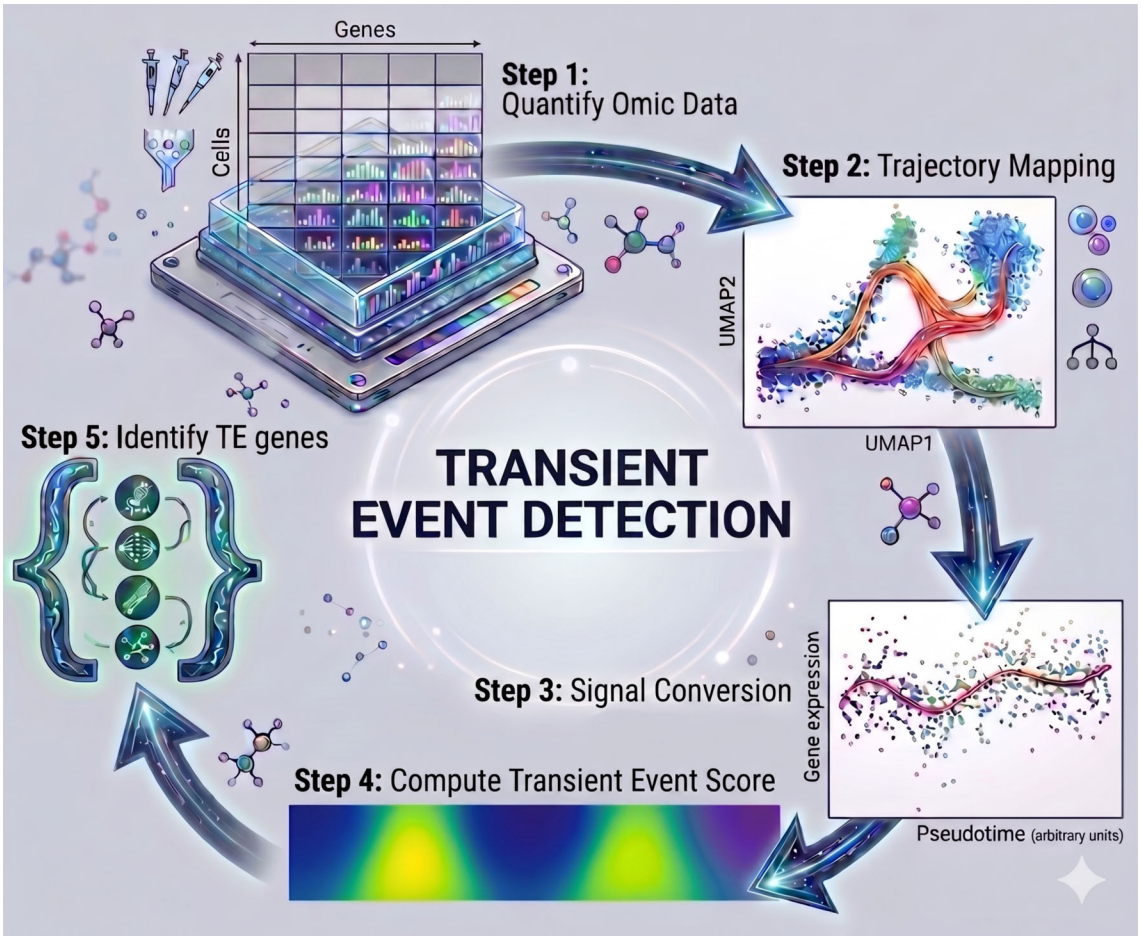


Figure 1. Schematic for detecting genes containing TEs. Starting with the gene quantifications, obtain a trajectory with pseudotime ordering. Convert samples along that trajectory to a pseudotime signal by windowing along pseudotime, then obtain the wavelet coefficients of the signal. From the coefficients, extract a set of genes by scoring the coefficients and selecting genes above a selected threshold.

single cell omics experiments can quantify the products from thousands of genes, making false positives more likely, (ii) ensuring that the pseudotimecourses relate to a particular process requires external process-related measurements, and (iii) it is not known *a priori* what these transitions should look like in pseudotime.

We aim to solve these challenges as follows:

(1) Compare the identified genes to known processes. We can validate the method by examining the set of genes we identify to those that are known to be involved in a particular process. However, we note that any method examining a large number of timecourses is ultimately susceptible to false positives, particularly where the signal-to-noise ratio (SNR) is low.
(2) Use known cell biomarkers that establish the progression of the single-cell samples in pseudotime along a known biological process. For example, cell cycle markers to identify the phase of different samples or cell type markers to identify differentiation along pseudotime. When available, the age of cultured cells is useful for informing pseudotime.
(3) Use a wavelet transform to obtain a scale-adjustable function that can detect transient changes. By finding peaks in the coefficients of the wavelet transform, we can identify positions in pseudotime corresponding to the TEs.

Methods

Trajectory inference and pseudotimecourses

Trajectory inference is the process of linking single-cell samples to one another by examining their omic profiles and inferring an ordering between them. Since omic data acquisition of single-cell samples is destructive, it is necessary to combine the measurements of different cells together to imitate monitoring over the course of a biological process. For this paper, we map cell samples to a pseudotime axis using a supervised pseudotime method, *psupertime* [3]. Using this, we can incorporate more information about a process to order the cells, which we expect results in the method being less sensitive to variations in gene expression that are irrelevant to the process being investigated.

Once cells are ordered on a pseudotime axis, we can convert the collection of samples into gene expression signals by using an averaging window along pseudotime. Windowing can be parameterized by the type, number, and width of the windows. Since this is a preprocessing step, parameter selection at this stage can significantly impact the signal and consequently the information that can be reliably extracted from it. Although

the selected values will depend on the problem being considered (e.g. how important resolution is), we offer some general guidance for determining parameters.

Window type

The type of window dictates the relative importance of data points. A rectangular window weighs all samples within it equally; although straightforward, it tends to over-emphasize changes between windows and samples can be overrepresented if there are overlapping windows. Gaussian windows allow for relative weighting of samples, with samples far from the window center getting downweighted compared to those nearer the center, allowing these to capture general trends while avoiding sample overrepresentation. We generally recommend using Gaussian windows (or other smoothly varying windows), even if they require considering window overlap when selecting the window count and width.

Number of windows and width

The number and width of windows are connected since the entire signal needs to be sampled, placing some bounds between the two parameters. The window count affects resolution; too few windows would fail to capture short transitions while too many would capture trends in noise. The width controls the number of points that are included in a window. Narrow windows provide more resolution but require more samples in the dataset, while wide windows can smooth out noisy signals by considering additional data. Ideally, the parameters are selected to be as narrow as possible while having all windows contain a comparable number of samples.

With the cells converted to a signal representing gene expression along a process, we can use signal processing to identify transient activity and localize it in pseudotime.

Wavelet transform

Signals can be deconstructed and represented in different ways, with different utility. The most useful and common tool is the fast Fourier transform (FFT), which allows us to deconstruct discrete signals into a combination of sinusoidal signals. Although the FFT considers the entire signal at once, modifications such as the short-time Fourier transform (STFT) allows for the FFT to be localized to part of the signal. Without knowing the width of the transient signal of interest, we would need to apply the STFT at multiple window widths. This can be done more efficiently by using the continuous wavelet transform (CWT) [4], which incorporates the width and position of the subsignal, making it ideal for analyzing signals with uncertain properties. The CWT decomposes the signal into signals that are parameterized by their position (location in pseudotime) and scale (width of the signal). For each gene, it produces coefficients for every combination of position and scale, and by identifying extrema in the coefficients we can use it to identify transient changes in gene expression.

Transient-event score

Gene expression signals are noisy, and given that there can be thousands of quantified genes in an omic dataset, it is not feasible to manually review the wavelet transform of every gene signal. Instead, we use a heuristic to score all genes and retain only the top-scoring genes. The heuristic, "transient-event score" (TES), consists of two terms derived from the coefficients obtained from the wavelet transform of the pseudotime signal. For every gene expression signal g, we compute the CWT and extract its set of coefficients C_g. From that, we use each coefficient $c_g \in C_g$ and compute its modified Z score (Z_{mod}, Eq. 1). We obtain the TES for each gene by taking the absolute value of the Hadamard product of Z_{mod} and C_g, then finding its maximum (Eq. 2).

$$Z_{\text{mod}}\left(c_g\right) = \frac{c_g - \text{median}\left(C_g\right)}{\text{std}\left(C_g\right)} \tag{1}$$

$$\text{TES}(g) = \max\left(\left|C_g \odot Z_{\text{mod},g}\right|\right) \tag{2}$$

Using a Ricker wavelet (second derivative of a Gaussian distribution), we can identify TEs that are similarly shaped; high-magnitude coefficients correspond to a better match between the signal and the wavelet being used. Signals with a Gaussian-shaped portion would be ranked higher than one with a gradual signal change. The Z_{mod} component can be interpreted as whether the coefficient value is meaningful; a signal with multiple large coefficients (e.g. noise) would be penalized compared to a signal with a single, large coefficient.

Synthetic data generation

Validating the proposed approach using real datasets is difficult as there are multiple confounding factors. First, establishing a verifiable pseudotime ordering is rarely possible; linking an obtained pseudotime to a specific biological process when cells undergo multiple processes at a given time [5] requires sample metadata that is not always available. In cases where such metadata is available (e.g. cell type labeling), it is not always clear whether cell labels are sufficiently reliable; although it is useful to discretize cell states for certain applications, the continuous nature of cellular omic profiles makes strict boundaries between states questionable. Second, we do not have ground truth for a given process and dataset. While a particular dataset might contain observations related to a particular process, it is possible that it does not capture the transient changes that the method is trying to identify. Third, among genes identified by a method, it is easy to dismiss genes that do not fit expectations and focus on identified genes that make sense for a particular dataset. This is particularly problematic as the intent of the method is to identify new genes that may not be known to fit a particular process.

To address some of these confounds, we start by evaluating the method on synthetic data. For any dataset, we can dictate the pseudotime ordering to be correct and related to a process of interest. It also allows us to set which genes contain a valid transient signal, along with exploring the effect of different processing steps on downstream conclusions. Synthetic data also grants us the ability to evaluate the impact of different aspects of the dataset on performance metrics such as precision and recall. We emphasize that the utility of a synthetic dataset is limited by how well it represents real data. Even with a coarse correspondence between real and synthetic data, it is useful as a first-pass validation.

Synthetic data parameters

We parameterize the synthetic data generation as follows. We first set the total number of cells and genes. Pseudotime is assumed to be in the range of 0–1. Along pseudotime, we produce pseudotime values by sampling from the uniform distribution. When it is nonuniform, we produce low-density regions by subtracting a Gaussian (mean at pseudotime = 0.5, std = 0.1) from the uniform distribution; the parameter con-

trolling the minimum value of this distribution is set to have a range of 0–1 such that "0" means that no samples will be produced at pseudotime = 0.5 and "1" retains the uniform distribution. Gene counts are sampled from two distributions, depending on the experiment: either a Gaussian (mean = 0, std = 0.1) or a negative binomial (r = 10, $p = 0.5$). Genes are separated into two categories: either "signal" or "noise" genes. Signal genes contain the process-related transient change that the method is trying to detect; noise genes contain unmodified counts. The signal genes contain a TE located at pseudotime = 0.5 for single TEs, or uniformly spread in the range [0.1, 0.9] for multiple TEs. Each TE has a parameterized width and amplitude, which is scaled to match the gene count distribution. Autocorrelation along pseudotime is introduced by post-processing the signal. For a gene expression signal $g(pt_{idx})$ as a function of pseudotime index, the autocorrelated signal $g'(pt_{idx})$ is the weighted sum described in Eq. 3, where A is a parameter for controlling the degree of autocorrelation (A = 0 has none, A = 1 is a constant signal). This is done iteratively for each point in pseudotime.

$$g'(pt_{idx}) = g(pt_{idx-1}) * (A) + g(pt_{idx}) * (1 - A) \quad (3)$$

The number of cells for signals were selected from 10, 20, 50, 100, 500; although we do not expect the methods to work well for extremely low sample counts, examining these allows us to identify where the methods start to identify TEs. TE widths were set to 0.02 in pseudotime. TE amplitudes were between 0.01 and 0.6. Samples were sampled uniformly in the pseudotime range of 0–1, except when examining the effects of a nonuniform distribution.

Comparison against other methods

We evaluate the performance of scTransient against tradeSeq [6] and GPfates [7]. We compared the methods using the synthetic datasets produced as described in the previous section. The synthetic datasets were produced with different samples (n = 20 50,100 500), different widths for the TEs (0.001, 0.005, 0.01, 0.05, 0.1 in pseudotime), and TE amplitudes (0.1–6.0 relative to Gaussian noise with 0.1 standard deviation). Each setting was repeated 20 times. The scoring methods between methods are not directly comparable, so we instead opt to produce a known number N of signal genes, select the best N genes according to each method, and calculate the precision.

For nonsynthetic datasets, we compare the genes identified by each method, compare their overlaps, and look at the kurtosis of the signals of the genes identified by each method where we would gene signals with TEs to have higher kurtosis than those without. We also examine the computational time associated with both methods as a function of sample size to determine the feasibility of using the methods on datasets with increasing sample counts and sampling depth.

We note additional methods that are not evaluated in this manuscript: PseudotimeDE [8], ImpulseDE2 [9], Monocle 2 [10], and Monocle 3 [11]. PseudotimeDE modifies tradeSeq to improve the P-values produced by the method, with the tradeoff of long computational times. As stated in the tradeSeq paper, the authors view the produced P-values as numerical summaries rather than a measure with statistical meaning (i.e. a heuristic); as such, we largely ignore the actual P-values. The better-distributed P-values produced by PseudotimeDE do not provide a benefit for the purposes of this paper because the P-values are only used to order genes. ImpulseDE2 is also used to establish differences in gene expression signals but was designed for experiments where time points are known (e.g. every hour after treatment) and uses technical replicates to establish which genes are likely to be different. Consequently, modifying the data to fit ImpulseDE2's assumptions (e.g. binning pseudotime) and evaluating its performance in the context of trajectory inference would not be a fair evaluation. Monocle 2 and 3 could be used to detect transient changes by first computing pseudotime, isolating a lineage, fitting splines using pseudotime as the independent variable, then post-processing the results to identify local extrema and determine which genes are significant. Since this approach is effectively what tradeSeq does and would require substantial custom post-processing code that may not be optimal for the method, we do not evaluate its performance. Alternatively, the approach described in the tradeSeq paper using Moran's I test [12] could also be used; since the comparison has already been made, we do not repeat it.

tradeSeq

The quantities matrix was transposed so that rows represent genes and columns represent samples. For each dataset, the number of knots to use for the GAM was selected by using tradeSeq's *evaluateK* function across k = 3 to k = 14 and selecting the value that produced the lowest AIC [13]. The *cellWeights* parameter was set to a vector with each element equal to 1 to indicate that all cells belonged to the same lineage. tradeSeq's *associationTest* method was used to produce P-values. Using default parameters resulted in poor performance with NaN values for almost all signal genes; changing the inversion method to "*eigen*" produced substantially better results and was used to evaluate tradeSeq. We note that in practice with real datasets, it would not be possible to know that this change improves performance. From the results produced by tradeSeq, the $N = 100$ genes with the lowest P-values were selected and compared to the known signal genes to calculate the precision. Signal genes with NaN P-values were excluded from the calculation.

GPfates

The GPfates Python package was modified to update calls to deprecated methods in its dependencies, but its core functionality remains unchanged. Its dependency on GPclust [14–17] requires NumPy [18] versions below 2.0.0, which is likely to become increasingly problematic with time as the project appears to no longer be maintained. GPfates fits an Overlapping Mixture of Gaussian Processes, and as described in the GPfates paper, we compute the D-statistic (max. likelihood ratio between the models). Precision is computed as the fraction of the signal genes that have the largest values for the D-statistic.

Preprocessing pipelines

The pipelines described here are available on PSCS [19], a platform for producing reproducible analyses for single-cell experiments. Note that in our analyses, we kept zero values, but users of scTransient may prefer to exclude missing values.

Cell cycle

The dataset was produced by processing the raw data made available through the ProteomeXchange Consortium [20] via PRIDE [21] with ID PXD049412. Cells were filtered to have at least 2000 genes, and genes were required to be present in

at least 3 cells. Normalization and and $\log_e$ were applied before regressing out the number of genes as a confound. Data were rescaled to 0 mean and 1 standard deviation with a value capped at 10. Principal Component Analysis (PCA) was computed using the ARPACK solver, sample neighborhoods were computed with n_neighbors = 10 and n_pcs = 40. Cell cycle phase labels were predicted by selecting known cell-cycle-related genes from [22], assigning phase labels to each sample using Scanpy's *score_genes_cell_cycle* function, then setting the phase order as G1 = 1, S = 2, and G2M = 3. Using these ordinal labels, *psupertime* was used to predict a pseudotime value for each value.

Culture of iNeurons and time-course proteomic analysis

Generation of the genome-edited wild-type induced pluripotent stem cells (iPSCs) line KOLF2.1J carrying a doxycycline-inducible NGN2 cassette and the neuronal induction strategy were previously described [23]. Briefly, iPSCs were dissociated to single cells using Accutase, and ~1.8×10^6 cells were plated onto Matrigel-coated 10-cm dishes in Dulbecco's modified Eagle's medium/F12 supplemented with N2, GlutaMAX (1×), NEAA (1×), the ROCK inhibitor Y-27632, and doxycycline (2 μg/ml). After 24 h, the medium was replaced with a fresh induction medium containing doxycycline but lacking the ROCK inhibitor, and cells were maintained for a total of 48 h with daily media changes (neuronal induction period).

iPSC-derived neuronal progenitors were subsequently dissociated with Accutase and cryopreserved in liquid nitrogen at 1.2×10^6 cells per vial until use. The cell yield from a 10 cm plate was a 2-to-3-fold increase. Prior to plating at each time point, cells from vials stored in liquid nitrogen were recovered, and cell viability was assessed by trypan blue staining, which was consistently above 90%. A total of 1×10^5 cells were plated per well of poly-L-ornithine-coated 96-well plates ($n = 8$ technical replicates per condition) in BrainPhys Neuronal Medium supplemented with BDNF (10 ng/ml), NT-3 (10 ng/ml), B27 (1×), mouse laminin (1 μg/ml), and doxycycline (3 μg/ml) and gamma-Secretase Inhibitor IX (DAPT) (3 μM) for 24 h only. Doxycycline was then withdrawn, cultures were rinsed, and cells were maintained with half-media changes twice weekly (neuronal maturation period).

Neuronal cultures (iNeurons) were harvested collectively at 1, 2, 4, 6, or 8 weeks post-induction. For all time points, cells were lysed in radioimmunoprecipitation assay (RIPA) buffer supplemented with protease inhibitors and prepared for bulk proteomic analysis as previously reported [23]. After measuring protein concentration with bicinchoninic acid (BCA) Assay (Pierce), 20 μg of protein per sample was processed using a modified SP3 workflow [24], including reduction and alkylation (dithiothreitol, DTT/iodoacetamide, IAA), protein capture on silica magnetic beads in 70% acetonitrile, bead-based washing, and overnight on-bead tryptic digestion. Peptides were acidified to quench digestion and analyzed by data-independent acquisition (DIA) liquid chromatography-tandem mass spectrometry (LC–MS/MS) using an Orbitrap Astral mass spectrometer. Peptides were separated on a PepSep C18 column using a 30-min gradient, with randomized injection order and periodic HeLa digest quality control (QC) injections.

Raw DIA data were analyzed using DIA-NN (v2.0) in library-free mode against the human UniProt reference proteome (downloaded 12/01/2025). Searches used trypsin/P specificity with up to one missed cleavage; carbamidomethylation of cysteine was set as a fixed modification. Peptide length was restricted to 7–30 amino acids, and precursor charge states to 1–4. Precursor- and protein-level false discovery rates were controlled at 1%. Quantification was performed using the QuantUMS algorithm with retention time (RT)-dependent cross-run normalization and match-between-runs enabled. Protein inference was performed at the gene level using only proteotypic peptides. In the whole experiment, we included mutant cell lines that are not relevant to this study. We extracted only the wild-type (WT) columns from the DIA-NN protein group matrix and shared them here.

Results

We present three sets of results: synthetic data, newly acquired data on iPSC cultured into neurons, and an existing public single-cell dataset characterizing the cell cycle.

Synthetic data results

Results from the synthetic datasets are intended to evaluate the method under known conditions. These serve as a sanity check to verify that the method functions as expected and to explore the impact of changes in the data on downstream analyses and conclusions.

The explicit assumption for this method is that there are detectable transient changes in gene expression. We explore some of the major effects present in single-cell experiments, including sample count, noisy samples, and sample density along pseudotime.

Sample count and SNR

We simulated datasets to investigate the effect of both sample count and the SNR on the detectability of TEs. Counts were generated using a Gaussian for all experiments shown in Fig. 2 before adding the signal of interest. The number of acquired cells has a number of impacts on the detectability of a signal. As demonstrated in Fig. 2A and B the cellular resolution in pseudotime determines whether a TE is detectable even when it is known to exist. More importantly, in the typical scenario with a thousand quantified genes, automatic detection above statistical thresholds is important. We generated datasets with multiple signal genes of varying SNR and sample counts (Fig. 2C). Examining the score, we see a steady increase in event score with increasing SNR. However, low-sample datasets ($n = \{10,20\}$) also had an increasing variance in scoring. For the noise genes in those same datasets, we see a generally decreasing trend in both the mean score and the score variance as the number of samples increases (Fig. 2D). For low-sample experiments, the score distribution is bimodal (Fig. 2E), where the higher scores are from instances where the pseudotime samples were located at the TE. Although the likelihood of a false negative decreases with SNR, it reaches a minimum of 50%–60% for 10 samples, 30%–40% for 20 samples, and 0% for 50 samples with a sufficient SNR (Fig. 2F).

Multiple TEs

The proposed heuristic consists of two components: the coefficient magnitude and the modified Z score. We investigated the impact of the presence of multiple TEs on the overall score assigned to signal genes, reasoning that additional TEs would increase the standard deviation of coefficients and lower the overall score. The results are shown in Fig. 2G, where starting

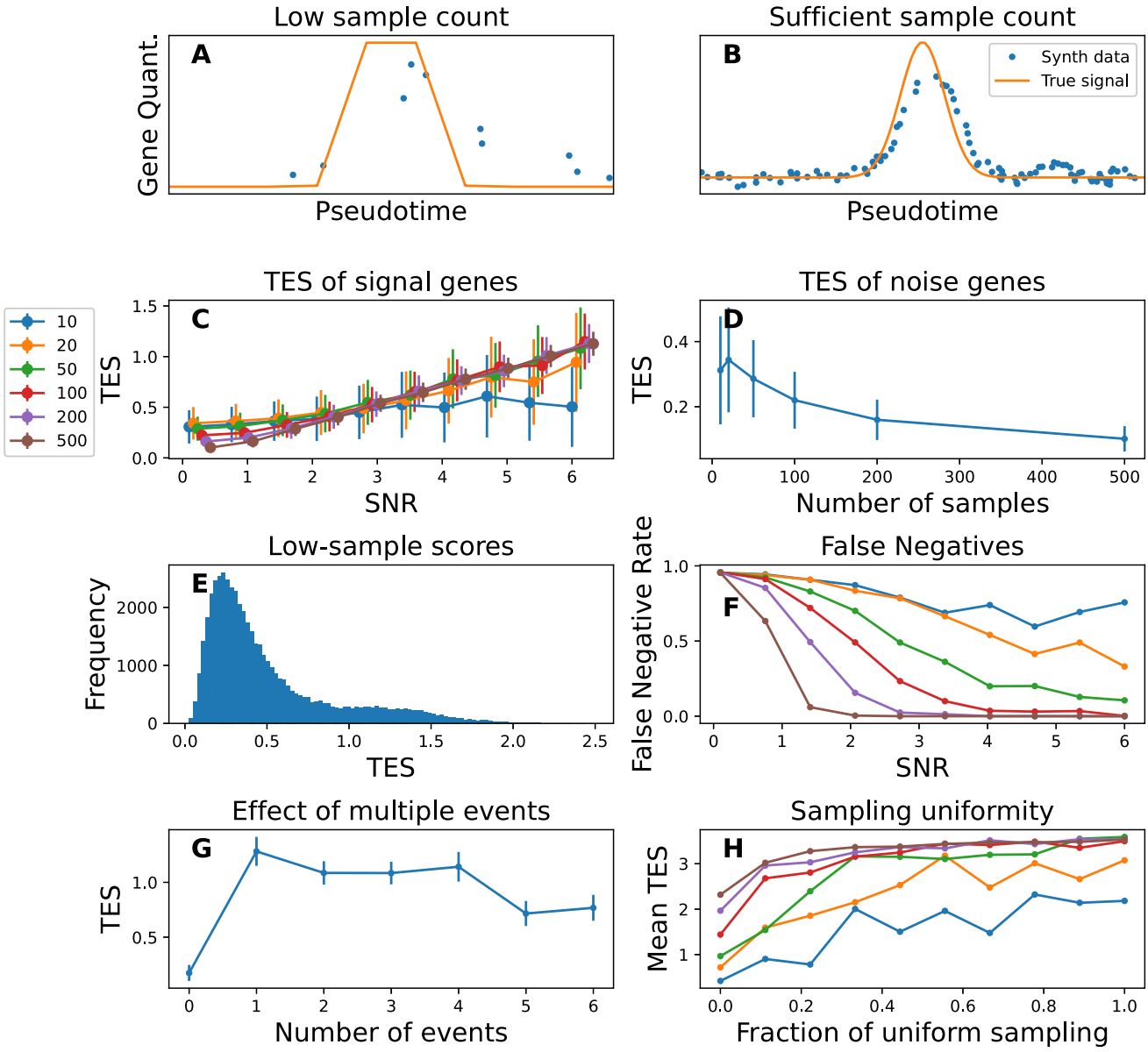


Figure 2. TE detection with the synthetic datasets. **(A, B)** The number of cell samples in a dataset impacts whether a transient change in expression is observable. **(C)** Evaluation of the impact of the number of samples and the SNR on the score given to signal genes. **(D)** The sample size directly affects the scores given to noise-only genes, establishing a noise floor for different sample counts. **(E)** For multiple random instantiations of a dataset, the distribution of scores for low-sample datasets produces a bimodal distribution with higher scores being predicted when samples appear during transient expression change. **(F)** The rate of signal genes being labeled as noise genes. The score threshold is set as two standard deviations above the mean noise gene score for a dataset. **(G)** The effect of genes demonstrating multiple events in a single pseudotime course is to decrease the overall score. **(H)** Rapid cell state transitions are modeled as a decrease in the uniformity of sampling across pseudotime. Starting on the left with a reduction to no samples obtained at the peak of the transition and moving to the right to be completely uniform, low-sample counts are greatly impacted.

with the first event, we see a general decreasing trend in the gene score as the number of TEs increases.

Duration of TEs

The pseudotime axis represents a measure of change in the gene expression and does not have a direct correspondence to time. Despite that, the timing of sample acquisition has an important effect on data acquired. We model the speed in transitions by the density of cell samples; faster transitions have fewer samples in the area of pseudotime where the signal is present. The duration of the TE is particularly impactful in cases where there are only a few samples. We reduce the probability density proportional to the magnitude of the signal, making it less likely to acquire samples located at the peak of the TE signal. Figure 2H shows the impact of the sampling density for different sample counts, with 0 indicating that no samples are found at the signal's peak, and 1 indicating that it is as likely as elsewhere in pseudotime (i.e. uniform distribution). We see a general increase for all sample counts as sampling becomes more uniform.

Comparison with other methods

Synthetic datasets were generated as previously described, and the data used a negative binomial distribution to meet trade-

Seq's assumptions. Data were preprocessed to be 0 mean before processing with scTransient. We compared the precision of scTransient, GPfates, and tradeSeq in identifying which genes contained TEs. For every dataset that was evaluated, scTransient either performed better or as well as tradeSeq, and performed comparably with GPfates. Figure 3 contains representative figures showing the performance of scTransient, GPfates, and tradeSeq for TEs of widths 0.01 and 0.005; runs for other parameters are included in the supplementary material. Each experiment was repeated 20 times, and the standard error is displayed around each point. For sufficiently large datasets ($n > 200$), performance across the methods converges.

The time to complete an analysis was tracked across sample counts ranging from 20–200 000 to examine the complexity of each approach empirically. As shown in Fig. 4, scTransient outperforms GPfates and tradeSeq by a wide margin. Experiments were halted if the execution time exceeded 6h or if they required more than 32GB of memory. Since scTransient uses Fourier-based methods, it has a time complexity of O(n log n), while GPfates requires the use of a Gaussian Process regression with a time complexity of $O(n^3)$. The generalized additive model used in tradeSeq appears to scale somewhere between the two.

Induced neuron dataset

We applied scTransient to a bulk proteomic dataset of iPSC-derived neurons at different maturation stages. Figure 5A shows two representative bright field images of NGN2-induced iPSC-derived neuronal cultures (iNeurons) maintained during 1, 2, 4, 6, or 8 weeks in culture (scale Bar, 40 μm) and also shows the experimental design used to profile proteomic dynamics during neuronal maturation. The iNeurons were generated using genome-edited KOLF2.1J iPSCs[25] harboring a doxycycline-inducible NGN2 cassette [23, 26] and were maintained under defined maturation conditions for up to 8 weeks. Neuronal cultures were processed for bulk proteomics to capture molecular events from early neuronal commitment through later stages of maturation. This strategy enabled longitudinal quantification of protein abundance changes across neuronal development. We then applied scTransient to our in-house-generated iNeuron time-course dataset, enabling the detection of transient molecular events (spiking proteins) that drive neuronal fate commitment and neuronal maturation of NGN2-induced iPSC cultures. Figure 5B demonstrates fluorescence for two proteins exclusive to neuronal development as confirmation of cell identity: CALB2 [27, 28] and MAP2 [29, 30]. As before, we used psupertime [3] to obtain pseudotime orderings for the acquired samples, and the distribution of samples along pseudotime are shown in Fig. 5C: we obtain distributions around the data acquisition timepoints (Weeks 1–8), ordered as expected. Figure 5D demonstrates the top six genes used for predicting pseudotime ordering. Positive values indicate that the gene pushes the pseudotime ordering to later pseudotime values; negative values for earlier ones. Each of the identified genes is known to be expressed along neuronal development: DCC [31], VCL [32, 33], TH [34, 35], PHF24 [36], AGPAT4 [37, 38], and ANXA2 [39–41]. The result is consistent with existing literature, and is a strong indication that the pseudotime ordering is biologically motivated. After windowing, clustering, and averaging protein trajectories within clusters, we obtain four representative trajectories shown in Fig. 5E. We note that there is a sample gap between the Week 1 and Week 2 samples (Fig. 5C), which makes interpretations of TEs in that area of pseudotime unreliable. For the three largest clusters (0, 1, 3), we see a transition across 0, indicating that the Week 1 and Week 2 samples have opposite Z-scores. Top-scoring proteins as defined by TES have their timecourses shown in Fig. 5F, which match the shape of TEs that scTransient aims to detect. Figure 5G shows representative examples of proteins that exhibit transient behavior across pseudotime based on their abundance (Z-score, *y*-axis). This includes early peaks with varying amplitudes (GRID1, NEFH) and increases at a later stage (GPRIN3, CALB1), suggesting that the time-course highlights specific expression programs occurring at distinct times during neuronal maturation. For instance, NEFH (Fig. 5H) shows a localized transient elevation in abundance at mid-psupertime, which would be difficult to identify from single punctual maturation events, as denoted by the grouped dots along psupertime (right panel, $n = 8$ technical replicates). Transient increases in GRID1 and NEFH indicate expression of glutamate receptors and mature neuronal markers involved in axonal growth in NGN2-induced iPSC neurons [42], supporting cortical excitatory identity of iNeuron cultures. Late-stage increases in GPRIN3 [43] and CALB1 [44] in iNeurons remain unexplored despite their high expression in the brain or may indicate expression of inhibitory neuronal markers. In support of the latter, we detected by immunofluorescence a low proportion of CALB2 + cells [27, 44] in iNeuron cultures at 4 weeks of maturation (Fig. 5B), suggesting the presence of inhibitory interneurons captured during neuronal maturation. In Fig. 5I, we evaluate whether TES ranking prioritizes proteins that drive biologically meaningful maturation programs. We performed GSEA on the TES-ranked protein list derived from proteins identified in the time course of iNeuron cultures and examined the cumulative recovery of pathway-leading proteins across the ranking. For each enriched pathway, the fraction of leading-edge proteins recovered was computed as progressively larger portions of the TES-ranked list were traversed. The diagonal baseline in the plot indicates random ordering. Results indicate that pathways associated with stem cell maintenance and negative regulation of developmental processes exhibited rapid leading-edge recovery within the top fraction of the TES ranking, indicating that their enrichment is driven by a small number of proteins with early temporally localized expression changes. These findings support the conversion of iPSCs into neurons by halting cell growth and facilitating neuronal fate commitment. In contrast, pathways related to translation, mitochondrial gene, and peptide biosynthetic expression showed more gradual recovery, consistent with actively expressed programs whose constituent proteins vary across neuronal maturation. Together, these results indicate that TES-based prioritization helps identify enrichment of proteins with stage-specific transient dynamics along psupertime.

Comparison with tradeSeq

We evaluated our dataset using tradeSeq for comparison. We computed pseudotime as with scTransient using psupertime. To meet the assumptions made by tradeSeq, protein abundance, and pseudotime values were shifted to be positive by subtracting the minimum value. Of the 5186 genes, 3239 produced *P*-values that were not NaN. Of those, only 33 had *P*-

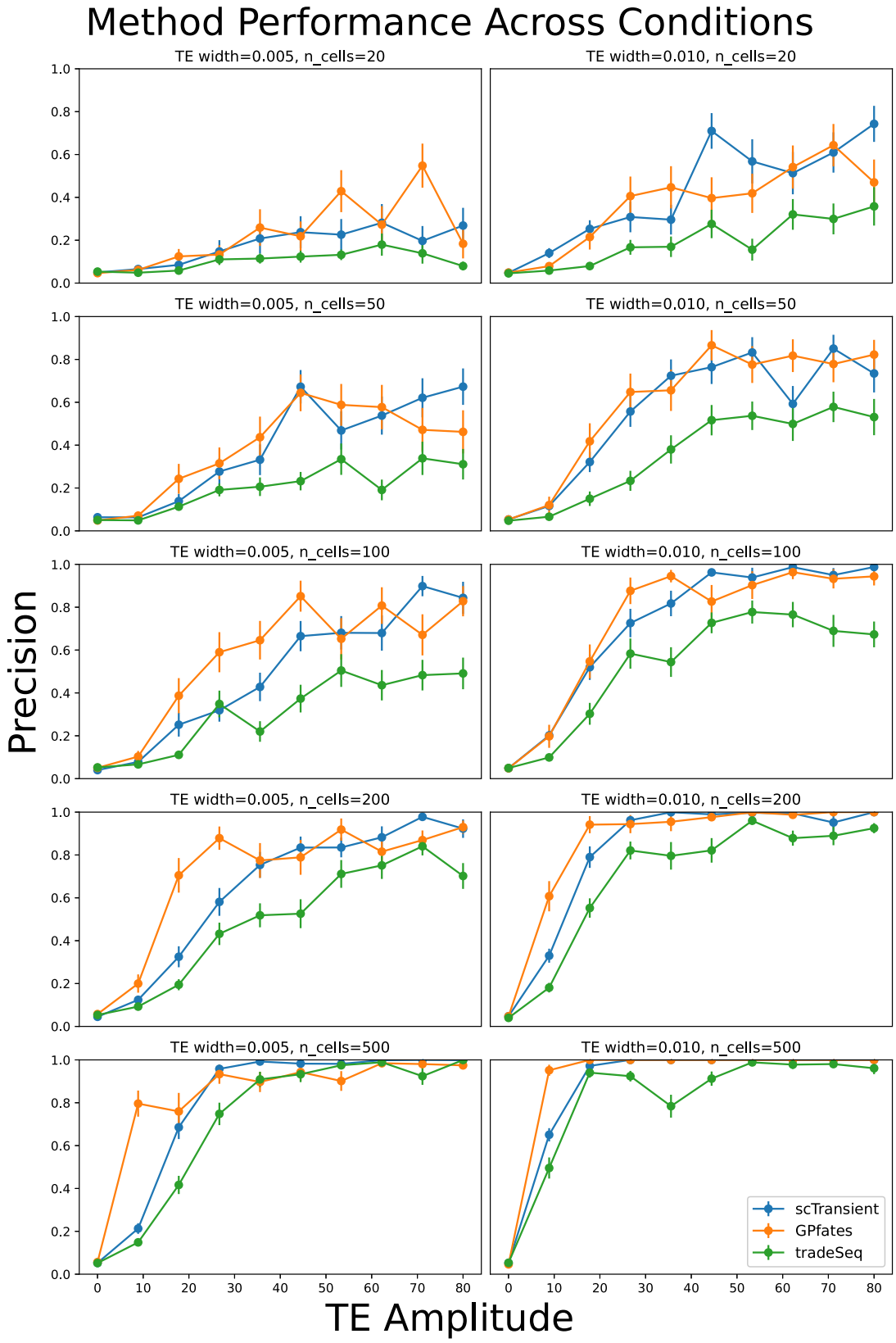


Figure 3. Performance of scTransient, GPfates, and tradeSeq across parameterized synthetic datasets. Datasets containing TEs with different properties (duration, amplitude) across multiple sample sizes were analyzed. Each dataset had 100 genes containing a TE and 1900 without. Counts were generated by sampling a negative binomial with r = 10, $P = 0.5$. The top 100 scoring genes for each method were used to determine how many of the TEs were detected by each method. Each experiment was repeated 20 times and the standard error is plotted around each point. Across conditions, scTransient and GPfates consistently outperform tradeSeq.

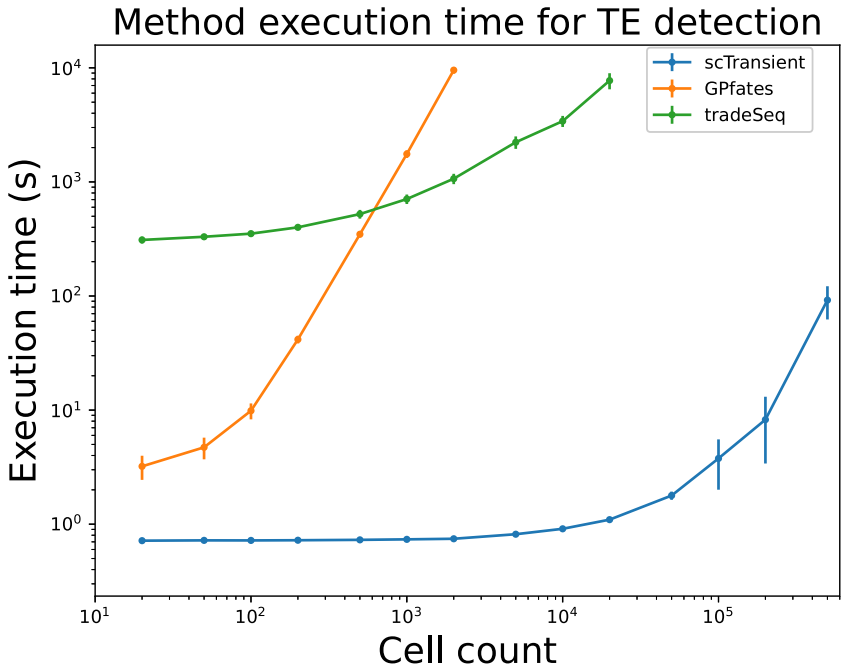


Figure 4. Observed time complexity of the methods used to detect TEs. Note the log–log axes. The missing points are due to experiments being halted after 6 h. scTransient outperforms other methods by orders of magnitude for modest dataset sizes (~1000), appears to be the only realistic option for larger datasets (>10 000) as the other methods either take longer than 6 h or have excessive memory requirements.

values <.05, and 162 had *P*-values <.1. As stated in the tradeSeq paper, the reported *P*-values are not statistically meaningful, so we retain the 162 genes with $P < .1$. Using those genes, we obtained the gene enrichment pathways from g:Profiler [45], shown in Table 1; although most are not specific to neuronal development, terms related to actin cytoskeleton development and neuron recognition indicate that the findings are consistent with development. The top-scoring genes from Fig. 5F all had *P*-values above 0.2, with half (GRID1, RETREG3, TMEFF1, INSR, LINGO1, TMEM245) having NaN values.

Cell cycle dataset

Having established that scTransient can reveal clear development programs in the control bulk proteomics iNeurons dataset, we expanded to a true single cell proteomics dataset from human lung adenocarcinoma A549 cells produced by Bubis *et al.* [46]. One of the advantages stated in the study is the ability to study the dependency between a cell's size and its cell cycle phase, making the data well-suited for evaluating our proposed method. Using default parameters for Leiden clustering, four clusters were detected for the dataset (Fig. 6A) that were not clearly associated with the proteome depth in those cells (Fig. 6B). We identify the cell cycle phase using Scanpy's *score_genes_cell_cycle* function that applies gene scoring as described in Seurat [47] by supplying known markers for S and G2/M phases, and labelling the rest as G1. The cell cycle showed coarse agreement with the Leiden clusters (Fig. 6C). We set ordinal labels of 1–3 for the G1, S, and G2/M phases and used psupertime [3] to obtain pseudotime orderings. Proteins found to be significant in predicting pseudotime based on the specified cell cycle phase are shown in Fig. 6D, and plotting the cell phase as a function of pseudotime produces the expected transitions between phases (Fig. 6E). The trajectory of some well known cell cycle proteins PCNA and UNG over pseudotime is consistent with their role in the S phase (Supplementary Fig. 1).

We apply windowing as previously described, and use scTransient to identify TEs. Although no clear pattern is apparent in the highest-scoring genes, obtaining ~200 genes and clustering them using kmeans ($k = 4$) clustering, we see two relevant patterns emerge. One cluster consists of genes ($n = 83$) spiking in the area of pseudotime (Fig. 6F) related to the S-phase, with functional gene enrichment analysis using g:Profiler [45, 48] (Table 2) showing links to DNA replication, RNA processing, and nucleus-related activity. The second cluster consists of genes ($n = 32$) genes spiking near the G1/S transition region (Fig. 6G), with gene enrichment showing genes related to extracellular space among others (Table 3). We note that there exists literature suggesting that components listed in Table 3 are related to the cell cycle in different cell types: zinc sequestering [49–51], arachidonate binding in macrophages [52], RAGE binding (via S100 ligands) [53], calprotectin in head and neck squamous cell carcinomas [54], but we are unable to experimentally verify that this is the case for the dataset. Of the two remaining clusters, one showed no significant enriched pathways, and the other had pathways related to NADPH and extracellular components (Table 4). We note that for the cluster containing 83 proteins, 22 of these did not have previously published literature supporting a connection to the cell cycle (Supplementary Table 1). Experimentally validating the impact of these proteins is beyond the scope of the current paper. The cell cycle genes used to order the samples could plausibly be identified as having transient expression, which could limit the relevance of the method; we explore this in the discussion section.

Comparison with tradeSeq

The cell cycle dataset from Bubis *et al.* [46] was used to evaluate which genes were identified by tradeSeq. Of the

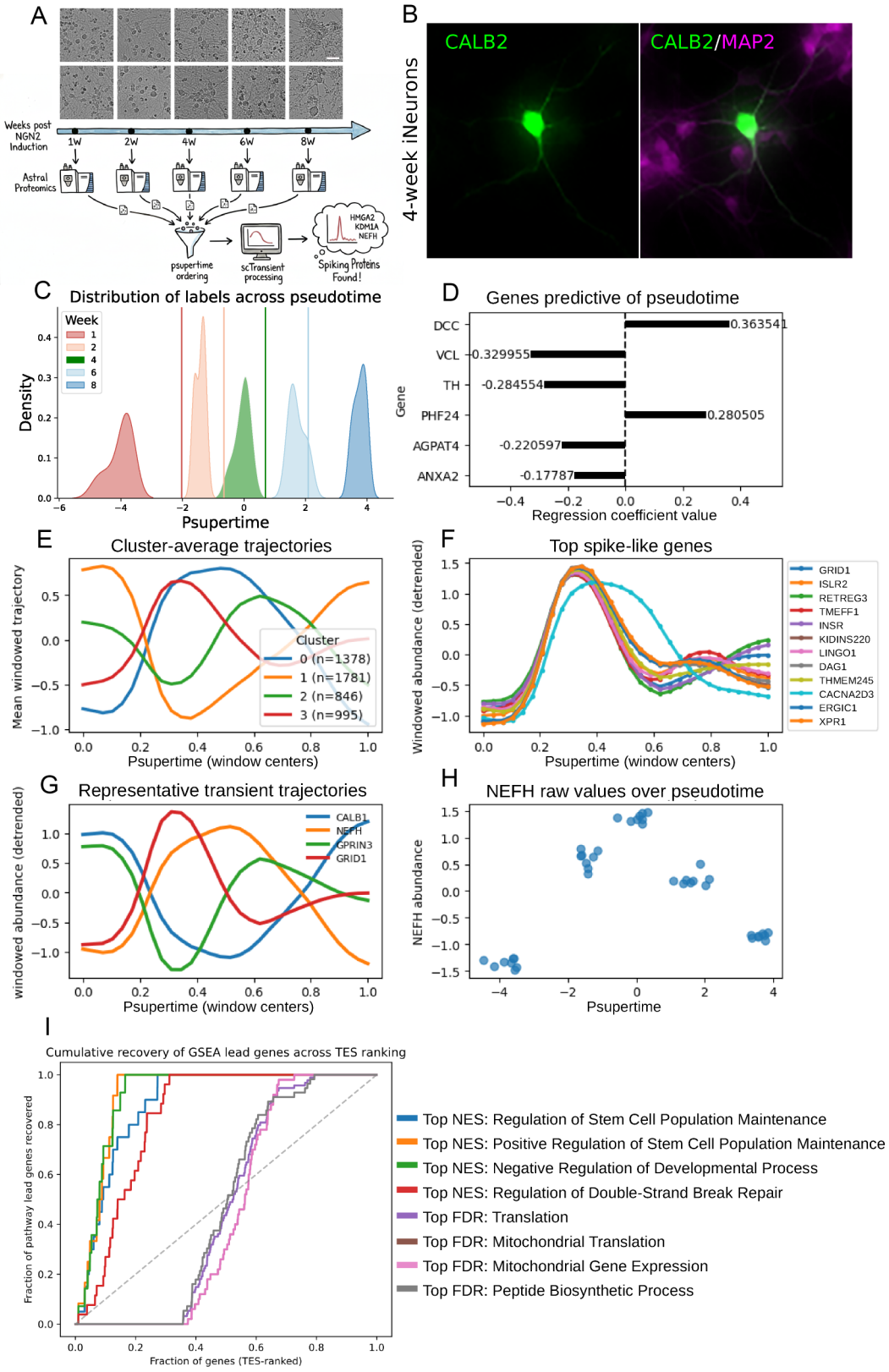


Figure 5. Experimental results for iNeuron datasets. **(A)** Experiment schematic. **(B)** CALB2 and MAP2 fluorescence, indicating the presence of inhibitory neurons. **(C)** Distribution of samples across pseudotime. **(D)** Genes predictive of pseudotime. **(E)** Cluster averages. **(F)** Trajectories of top-scored genes. **(G)** Selected genes from each cluster. **(H)** Sample NEFH abundance along pseudotime. **(I)** Using GSEA to show the identification of pathway genes using a TES-sorted genes.

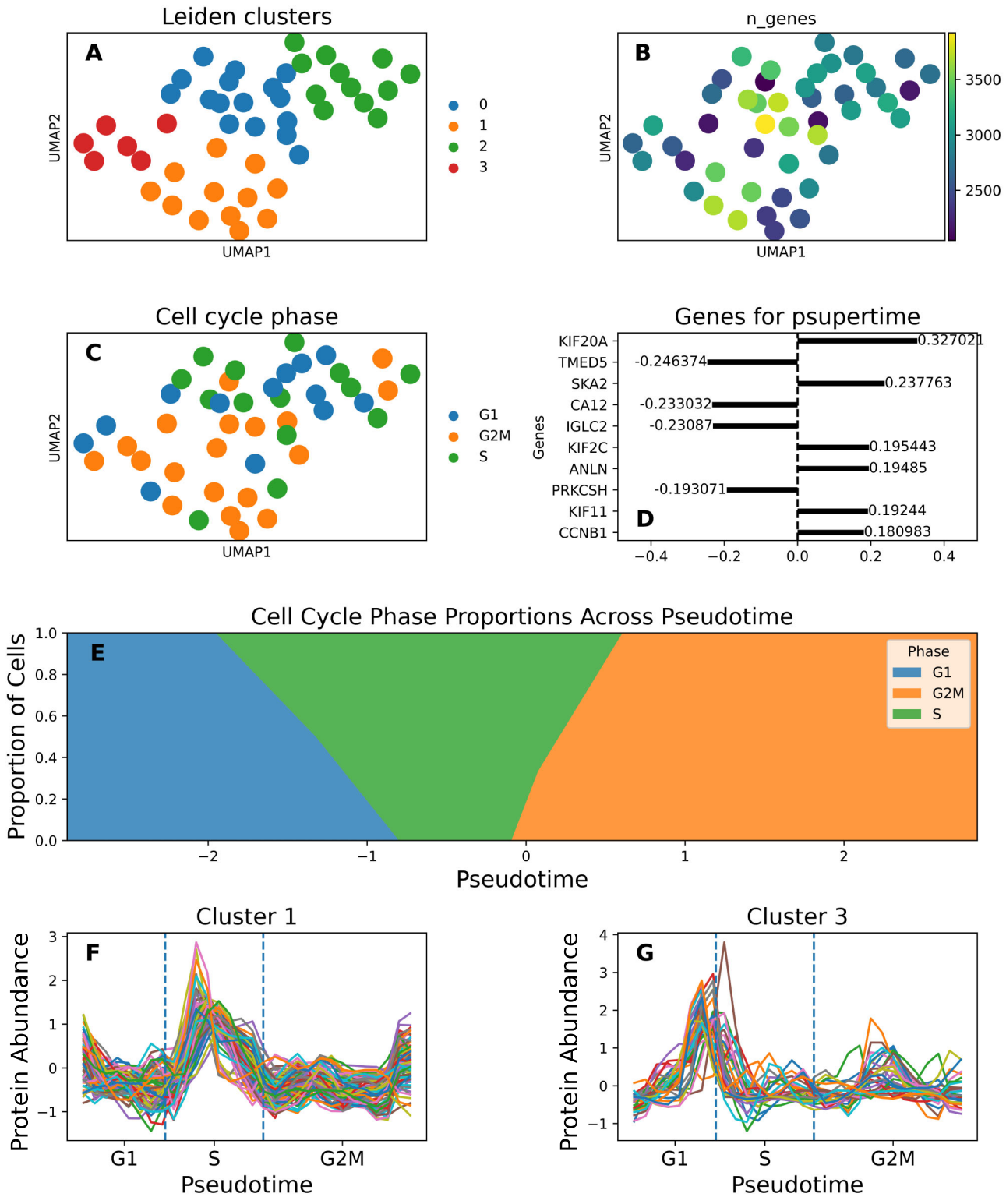


Figure 6. Analysis of the cell cycle dataset. **(A)** Clusters obtained by unsupervised clustering via Leiden algorithm overlaid on a 2D UMAP. **(B)** Number of genes with nonzero quantities overlaid on the same 2D UMAP as panel (A). **(C)** Cell cycle phase as determined by known biomarkers overlaid on the same 2D UMAP as panel (A). **(D)** Weighting of genes used by psupertime to order samples in pseudotime; these genes were found to be useful in predicting pseudotime ordering. **(E)** Proportions of cell cycle phase proportions across pseudotime. The G1, S, and G2 phases are in the expected order and there is a smooth transition between adjacent phases. **(F)** Pseudotimecourse for cluster 1, with increased protein quantities for cells in the S-phase. The enriched pathways listed in Table 1 are consistent with DNA replication. **(G)** Pseudotimecourse for cluster 3, showing increased protein quantities for cells near the end of the G1 phase.

Table 1. Term enrichment analysis of the genes from tradeSeq analysis of the neuron dataset

Source	Term ID	Term name	P_adj
GO:CC	GO:0 070 062	Extracellular exosome	4.909×10^{-8}
GO:CC	GO:0 005 737	Cytoplasm	6.918×10^{-8}
GO:MF	GO:0 003 779	Actin binding	9.601×10^{-6}
GO:MF	GO:0 050 839	Cell adhesion molecule binding	1.436×10^{-5}
GO:CC	GO:0 015 629	Actin cytoskeleton	1.793×10^{-5}
GO:CC	GO:0 030 054	Cell junction	2.048×10^{-4}
GO:BP	GO:0 030 029	Actin filament-based process	2.664×10^{-4}
GO:CC	GO:0 009 898	Cytoplasmic side of plasma membrane	9.772×10^{-4}
GO:BP	GO:0 008 038	Neuron recognition	1.955×10^{-3}
GO:CC	GO:0 099 080	Supramolecular complex	3.956×10^{-3}
GO:CC	GO:0 005 635	Nuclear envelope	7.348×10^{-3}
GO:CC	GO:0 005 684	U2-type spliceosomal complex	1.925×10^{-2}
GO:CC	GO:0 031 252	Cell leading edge	2.234×10^{-2}
GO:BP	GO:0 043 087	Regulation of GTPase activity	3.098×10^{-2}

4660 genes, 4381 were NaN and 279 produced numeric *P*-values. Only 23 of these had values below 0.5 (one half) and 96 were above 0.9. Since the values themselves are not statistically meaningful, we retain all genes with numeric *P*-values and used kmeans to cluster their expression signals (Supplementary Fig. 2). Whereas scTransient produces clusters with clear TEs, genes identified using tradeSeq had changes that were more diffuse along pseudotime. We quantify the signal spread using the mean kurtosis of the gene signals within each cluster; we exclude the first and last two points in pseudotime. The mean kurtosis for the genes identified by tradeSeq is lower than for genes found by scTransient. Using g:Profiler [45, 48] to perform functional gene enrichment, we find that one cluster from tradeSeq (cluster 0) had enriched nucleus-related pathways (Table 5). Of the 207 genes identified by scTransient and the 297 identified by tradeSeq, only 13 genes overlapped.

Discussion

Synthetic experiments

The first set of experiments examined the SNR and showed that, for a given sample count, the TES of genes containing a TE increased with increasing SNR (Fig. 2C). Increasing the sample count decreased the TES of noise genes, as expected. When the data is converted from samples along pseudotime into a signal, the averaging window considers points over a stretch of pseudotime; having additional points averages out the noise component, resulting in a flatter signal (which in turn reduces the Z_{mod} component of the heuristic). Although the results in Fig. 2C might suggest that a stronger signal can overcome low sample counts, the bimodal distribution in Fig. 2E confirms that this is not the case. For low-sample datasets, the improved score is driven by iterations in which samples happen to be at the TE; experimentally, this would translate to hoping that one of your cells happened to be in the middle of the process being investigated. In that case, we would not expect detectability to improve beyond some SNR threshold, which is consistent with the results shown in Fig. 2F. In that panel, the false negative rate is high for all signals when no signal is detectable (SNR = 0), shows an initial decrease as some instances have samples land in the right region, then tapers off to a minimum that is dependent on the number of samples. Taken together, these results confirm that using pseudotime ordering requires sufficient samples to derive reliable conclusions from the pseudotime signal. These results are based on multiple idealized conditions and should be treated as optimistic if the specified parameters are used to inform the design of biological experiments.

For biological events that could involve multiple TEs, we showed that the proposed heuristic exhibited a gradual decrease as the number of events increased, due to a corresponding decrease in the modified Z score (Fig. 2G). For the specified conditions, the overall score did not drop to the noise floor, but the combination of multiple conditions (lower SNR, fewer samples) might do so.

Lastly, transition time (sampling density during the TE) decreased the heuristic score, particularly for low-sample datasets (Fig. 2H). With fewer samples to describe the change in signal intensity, the peak in gene expression is less likely to be caught, which in turn decreases the wavelet coefficients and the overall score. The effect is most notable in low-sample datasets, where it drastically reduced the mean TES. There is a decreasing utility with the number of samples; even if a peak is not completely described, a transient change that can be reliably detected is not detected better with additional samples.

iNeuron dataset

Samples of neurons induced from iPSCs were acquired using bulk proteomics. Although pseudotime was developed for single-cell datasets, there is precedent for its application to bulk samples [55], and the underlying metrics are not adversely affected. The newly acquired dataset clearly demonstrates that scTransient can detect TEs, and that the detected events are in line with existing literature. The availability of neuron age labels allowed us to obtain the proteomic distribution of samples: without pseudotime, the samples would be strictly positioned at their time label instead of across pseudotime, which would in turn limit the possible resolution of TE detection.

Cell cycle dataset

The cell cycle dataset produced by Bubis *et al.* [46] was used to evaluate scTransient on a public dataset. Cell cycle markers were used to order the cells, which might improve their downstream correspondence to known pathways compared to unsupervised pseudotime. TEs were detected that corresponded with the main processes of the S phase (Fig. 6F and Table 2)

Table 2. Significant pathways for the cluster 1 of the genes found to have TEs in the cell cycle dataset

Source	Native	Name	*P*-value
GO:CC	GO:0 005 654	Nucleoplasm	7.96E-12
GO:CC	GO:0 031 981	Nuclear lumen	3.52E-10
GO:CC	GO:0 005 634	Nucleus	2.35E-09
GO:CC	GO:0 031 974	Membrane-enclosed lumen	7.11E-09
GO:CC	GO:0 043 233	Organelle lumen	7.11E-09
GO:CC	GO:0 070 013	Intracellular organelle lumen	7.11E-09
GO:MF	GO:0 003 723	**RNA binding**	6.66E-08
GO:CC	GO:0 043 231	Intracellular membrane-bounded organelle	1.19E-07
GO:MF	GO:0 140 640	**Catalytic activity, acting on a nucleic acid**	2.30E-07
GO:CC	GO:0 043 232	Intracellular membraneless organelle	3.95E-07
GO:CC	GO:0 043 228	Membraneless organelle	3.97E-07
GO:CC	GO:0 043 229	Intracellular organelle	7.75E-07
GO:CC	GO:0 005 622	Intracellular anatomical structure	8.78E-06
GO:CC	GO:0 043 227	Membrane-bounded organelle	1.40E-05
GO:MF	GO:0 003 676	**Nucleic acid binding**	1.42E-05
GO:CC	GO:0 043 226	Organelle	2.55E-05
GO:BP	GO:0 016 072	**Ribosomal RNA (rRNA) metabolic process**	0.000 137
GO:MF	GO:0 008 173	**RNA methyltransferase activity**	0.000 142
GO:MF	GO:0 004 386	**Helicase activity**	0.000 144
GO:MF	GO:0 140 098	**Catalytic activity, acting on RNA**	0.000 341
GO:CC	GO:0 005 730	Nucleolus	0.000 379
GO:BP	GO:0 006 364	**rRNA processing**	0.000 421
GO:MF	GO:0 097 159	Organic cyclic compound binding	0.000 687
GO:BP	GO:0 042 254	**Ribosome biogenesis**	0.000 749
GO:BP	GO:0 071 840	**Cellular component organization or biogenesis**	0.001 483
GO:BP	GO:0 001 510	**RNA methylation**	0.001 618
GO:CC	GO:0 035 770	Ribonucleoprotein granule	0.001 713
GO:MF	GO:0 005 515	Protein binding	0.005 665
GO:BP	GO:0 090 304	**Nucleic acid metabolic process**	0.007 752
GO:BP	GO:0 009 451	**RNA modification**	0.009 092
GO:CC	GO:0 036 464	Cytoplasmic ribonucleoprotein granule	0.010 392
GO:CC	GO:0 030 684	Preribosome	0.011 902
GO:CC	GO:0 140 513	Nuclear protein-containing complex	0.01 326
GO:CC	GO:0 005 829	Cytosol	0.018 899
GO:BP	GO:0 043 414	Macromolecule methylation	0.021 186
GO:BP	GO:0 043 170	Macromolecule metabolic process	0.026 077
GO:BP	GO:0 006 260	**DNA replication**	0.026 904
GO:CC	GO:0 098 687	**Chromosomal region**	0.027 403
GO:BP	GO:0 006 139	Nucleobase-containing compound metabolic process	0.028 182
GO:CC	GO:0 032 991	Protein-containing complex	0.031 968
GO:MF	GO:0 008 757	S-adenosylmethionine-dependent methyltransferase activity	0.033 176
GO:CC	GO:0 005 663	**DNA replication factor C complex**	0.033 592
GO:MF	GO:0 140 097	**Catalytic activity, acting on DNA**	0.036 054
GO:CC	GO:0 032 040	Small-subunit processome	0.037 421
GO:BP	GO:0 007 275	Multicellular organism development	0.037 504
GO:CC	GO:0 099 080	Supramolecular complex	0.041 997
GO:MF	GO:0 016 787	Hydrolase activity	0.047 276
GO:BP	GO:0 006 401	**RNA catabolic process**	0.049 419

and the end of the G1 phase (Fig. 6G and Table 3). Detecting groups of genes that show transient expression changes aligned with cell cycle phases suggests the potential to use pseudotime to detect TEs across different biological processes. We note that the cell cycle dataset was acquired using sensitive instrumentation (Astral, 2023) that provided a deep profile and a reliable signal (higher SNR). Comparing these to the scTransient results obtained for other datasets, it is unclear whether the detectability of TEs requires high-sensitivity instruments or if the difference is inherent to biological processes.

Since the cell cycle phase was determined using a set of known cell cycle markers, it is reasonable to suspect that the same genes identified by scTransient could be the ones responsible. Of the 97 genes known to be relevant to the cell cycle [56], 63 were present in the dataset. Clustering the pseudotimecourses of those genes into four groups using kmeans clustering [57], we show the groupings in Supplementary Fig. 3 and note that they do not have the increase in late G1 or throughout S as seen as in the clusters of the genes identified by scTransient.

Comparison with other methods

Two sets of results were produced for comparing scTransient, GPfates, and tradeSeq: one using synthetic data, and another

Table 3. Significant pathways for the cluster 3 of the genes found to have TEs in the cell cycle dataset

Source	Native	Name	*P*-value
GO:CC	GO:0 070 062	Extracellular exosome	3.46E-05
GO:CC	GO:1 903 561	Extracellular vesicle	3.98E-05
GO:CC	GO:0 065 010	Extracellular membrane-bounded organelle	4.00E-05
GO:CC	GO:0 043 230	Extracellular organelle	4.00E-05
GO:CC	GO:1 990 660	**Calprotectin complex**	3.72E-04
GO:BP	GO:0 051 238	Sequestering of metal ion	5.03E-04
GO:CC	GO:0 031 982	Vesicle	5.39E-04
GO:CC	GO:0 005 615	Extracellular space	1.38E-03
GO:BP	GO:0 070 488	Neutrophil aggregation	3.28E-03
GO:BP	GO:0 032 119	**Sequestering of zinc ion**	3.28E-03
GO:MF	GO:0 035 662	Toll-like receptor 4 binding	5.46E-03
GO:BP	GO:0 002 523	Leukocyte migration involved in inflammatory response	5.52E-03
GO:CC	GO:0 034 774	Secretory granule lumen	1.38E-02
GO:CC	GO:0 060 205	Cytoplasmic vesicle lumen	1.44E-02
GO:CC	GO:0 031 983	Vesicle lumen	1.47E-02
GO:MF	GO:0 050 544	Arachidonate binding	1.91E-02
GO:MF	GO:0 050 543	Icosatetraenoic acid binding	0.019 062
GO:MF	GO:0 050 542	Icosanoid binding	0.025 392
GO:CC	GO:0 005 576	Extracellular region	0.028 535
GO:MF	GO:0 050 786	**RAGE receptor binding**	0.04 073
GO:MF	GO:0 030 234	Enzyme regulator activity	0.04 203

Table 4. Significant pathways for the cluster 2 of the genes found to have TEs in the cell cycle dataset

Source	Native	Name	*P*-value
GO:MF	GO:0 004 090	Carbonyl reductase (NADPH) activity	1.79E-02
GO:CC	GO:0 070 062	Extracellular exosome	2.32E-02
GO:CC	GO:1 903 561	Extracellular vesicle	2.55E-02
GO:CC	GO:0 043 230	Extracellular organelle	2.56E-02
GO:CC	GO:0 065 010	Extracellular membrane-bounded organelle	2.56E-02

using the cell cycle dataset. In the case of the synthetic data where ground truth is known, our results show that scTransient and GPfates generally perform better in identifying genes with TEs than tradeSeq, but that all three methods perform comparably well with sufficiently large datasets, though both GPfates and tradeSeq quickly require excessive computation time with increasing sample size. For smaller datasets, GPfates performs better for low-amplitude TEs.

Using newly acquired data from induced neuronal cells, we demonstrated scTransient's ability to detect TEs throughout neuronal development. The top-scoring genes were consistent with those known to be associated with neuronal development, and examining pathway genes identified by sorted TES values (Fig. 5I) shows that the top-scoring genes are strongly associated with stem cell regulation. The comparison with tradeSeq further indicates that scTransient is able to better-identify TEs.

The cell cycle dataset results summarized in Fig. 6 show that scTransient is able to identify genes that have TEs and that they can be grouped together to demonstrate peaks at different points in the biological process (end of the G1 phase, throughout the S phase). Groups of genes identified by tradeSeq do not display clear peaks along pseudotime, which is reflected by their kurtosis: values closer to 0 indicate signal distributions closer to the Gaussian distribution which would be consistent with noise. Given that both methods identified over 200 genes but had only 13 in common, we do not expect the kurtosis values to be a function of any particular grouping. We note that the minimum *P*-value produced by tradeSeq was ∼0.17, and that more than a third of the numeric *P*-values were above 0.9. For users that do not realize that the *P*-values do not represent the likelihood of the result belonging to the null distribution, the typical cutoff of $P = 0.05$ would result in none of the genes being identified as having TEs.

As shown in Fig. 4, one of the main benefits of scTransient is its execution time. scTransient can process 200k cells faster than either GPfates or tradeSeq can process 500. Since scTransient uses wavelets whose transforms are computed using Fourier transforms, it is able to scale exceptionally well to large datasets. Scaling experiments were stopped when they required more than 6h or required more than 32GB of memory.

Hypothesis testing

Although the TES is designed as a ranking statistic rather than a formal hypothesis test, gene-specific *P*-values can be obtained by constructing an empirical null distribution that preserves the structure of each gene's expression signal while breaking its alignment to the inferred biological progression. One straightforward approach is to repeatedly permute the pseudotime ordering of cells, recompute the windowed pseu-

Table 5. Significant pathways for cluster 0 of the genes identified by tradeSeq in the cell cycle dataset

Source	Native	Name	*P*-value
GO:CC	GO:0 031 981	Nuclear lumen	2.43E-10
GO:CC	GO:0 043 233	Organelle lumen	3.08E-10
GO:CC	GO:0 070 013	Intracellular organelle lumen	3.08E-10
GO:CC	GO:0 031 974	Membrane-enclosed lumen	3.08E-10
GO:CC	GO:0 005 654	Nucleoplasm	5.22E-10
GO:MF	GO:0 003 723	RNA binding	1.35E-09
GO:CC	GO:0 070 062	Extracellular exosome	3.34E-09
GO:CC	GO:1 903 561	Extracellular vesicle	4.47E-09
GO:CC	GO:0 043 230	Extracellular organelle	4.53E-09
GO:CC	GO:0 065 010	Extracellular membrane-bounded organelle	4.53E-09
GO:CC	GO:0 005 829	Cytosol	9.16E-08
GO:MF	GO:0 097 159	Organic cyclic compound binding	9.66E-08
GO:MF	GO:0 003 676	Nucleic acid binding	2.03E-07
GO:BP	GO:0 042 254	Ribosome biogenesis	6.46E-07
GO:CC	GO:0 032 991	Protein-containing complex	1.81E-06
GO:CC	GO:0 005 615	Extracellular space	3.41E-06
GO:CC	GO:0 005 622	Intracellular anatomical structure	4.14E-06
GO:CC	GO:0 022 626	Cytosolic ribosome	4.39E-06
GO:CC	GO:0 043 227	Membrane-bounded organelle	4.76E-06
GO:CC	GO:0 031 982	Vesicle	1.06E-05
GO:MF	GO:0 017 111	Ribonucleoside triphosphate phosphatase activity	2.62E-05
GO:BP	GO:0 002 181	Cytoplasmic translation	2.78E-05
GO:BP	GO:0 048 024	Regulation of mRNA splicing, via spliceosome	3.29E-05
GO:MF	GO:0 140 657	ATP-dependent activity	5.39E-05
GO:CC	GO:0 043 231	Intracellular membrane-bounded organelle	5.50E-05
GO:MF	GO:0 016 887	ATP hydrolysis activity	5.66E-05
GO:MF	GO:0 045 296	Cadherin binding	6.81E-05
GO:MF	GO:0 016 462	Pyrophosphatase activity	7.50E-05
GO:MF	GO:0 016 818	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	7.63E-05
GO:MF	GO:0 016 817	Hydrolase activity, acting on acid anhydrides	7.63E-05
GO:CC	GO:0 043 226	Organelle	9.41E-05
GO:CC	GO:0 005 634	Nucleus	0.00 011 802
GO:BP	GO:0 050 684	Regulation of mRNA processing	0.00 012 376
GO:CC	GO:0 043 229	Intracellular organelle	0.00 014 362
GO:CC	GO:0 005 576	Extracellular region	0.00 014 852
GO:CC	GO:0 022 627	Cytosolic small ribosomal subunit	0.00 019 809
GO:CC	GO:0 044 391	Ribosomal subunit	0.00 020 235
GO:CC	GO:0 005 737	Cytoplasm	0.00 038 798
GO:CC	GO:0 043 232	Intracellular membraneless organelle	0.0 004 907
GO:CC	GO:0 043 228	Membraneless organelle	0.00 049 298
GO:BP	GO:0 071 840	Cellular component organization or biogenesis	0.00 050 337
GO:CC	GO:0 098 687	Chromosomal region	0.00 087 814
GO:CC	GO:0 005 925	Focal adhesion	0.00 134 782
GO:BP	GO:0 043 484	Regulation of RNA splicing	0.00 157 119
GO:BP	GO:0 006 412	Translation	0.00 163 085
GO:CC	GO:0 030 055	Cell–substrate junction	0.00 165 532
GO:MF	GO:0 050 839	Cell adhesion molecule binding	0.00 166 669
GO:MF	GO:0 004 386	Helicase activity	0.00 226 405
GO:MF	GO:0 000 166	Nucleotide binding	0.00 237 746
GO:BP	GO:0 090 329	Regulation of DNA-templated DNA replication	0.00 254 835
GO:MF	GO:0 060 090	Molecular adaptor activity	0.00 271 763
GO:MF	GO:1 901 265	Nucleoside phosphate binding	0.00 274 435
GO:MF	GO:0 005 515	Protein binding	0.00 336 129
GO:CC	GO:0 015 935	Small ribosomal subunit	0.00 370 444
GO:MF	GO:0 044 389	Ubiquitin-like protein ligase binding	0.0 041 149
GO:MF	GO:0 019 843	rRNA binding	0.00 422 516
GO:BP	GO:0 042 274	Ribosomal small subunit biogenesis	0.00 593 403
GO:BP	GO:1 903 311	Regulation of mRNA metabolic process	0.00 617 097
GO:CC	GO:0 030 054	Cell junction	0.00 640 336
GO:MF	GO:1 901 363	Heterocyclic compound binding	0.00 758 671
GO:MF	GO:0 035 639	Purine ribonucleoside triphosphate binding	0.00 797 215
GO:CC	GO:0 034 774	Secretory granule lumen	0.00 883 716
GO:BP	GO:0 080 090	Regulation of primary metabolic process	0.00 902 492
GO:CC	GO:0 060 205	Cytoplasmic vesicle lumen	0.00 944 068
GO:CC	GO:0 031 983	Vesicle lumen	0.00 964 937
GO:MF	GO:0 032 555	Purine ribonucleotide binding	0.01 163 418
GO:MF	GO:0 032 553	Ribonucleotide binding	0.01 329 471
GO:BP	GO:0 044 238	Primary metabolic process	0.01 473 704
GO:BP	GO:1 901 798	Positive regulation of signal transduction by p53 class mediator	0.01 494 761

Table 5. Continued

Source	Native	Name	*P*-value
GO:CC	GO:0 030 684	Preribosome	0.01 681 041
GO:MF	GO:0 097 367	Carbohydrate derivative binding	0.01 950 646
GO:CC	GO:0 042 470	Melanosome	0.01 999 278
GO:CC	GO:0 048 770	Pigment granule	0.01 999 278
GO:MF	GO:0 031 625	Ubiquitin protein ligase binding	0.02 069 957
GO:MF	GO:0 016 787	Hydrolase activity	0.02 306 334
GO:MF	GO:0 017 076	Purine nucleotide binding	0.02 382 525
GO:BP	GO:0 006 403	RNA localization	0.02 599 176
GO:MF	GO:0 140 640	Catalytic activity, acting on a nucleic acid	0.02 691 313
GO:MF	GO:0 044 877	Protein-containing complex binding	0.03 152 197
GO:CC	GO:1 904 813	Ficolin-1-rich granule lumen	0.03 238 097
GO:MF	GO:0 043 021	Ribonucleoprotein complex binding	0.03 497 318
GO:CC	GO:0 045 202	Synapse	0.03 501 121
GO:BP	GO:0 006 996	Organelle organization	0.04 203 329
GO:CC	GO:0 070 161	Anchoring junction	0.04 497 672
GO:CC	GO:1 990 904	Ribonucleoprotein complex	0.04 916 998

dotime signal and its wavelet coefficients for each permutation, and recalculate the TES for that gene. Repeating this procedure 100–1000 times yields a gene-specific null distribution of TES values expected under the absence of a coherent TE aligned to pseudotime. The observed TES can then be assigned an empirical *P*-value based on its percentile within this null distribution. This strategy directly tests whether the observed transient signal is stronger than would be expected from noise structure alone, while remaining agnostic to the absolute scale of TES across genes. More conservative variants could restrict permutations to local blocks of pseudotime to preserve sample density and autocorrelation or alternatively use circular shifts of the pseudotime signal rather than full randomization. In practice, such permutation-based calibration is computationally tractable due to the efficiency of the wavelet transformation and provides a principled way to complement TES-based ranking with statistical confidence estimates when required.

Suitable datasets

scTransient assumes that cells (or samples) can be meaningfully ordered along a single, continuous progression so that each feature's abundance can be interpreted as a function of pseudotime rather than as unrelated heterogeneity. In practice, this means the dataset should satisfy the core trajectory-inference prerequisites and validation checks described in our trajectory-inference review (Hutton and Meyer), including evidence that the inferred ordering reflects a real biological process rather than batch structure or unrelated variation [58].

The underlying biology should plausibly trace a continuous manifold between endpoints (for example, differentiation, activation, cell-cycle progression, maturation), with sufficiently dense sampling across the transition so that short-lived spikes are not missed. Sparse coverage, large gaps between states, or strong multimodal mixtures of unrelated processes will tend to produce unstable pseudotime signals and inflate false positives from windowing and wavelet coefficients. Datasets with strong branching can still be usable, but only if a single lineage is isolated (or cell weights/lineage assignments are used), so that pseudotime reflects a single dominant path rather than an averaged mixture.

Because transient-event detection is only meaningful if pseudotime corresponds to the intended process, suitable datasets should include orthogonal process information to confirm the ordering. This can include experimental time labels (true time course), known phase labels (cell cycle), lineage tracing, RNA velocity consistency, or independent markers that vary monotonically with progression (surface markers, morphology, reporter intensity, protein immunofluorescence, EdU incorporation, etc.). Following best practices in trajectory inference, these anchors should be used to check directionality, continuity, and robustness of the ordering across parameterizations.

scTransient is explicitly designed to surface transient spikes, so a dataset is most convincing when at least one of the following is true:

1. The process has expected transient regulators (canonical spiking genes or proteins) that provide a partial ground truth for benchmarking, or
2. There is an independent readout that can validate candidate spikes without relying on the same omics modality (for example, imaging, flow cytometry, targeted proteomics, perturbation-response timing, reporter assays, chromatin accessibility at known transient enhancers, or functional assays tied to a narrow window of progression). Without known spiking features or orthogonal validation, it becomes difficult to distinguish true transient programs from noise, windowing artifacts, or pseudotime uncertainty.

Finally, the dataset should have adequate per-cell (or per-sample) depth and signal-to-noise for the modality under analysis, and the experimental design should minimize batch effects that could be misinterpreted as progression. When technical covariates are unavoidable, they should be modeled or regressed prior to trajectory inference so that ordering is not driven by library size, missingness, or run-order effects.

Conclusion

scTransient bridges a critical gap in single-cell analytics by turning trajectory inference from a descriptive ordering tool into a quantitative detector of fleeting regulatory programs, and it can do so at scale. Across synthetic benchmarks and a high-depth public dataset, our wavelet-based TES consistently highlighted short-lived surges in gene or protein expression that map to process-specific pathways yet were over-

looked by standard differential-expression approaches. Because scTransient is deployed on the PSCS web platform, researchers can now explore TEs without local installation, iteratively tune parameters, and share reproducible workflows. Future extensions—including adaptive windowing, refined TES heuristics, and compatibility with multimodal atlases—promise to widen the landscape of discoverable biological TEs. We anticipate that charting these transient signals will reveal new regulatory nodes and therapeutic entry points in development, immunology, cancer, and beyond.

Acknowledgements

The authors thank Lahiruni M. Halwatura for collecting the iNeuron proteomics data.

Author contributions: Alexandre Hutton (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Software [lead], Validation [lead], Visualization [lead], Writing – original draft [lead], Writing – review & editing [lead]), Jose de Jesús Muñoz-Estrada (Data curation [supporting], Investigation [supporting], Validation [supporting], Writing – review & editing [supporting]), Jesse Gerard Meyer (Conceptualization [equal], Data curation [supporting], Formal analysis [supporting], Funding acquisition [lead], Investigation [supporting], Methodology [supporting], Project administration [lead], Resources [lead], Software [supporting], Supervision [lead], Validation [supporting], Visualization [supporting], Writing – original draft [supporting], Writing – review & editing [supporting])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was funded in part by the NIGMS (R35GM142502). Funding to pay the Open Access publication charges for this article was provided by startup funds from Cedars-Sinai Medical Center.

Data availability

Datasets were downloaded from external sources: the cell cycle dataset [46] is available through the ProteomeXchange Consortium [20] via PRIDE [21] with ID PXD049412. The iNeuron raw and processed data is available on massive under ID MSV000100760 with DOI: https://doi.org/doi:10.25345/C5FQ9QK2G. The synthetic data generation code is available from: https://github.com/xomicsdatascience/scTransient_notebooks.

The scTransient code is available from Zenodo and the PSCS [19] interface includes scTransient analysis capabilities. The notebooks for data analysis to produce the figures in this paper are on Zenodo.

References

1. Otto DJ, Jordan C, Dury B *et al.* Quantifying cell-state densities in single-cell phenotypic landscapes using Mellon. *Nat Methods* 2024;21:1185–95. https://doi.org/10.1038/s41592-024-02302-w
2. Martinson WD, Volkening A, Schmidtchen M *et al.* Linking discrete and continuous models of cell birth and migration. 2023. https://doi.org/10.48550/ARXIV.2308.16093
3. Macnair W, Gupta R, Claassen M. psupertime: supervised pseudotime analysis for time-series single-cell RNA-seq data. *Bioinformatics* 2022;38:i290–8. https://doi.org/10.1093/bioinformatics/btac227
4. Grossmann A, Kronland-Martinet R, Morlet J. Reading and understanding continuous wavelet transforms. In: Combes J-M, Grossmann A, Tchamitchian P. (eds), *Wavelets, inverse problems and theoretical imaging*. Berlin, Heidelberg: Springer Berlin Heidelberg.1990, pp. 2–20.
5. Hsiao CJ, Tung P, Blischak JD *et al.* Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome Res* 2020;30:611–21. https://doi.org/10.1101/gr.247759.118
6. Van Den Berge K, Roux De Bézieux H, Street K *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* 2020;11:1201. https://doi.org/10.1038/s41467-020-14766-3
7. Lönnberg T, Svensson V, James KR *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves T_H 1/T_{FH} fate bifurcation in malaria. *Sci Immunol* 2017;2:eaal2192. https://doi.org/10.1126/sciimmunol.aal2192
8. Song D, Li JJ. PseudotimeDE: inference of differential gene expression along cell pseudotime with well-calibrated p-values from single-cell RNA sequencing data. *Genome Biol* 2021;22:124. https://doi.org/10.1186/s13059-021-02341-y
9. Fischer DS, Theis FJ, Yosef N. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res* 2018;46:e119. https://doi.org/10.1093/nar/gky675
10. Qiu X, Mao Q, Tang Y *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14:979–82. https://doi.org/10.1038/nmeth.4402
11. Cao J, Spielmann M, Qiu X *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 2019;566:496–502. https://doi.org/10.1038/s41586-019-0969-x
12. Moran PAP. A test for the serial independence of residuals. *Biometrika* 1950;37:178–81. https://doi.org/10.1093/biomet/37.1-2.178
13. Hirotugu A. Information theory and an extension of the maximum likelihood principle. *2nd Int Symp Inf Theory* 1973, 199–213.
14. Hensman J, Rattray M, Lawrence N. Fast variational inference in the conjugate exponential family. In: Pereira F., Burges CJ, Bottou L., Weinberger KQ. (eds), *Advances in Neural Information Processing Systems*. Vol. 25, La Jolla, California, United States: Curran Associates, Inc., 2012.
15. Hensman J, Rattray M, Lawrence ND. Fast nonparametric clustering of structured time-series. *IEEE Trans Pattern Anal Mach Intell* 2015;37:383–93. https://doi.org/10.1109/TPAMI.2014.2318711
16. Hensman J, Lawrence ND, Rattray M. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics* 2013;14:252. https://doi.org/10.1186/1471-2105-14-252
17. Lázaro-Gredilla M, Van Vaerenbergh S, Lawrence ND. Overlapping mixtures of Gaussian processes for the data association problem. *Pattern Recognit* 2012;45:1386–95. https://doi.org/10.1016/j.patcog.2011.10.004
18. Harris CR, Millman KJ, Van Der Walt SJ *et al.* Array programming with NumPy. *Nature* 2020;585:357–62. https://doi.org/10.1038/s41586-020-2649-2
19. Hutton A, Ai L, Meyer JG. PSCS: unified sharing of single-cell omics data, analyses, and results. *J Proteome Res* 2025;24:4825–30. https://doi.org/10.1021/acs.jproteome.5c00178

20. Deutsch EW, Bandeira N, Perez-Riverol Y *et al.* The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res* 2023;51:D1539–48. https://doi.org/10.1093/nar/gkac1040
21. Perez-Riverol Y, Bandla C, Kundu DJ *et al.* The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res* 2025;53:D543–53. https://doi.org/10.1093/nar/gkae1011
22. Kowalczyk MS, Tirosh I, Heckl D *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* 2015;25:1860–72. https://doi.org/10.1101/gr.192237.115
23. Muñoz-Estrada J, Mostafania A, Halwatura L *et al.* Extended NGN2 expression in iPSCs dramatically enhances purity of neuronal cultures. *bioRxiv* 2025. https://doi.org/10.1101/2025.06.20.660764
24. Hughes CS, Moggridge S, Müller T *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* 2019;14:68–85. https://doi.org/10.1038/s41596-018-0082-x
25. Pantazis CB, Yang A, Lara E *et al.* A reference human induced pluripotent stem cell line for large-scale collaborative studies. *Cell Stem Cell* 2022;29:1685–1702.e22. https://doi.org/10.1016/j.stem.2022.11.004
26. Fernandopulle MS, Prestil R, Grunseich C *et al.* Transcription factor-mediated differentiation of human iPSCs into neurons. *Curr Protoc Cell Biol* 2018;79:e51. https://doi.org/10.1002/cpcb.51
27. Kobayashi Y, Amaral DG. Chapter IV – Chemical neuroanatomy of the hippocampal formation and the perirhinal and parahippocampal cortices. In: Bloom F. E., Björklund A., Hökfelt T.. (eds), *Handbook of Chemical Neuroanatomy*. Vol. 15, Amsterdam, Netherlands, Elsevier, 1999, 285–401.
28. Gutierrez-Mecinas M, Davis O, Polgár E *et al.* Expression of calretinin among different neurochemical classes of interneuron in the superficial dorsal horn of the mouse spinal cord. *Neuroscience* 2019;398:171–81. https://doi.org/10.1016/j.neuroscience.2018.12.009
29. Soltani MH, Pichardo R, Song Z *et al.* Microtubule-associated protein 2, a marker of neuronal differentiation, induces mitotic defects, inhibits growth of melanoma cells, and predicts metastatic potential of cutaneous melanoma. *Am J Pathol* 2005;166:1841–50. https://doi.org/10.1016/S0002-9440(10)62493-5
30. DeGiosio RA, Grubisha MJ, MacDonald ML *et al.* More than a marker: potential pathogenic functions of MAP2. *Front Mol Neurosci* 2022;15:974890. https://doi.org/10.3389/fnmol.2022.974890
31. Horn KE, Glasgow SD, Gobert D *et al.* DCC expression by neurons regulates synaptic plasticity in the adult brain. *Cell Rep* 2013;3:173–85. https://doi.org/10.1016/j.celrep.2012.12.005
32. Mandal P, Belapurkar V, Nair D *et al.* Vinculin-mediated axon growth requires interaction with actin but not talin in mouse neocortical neurons. *Cell Mol Life Sci* 2021;78:5807–26. https://doi.org/10.1007/s00018-021-03879-7
33. Liu L, Wang X, Liang M *et al.* Vinculin influences essential processes in enteric nervous system development and Hirschsprung disease pathogenesis. *J Clin Invest* 2025;136:e198531. https://doi.org/10.1172/JCI198531
34. Daadi MM, Weiss S. Generation of tyrosine hydroxylase-producing neurons from precursors of the embryonic and adult forebrain. *J Neurosci* 1999;19:4484–97. https://doi.org/10.1523/JNEUROSCI.19-11-04484.1999
35. Kim SM, Yang JW, Park MJ *et al.* Regulation of human tyrosine hydroxylase gene by neuron-restrictive silencer factor. *Biochem Biophys Res Commun* 2006;346:426–35. https://doi.org/10.1016/j.bbrc.2006.05.142
36. Numakura Y, Uemura R, Tanaka M *et al.* PHF24 is expressed in the inhibitory interneurons in rats. *Exp Anim* 2021;70:137–43. https://doi.org/10.1538/expanim.20-0105
37. Bradley RM, Mardian EB, Bloemberg D *et al.* Mice deficient in *lysophosphatidic acid acyltransferase delta* (*Lpaat δ*)/*acylglycerophosphate acyltransferase 4* (*Agpat4*) have impaired learning and memory. *Mol Cell Biol* 2017;37:e00245–17. https://doi.org/10.1128/MCB.00245-17
38. Karagiota A, Chachami G, Paraskeva E. Lipid metabolism in cancer: the role of acylglycerolphosphate acyltransferases (AGPATs). *Cancers* 2022;14:228. https://doi.org/10.3390/cancers14010228
39. Deng F-Y, Lei S-F, Zhang Y *et al.* Peripheral blood monocyte-expressed ANXA2 gene is involved in pathogenesis of osteoporosis in humans. *Mol Cell Proteomics* 2011;10:M111.011700. https://doi.org/10.1074/mcp.M111.011700
40. Wang F, Jiang J, Xu T *et al.* ANXA2 knockdown inhibits oxidative stress and pyroptosis to alleviate epilepsy by suppressing the toll-like receptor (TLR) pathway. *Cell Signalling* 2025;136:112140. https://doi.org/10.1016/j.cellsig.2025.112140
41. Partevian SA, Slominsky PA, Shadrina MI *et al.* ANXA2 protein and its role in neurodegeneration processes. *Life* 2025;15:402. https://doi.org/10.3390/life15030402
42. Lin W, Shiomoto S, Yamada S *et al.* Dendritic spine formation and synapse maturation in transcription factor-induced human iPSC-derived neurons. *iScience* 2023;26:106285. https://doi.org/10.1016/j.isci.2023.106285
43. Karadurmus D, Rial D, De Backer J-F *et al.* GPRIN3 controls neuronal excitability, morphology, and striatal-dependent behaviors in the indirect pathway of the striatum. *J Neurosci* 2019;39:7513–28. https://doi.org/10.1523/JNEUROSCI.2454-18.2019
44. Zhang B, Li L, Tang X *et al.* Distribution patterns of subgroups of inhibitory neurons divided by Calbindin 1. *Mol Neurobiol* 2023;60:7285–96. https://doi.org/10.1007/s12035-023-03542-1
45. Kolberg L, Raudvere U, Kuzmin I *et al.* g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* 2023;51:W207–12. https://doi.org/10.1093/nar/gkad347
46. Bubis JA, Arrey TN, Damoc E *et al.* Challenging the Astral mass analyzer to quantify up to 5,300 proteins per single cell at unseen accuracy to uncover cellular heterogeneity. *Nat Methods* 2025;22:510–9. https://doi.org/10.1038/s41592-024-02559-1
47. Satija R, Farrell JA, Gennert D *et al.* Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33:495–502. https://doi.org/10.1038/nbt.3192
48. Reimand J, Kull M, Peterson H *et al.* g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 2007;35:W193–200. https://doi.org/10.1093/nar/gkm226
49. Angus-Hill ML, Schlichter A, Roberts D *et al.* A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. *Mol Cell* 2001;7:741–51. https://doi.org/10.1016/S1097-2765(01)00219-2
50. Kim AM, Bernhardt ML, Kong BY *et al.* Zinc sparks are triggered by fertilization and facilitate cell cycle resumption in mammalian eggs. *ACS Chem Biol* 2011;6:716–23. https://doi.org/10.1021/cb200084y
51. Lo MN, Damon LJ, Wei Tay J *et al.* Single cell analysis reveals multiple requirements for zinc in the mammalian cell cycle. *eLife* 2020;9:e51107. https://doi.org/10.7554/eLife.51107
52. Shen Z, Ma Y, Ji Z *et al.* Arachidonic acid induces macrophage cell cycle arrest through the JNK signaling pathway. *Lipids Health Dis* 2018;17:26. https://doi.org/10.1186/s12944-018-0673-0
53. Piras S, Furfaro AL, Domenicotti C *et al.* RAGE expression and ROS generation in neurons: differentiation versus damage. *Oxid Med Cell Long* 2016;2016:9348651. https://doi.org/10.1155/2016/9348651
54. Khammanivong A, Wang C, Sorenson BS *et al.* S100A8/A9 (Calprotectin) negatively regulates G2/M cell cycle progression and growth of squamous cell carcinoma. *PLoS One* 2013;8:e69395. https://doi.org/10.1371/journal.pone.0069395

55. Pichet Binette A, Gaiteri C, Wennström M *et al*. Proteomic changes in Alzheimer's disease associated with progressive Aβ plaque and tau tangle pathologies. *Nat Neurosci* 2024;27:1880–91. https://doi.org/10.1038/s41593-024-01737-w
56. Tirosh I, Izar B, Prakadan SM *et al*. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016;352:189–96. https://doi.org/10.1126/science.aad0501
57. Lloyd S. Least squares quantization in PCM. *IEEE Trans Inform Theory* 1982;28:129–37. https://doi.org/10.1109/TIT.1982.1056489
58. Hutton A, Meyer JG. Trajectory inference for single cell omics [Internet]. arXiv, http://arxiv.org/abs/2502.09354, 13 February 2025, preprint: not peer reviewed.

Received: May 8, 2025. **Revised:** March 14, 2026. **Accepted:** March 20, 2026