

Deciphering splicing heterogeneity at single-cell resolution by SCSES

Received: 4 November 2024

Accepted: 19 September 2025

Published online: 27 October 2025

Xiao Wen^{1,2,5}, Xuan Lv^{1,2,3,5}, Dan Guo^{1,2}, Nan Han^{1,2,3}, Lei Zhou^{1,2,3},
Peizhuo Wang^{1,2} & Zhaoqi Liu^{1,2}✉

Alternative splicing (AS) plays a critical role in generating cellular transcriptomic heterogeneity. While single-cell RNA sequencing (scRNA-seq) has become a standard approach for exploring this heterogeneity, it remains challenging to accurately characterize splicing changes at the single-cell level due to high dropout rates, inevitable noise, and limited coverage. To address this, we developed SCSES (Single-Cell Splicing ESTimation), a computational framework designed to enhance the AS profiles. SCSES infers and completes the missing splicing changes by sharing information across similar cells and events with data diffusion. Through systematic simulation studies, SCSES outperforms existing algorithms in recovering percent spliced-in (PSI) values and diversity across cell populations. When applied to various datasets, SCSES uncovers substantial splicing heterogeneity and cell subgroups with exclusive splicing patterns, which cannot be captured by conventional single-cell gene expression clustering. Together, our study provides SCSES as a valuable tool in deciphering splicing heterogeneity and is widely capable of handling different biological scenarios, species and sequencing platforms.

Single-cell RNA sequencing (scRNA-seq) techniques make it possible to decipher cell transcriptional heterogeneity on a large scale of cells¹. While many studies have successfully characterized the gene expression specificity in different single-cell clusters^{2–6}, the complexity of post-transcriptional regulation has often been overlooked. Alternative splicing (AS), an essential post-transcriptional process during pre-mRNA maturation, substantially contributes to tremendous transcriptional diversity^{7–9}. AS events can be classified into five classical types: exon-skipping events (SE), alternative 3' splicing site events (A3SS), alternative 5' splicing site events (A5SS), retention intron events (RI), and mutually exclusive exons events (MXE)¹⁰ (Supplementary Fig. 1). AS generates various isoforms from the same pre-mRNA by jointing exons in different combinations¹¹, thereby directly determining proteins constitutions^{12–15}. In this way, more than 95% of human multiexon genes undergo AS, yielding over 300,000 isoforms derived from ~24,000 protein-coding genes^{16,17}. Thus, compared to overall gene expression, AS changes are not only more numerous but

also carry richer information in terms of transcript structure and functional diversity, which expands the dimension of feature space to define cell types and infer cell state trajectories^{8,15,16}.

A number of computational tools, including MISO¹⁸, rMATS¹⁹, MAJIQ²⁰, and IRFinder²¹, are developed to identify and quantify AS events in bulk sequencing data. However, due to technical limitations of single-cell sequencing such as high dropout rate, high sequencing errors, and biased low coverage^{9,22}, these tools are poorly adaptive for scRNA-seq datasets^{23,24}, no matter whether using full-length protocols like smart-seq2²⁵ or the droplet-based protocols like inDrop-seq²⁶. Several methods have been specifically designed to address this challenge. For example, Expedition introduces a computational pipeline for quantifying exon-skipping events by leveraging well-aligned reads and a model to detect differential splicing events²⁷. Similarly, BRIE/BRIE2 infers differentially spliced exon-skipping events through Bayes regression models incorporating either event sequence or cell features^{23,28}. Psix can identify AS events that are highly correlated with

¹Department of Computation Biology, China National Center for Bioinformation, Beijing 100101, China. ²Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China. ³University of Chinese Academy of Sciences, Beijing 100049, China. ⁴School of Life Science and Technology, Xidian University, 710071 Xi'an, Shaanxi, China. ⁵These authors contributed equally: Xiao Wen, Xuan Lv. ✉ e-mail: liuqz@big.ac.cn

cell state transition by a probabilistic model²⁹. SCASL concentrates on the junctions with alternative 3'/5' splicing sites and infers the missing values in the AS probability matrix from other cells³⁰. Nonetheless, there are still apparent deficiencies in each of the above methods. Firstly, some methods, such as BRIE2, heavily rely on cell type identities, which are unavailable before analysis. Secondly, methods including Psix use global gene expression to measure cell similarities, which may not fully reflect the complexity of heterogeneous splicing patterns among cells. Thirdly, the limited junction read counts often result in unreliable estimation of percent splice-in (PSI) values leading to inaccurate quantification in methods such as Expedition and SCASL (Supplementary Fig. 2). Lastly, most algorithms are limited in scope, primarily focusing on SE and MXE events, while lacking the ability to detect and quantify A3SS, A5SS and RI events. A detailed comparison of these methods is listed in Supplementary Data 1.

Here, we present Single Cell Splicing Estimation algorithm (SCSES), a network diffusion-based imputation method designed to accurately recover splicing changes across main types of splicing events at the single-cell level. SCSES is inspired by the data diffusion technique, which is widely applied in scRNA-seq data analysis, such as MAGIC³¹, DTFLOW³² and PHATE³³. SCSES infers splicing changes by utilizing both cell and event similarities with different options of imputation strategies. Through extensive simulation studies and applications to various scRNA-seq datasets, we demonstrated the power of SCSES in faithfully recovering the splicing features in individual cells and cell groups. In summary, SCSES is a promising tool to explore and interpret single-cell data from a post-transcriptional perspective concerning the conventional single-cell gene expression analysis.

Results

SCSES characterizes splicing changes on single-cell resolution via data diffusion

SCSES is a computational framework designed to identify and quantify alternative splicing events using scRNA-seq data (Fig. 1). It takes scRNA-seq alignments as input and outputs refined PSI values for every detected splicing event in each cell. To define a global set of all splicing events, SCSES firstly merges all aligned reads from every single-cell into a pseudo bulk sequencing file without cell identities and identifies main types of splicing events by conventional algorithms (Fig. 1a). According to this splicing reference, SCSES then counts the raw inclusion/exclusion junction reads in each cell, constructing the raw junction read count matrix (Raw RC) and calculating the raw PSI matrix (Raw PSI) (Fig. 1a). Due to the high dropout rate and technical limitations, such matrices are very sparse with limited read counts, leading to inaccurate PSI estimation (Supplementary Fig. 3, 4). To overcome this sparsity, SCSES uses a diffusion operator that propagates information across similar cells/events over a lower dimensional manifold, restoring missing junctions in this process. The framework is built on the assumption that cells exhibiting similar activity of splicing machinery harbor akin splicing patterns²⁹ (Supplementary Fig. 5), and that events with similar regulatory features result in comparable splicing outcomes²³.

To build the underlying manifold, SCSES constructs cell and event similarity networks using K-nearest neighbor algorithm (KNN) (Fig. 1b). It uses gene expressions of RNA-binding proteins (RBPs), Raw RC, or Raw PSI matrices to measure pairwise cell similarities and builds the network with adaptively optimized *K* values for each cell (Methods). Event similarities are defined by the RBP regulatory correlations and an embedding representation by integrating event sequence similarities, which includes the length, motif, conservation and k-mer features from the event sequence (Supplementary Note 1). Once these networks are established, SCSES aggregates the splicing information across highly similar cells or events to impute splicing junctions or PSIs, thereby correcting for dropout and other sources of noise (Fig. 1). We evaluate the degree of difference between two successive imputed matrices and terminate the process once the changes stabilize (Fig. 1c,

Methods). Based on our practice, we recommend three different data imputation strategies, matched to four types of biological scenarios, which are defined by the abundance of alternative junctions in both the target cell and its neighboring cells (Fig. 1c, Supplementary Fig. 6, Supplementary Note 7). Here, for a given AS event, cells with dropout in alternative splicing junctions are defined as WD (with dropout), while others as ND (non dropout). For ND cases, which have reads supporting alternative junctions, we suggest imputing the PSI value directly using a cell similarity network (Supplementary Fig. 7, 8). WD cases are further divided into BD (biological dropout) or TD (technical dropout) based on their read count patterns. In BD cases, both the target cell and its neighbors show abundant reads for one junction type (either inclusion or exclusion) but complete absence of reads for the other type, indicating that only a single isoform is expressed. TD cases, however, are characterized by low read depth in both harboring gene and splicing event in the target cell, likely due to technical limitations rather than biological truth. We further divided TD based on the information available from neighboring cells: TD+Info represents cases where neighboring cells have abundant junction reads that can guide imputation, while TD-Info represents cases where such information is lacking in the local neighborhood. For BD, TD+Info and TD-Info, we recommend data imputation on the raw junction matrix with cell similarity network. And specifically, TD-Info requires an additional round of data diffusion using the event similarity network on the PSI matrix derived from previous round of imputation (Methods, Supplementary Fig. 7–8). To determine the appropriate scenario for each cell-event pair belongs to, SCSES pre-trains a cascade decision model to predict the scenario probabilities (Methods). Lastly, SCSES estimates the final PSI value via a weighted linear combination of predictions from all different strategies, with weights determined by their corresponding probabilities (Methods, Fig. 1c).

SCSES recapitulates splicing characteristics in individual cells

To assess the accuracy of SCSES, we utilized bulk RNA-seq data from four cell lines (HCT116, HCC1954, HepG2, and HL-60) obtained from Cancer Cell Line Encyclopedia (CCLE)³⁴. From these bulk datasets, we selected a set of high-confidence splicing events with PSI values served as the biological truth. We then used real scRNA-seq data for the same cell lines^{35,36} to test if SCSES could accurately recapitulate the splicing landscape at the single-cell level (Supplementary Fig. 9a, Supplementary Fig. 10). We compared SCSES against five existing algorithms for PSI estimation in scRNA-seq data: BRIE1, BRIE2 (aggregated model), Expedition, Psix and SCASL, five algorithms designed for inferring PSI values in scRNA-seq data, as well as rMATS, a classical algorithm designed for bulk RNA-seq data. Due to the substantial variability in splicing events identified by different algorithms, we performed pairwise comparisons between SCSES and each of the other tools, focusing only on the overlapping splicing events to ensure a fair evaluation (Supplementary Data 2). Firstly, we evaluated the accuracy of PSI recovered by different methods. We calculated the Spearman correlation coefficients (SCC) between the PSI values estimated by the compared algorithms and the benchmark PSI values (derived from bulk RNA-seq of the matched cell type) across all events within each cell, as well as the root mean squared error (RMSE) between estimated and the benchmark PSI values across all cells for each splicing event. Overall, in real datasets, SCSES consistently outperformed the competing algorithms by achieving higher cell-wise PSI correlations and lower event-wise PSI estimation errors (Fig. 2a, b, Supplementary Figs. 11–13). In all cases, SCSES outperformed BRIE1, Expedition and rMATS, exhibiting a significant increase in median SCC values across cells, ranging from 0.1–0.6, and a reduction in median RMSE by > 19.4% across events (Fig. 2a, b, Supplementary Figs. 11–13). In most cases, SCSES achieved improved or comparable performance relative to algorithms relying on inference from similar cells, such as BRIE2, Psix and SCASL (Supplementary Figs. 11–13). Furthermore, we synthesized artificial scRNA-seq data to simulate a

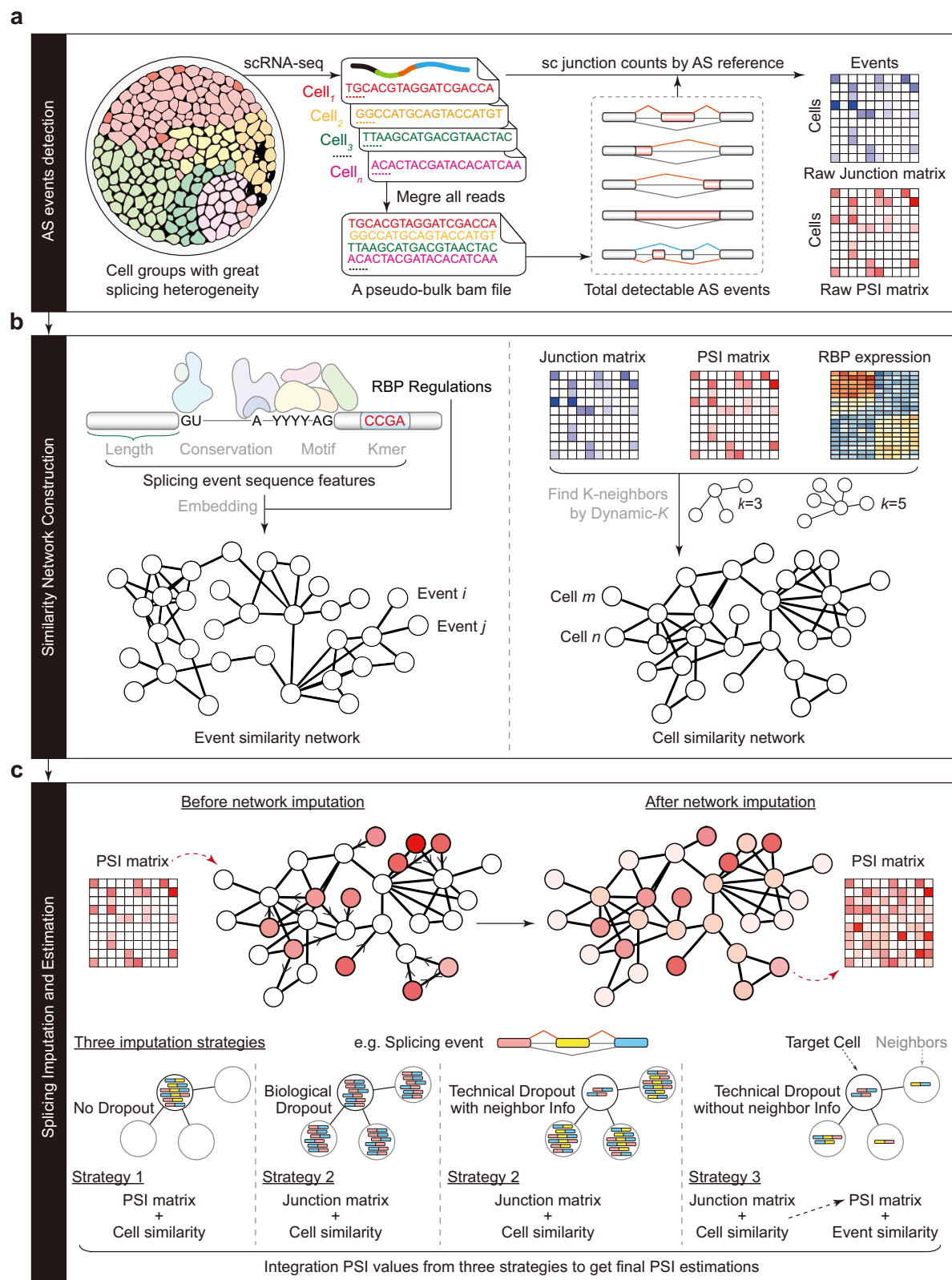
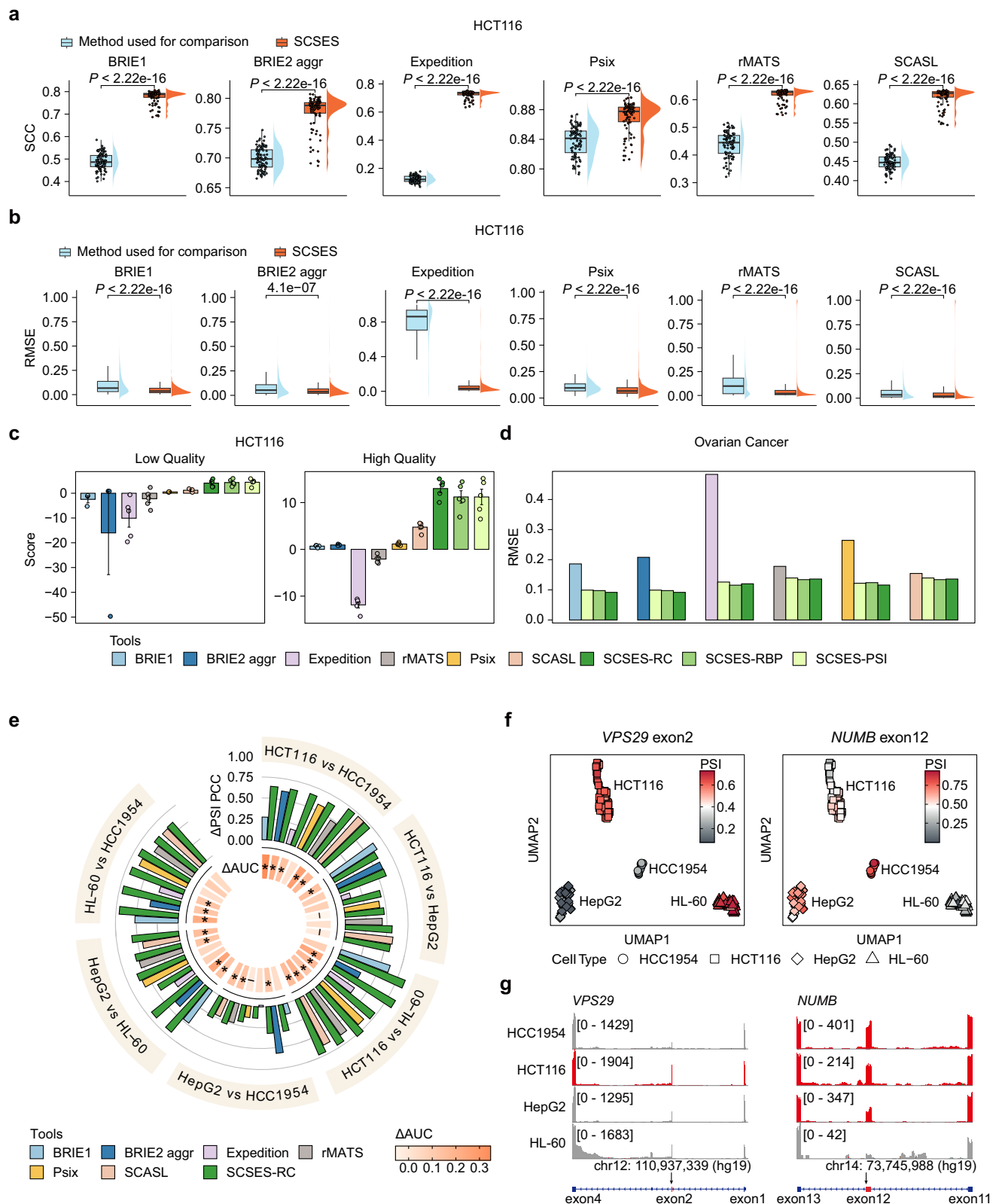


Fig. 1 | Overview of the SCSES framework. **a** SCSES creates a pseudo-bulk bam file by merging all single-cell reads. Conventional AS detection algorithms are used to detect all kinds of AS events. Raw junction read counts can then be retrieved based on this splicing reference. Raw PSI values can be calculated by raw junction read counts. **b** Cell splicing similarities can be built by three types of matrices. AS event similarities are measured by combining the RBP-event regulation relationship and event sequence features. KNN algorithm is used to create cell/event similarity

networks. Dynamic-K strategy is introduced to select optimal K for each cell. Random walk with restart (RWR) algorithm is used to capture the global topological similarities in both networks. **c** Event-cell pairs are divided into four types based on the dropout of target cell and neighbor cells. SCSES introduces three imputation strategies for different event-cell groups. Final imputed PSI values are the linear combination of different imputation values with predicted group probabilities as coefficients.



balanced population of the four cell types with varying sequencing qualities via Spanki simulator³⁷ (Supplementary Fig. 9b, Supplementary Note 2). By testing on these simulated datasets, SCSES obtained the best performance, effectively balancing the accuracy of AS identification and quantification across all cell lines, regardless of the dataset quality (Fig. 2c, Supplementary Fig. 14a–c). To assess the performance of SCSES in real tissues, we applied SCSES and other algorithms to single-cell

datasets from an ovarian cancer sample³⁸ and the human hippocampus³⁹, in which each cell was sequenced using paired short-read and long-read technologies (Supplementary Data 3–4, Methods, Supplementary Note 2). The PSI values of high-quality event-cell pairs derived from long-read data were used as benchmarks to calculate the RMSE of PSI estimates from each method (Supplementary Fig. 9d). Considering the substantial differences in detected events by different

Fig. 2 | SCSES recovers the splicing levels in individual cells. a, b Raincloud plots showing the performance comparison between SCSES-RC and other algorithms in real scRNA-seq data of HCT116. **a** The SCC value refers to the correlation between the estimated PSI values and benchmarks of all events in each cell. $N_{\text{HCT116}} = 91$. **b** The RMSE is calculated between the estimated PSI and benchmarks of an event among all cells. The events detected by both SCSES and the compared algorithm are considered for each comparison. The *P*-values are calculated by the Wilcoxon test (two-sided test) without any adjustments. $N_{\text{BRIE1}} = 556$, $N_{\text{BRIE2}} = 556$, $N_{\text{Expedition}} = 1458$, $N_{\text{Psix}} = 495$, $N_{\text{IMATS}} = 10,313$, $N_{\text{SCASL}} = 10,313$. The boxes indicate median (center), Q25, and Q75 (bounds of box), the smallest value within 1.5 times interquartile range below Q25 and the largest value within 1.5 times interquartile range above Q75 (whiskers). **c** Bar plots showing the accuracy scores of different algorithms in HCT116 synthetic dataset. The accuracy score is defined as the product of correlation between inference and benchmarks and AS events recall rate. Error bars represent the standard error of the

mean from five independent replicates. **d** Bar plots comparing the performance of different methods on Ovarian Cancer dataset. The RMSE is calculated between the estimated PSI values by different methods and benchmark values averaged on all cell-event pairs. **e** Comparisons of the detected DSEs between SCSES-RC with other algorithms in the real datasets. Bar plot shows the SCC of ΔPSI in DSEs from the benchmark in each comparison group. Colors represent different algorithms. The inner circle represents the difference of AUC for DSEs identification between SCSES and the compared algorithm. On the circle, *: $\text{AUC}_{\text{SCSES}} - \text{AUC}_{\text{ref}} > 0.1$, -: $\text{AUC}_{\text{SCSES}} - \text{AUC}_{\text{ref}} < 0$. **f** Scatter plot showing the inclusion level of *VPS29* exon 2 (left panel) and *NUMB* exon 12 (right panel) between 4 cell lines ($N = 235$). **g** Read coverage showing the inclusion of *VPS29* exon 2 (left panel) and *NUMB* exon 12 (right panel) on the bulk RNA-seq of four cell lines. Alternative exons are highlighted in red in the genome annotation track. Source data of panels is provided as a Source Data file.

methods (Supplementary Fig. 14e–h), we used only the common high-quality event–cell pairs between SCSES and each compared method for the evaluation. SCSES consistently achieved the lowest RMSE across all comparisons. Notably, RMSE was reduced by over 13% in the ovarian cancer dataset and by ~16% in the human hippocampus dataset by SCSES (Fig. 2d, Supplementary Fig. 14d). In general, the algorithms based on inference from similar cells (SCASL, Psix, BRIE2) perform better than others. However, these algorithms require external information to estimate PSI. For instance, BRIE2 incorporates cell identity (e.g., cell type or developmental stage) as a prior feature for Bayesian regression. In contrast, Psix and SCASL require users to predefine the number of nearest neighboring cells, which constrains the learning of splicing information to fixed-size local cell populations. These results suggest that SCSES can recapitulate the splicing features with higher accuracy automatically.

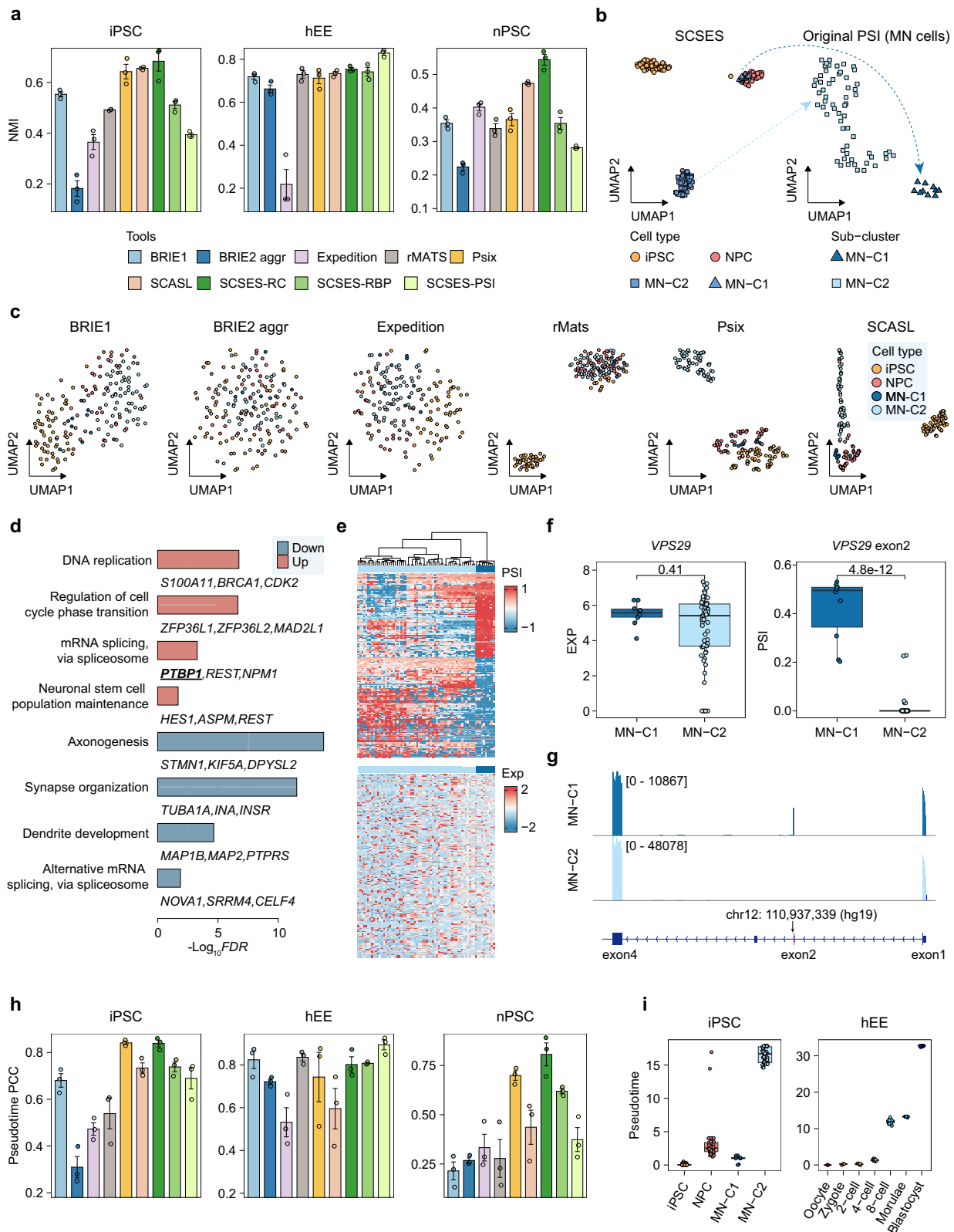
Subsequently, we assessed the capability of detecting differentially spliced events (DSEs) between these four cell types using the receiver operating characteristic curve (Methods, Supplementary Data 5). In general, SCSES demonstrated superior accuracy compared to other algorithms by promoting the area under curve (AUC) > 0.1 in most cases (Fig. 2e). Moreover, the splicing changes detected by SCSES aligned more closely with real cell bulk sequencing results than those reported by other methods (Fig. 2e, Supplementary Fig. 15). These outcomes suggest that SCSES can provide more reliable PSI values for DSE identification. Notably, SCSES could capture a broader range of splicing events and event types than algorithms designed for scRNA-seq data (Supplementary Fig. 16a). For example, SCSES successfully identified the inclusion of exon 2 in *VPS29* transcript in HCT116 cells, which is consistent with observations from bulk RNA-seq data of HCT116 and recent studies⁴⁰ (Fig. 2f, g). However, this splicing variant was missed by all other algorithms for scRNA-seq data. Additionally, SCSES was the only method to uncover that *NUMB* exon 12 was more frequently included in three solid tumor cell lines (HCC1954, HCT116 and HepG2) than the blood tumor cells (HL-60), in agreement with recent studies^{41–43} (Fig. 2f, g). Moreover, most methods are limited to detect cassette exons, thus cannot infer the splicing levels of other splicing types, including A3SS, A5SS, RI, and MXE. In contrast, SCSES accurately captured these additional event types (Supplementary Fig. 16b–d). Taken together, our evaluation demonstrated that SCSES can provide more reliable and comprehensive splicing changes of all types on single-cell level than other methods.

SCSES reproduces the splicing status across cell populations

To evaluate the biological implications of SCSES, we assessed the capability of imputed splicing profiles to identify cell types and infer cell pseudo-time trajectories. For this analysis, we collected three public datasets with high sequencing quality, including induced human naïve pluripotent stem cells (nPSC)⁴⁴, human early embryos (hEE)⁴⁵, and induced human pluripotent stem cells (iPSC)²⁷ (Methods). We generated lower-quality datasets by down-sampling reads from

original BAM files and tested if SCSES could recover the splicing changes from the original scRNA-seq data of relatively high sequencing quality (Supplementary Fig. 9d, Methods). The specific *K* value for each cell was shown in Supplementary Fig. 17.

To evaluate the accuracy of cell clustering based on splicing levels estimated by different algorithms, we calculated the normalized mutual information (NMI) between the *K*-means clustering results derived from the estimated PSI profiles and the cell type annotations provided in the original publications of the test datasets. Overall, SCSES showed the best performance, with an average NMI score improvement of >10% on both the nPSC and hEE datasets (Fig. 3a). For the iPSC dataset, the performance of SCSES was slightly higher than Psix and SCASL, and substantially higher than the remaining methods. Interestingly, low dimensional reduction analysis showed that SCSES divided the motor neurons (MNs) into two more subclusters (MN-C1, MN-C2), a distinction that was also captured by the Psix and SCASL but was not as clearly delineated by other methods or by gene expression values (Fig. 3b, c, Supplementary Fig. 18a, b). These two splicing subclusters could also be detected by using the original splicing profile from MN cells (Fig. 3b right panel). Moreover, a set of genes associated with mRNA splicing and neuron development were differentially expressed between MN-C1 and MN-C2 cells (Fig. 3d). Among them, *PTBP1* emerged as the most significantly upregulated splicing factor in MN-C1. This gene has been extensively characterized for its role in regulating neuronal development in neural stem cells and progenitors^{46,47} (Supplementary Fig. 18d). In addition, the pseudotime analysis by Monocle3⁴⁸ and CytoTrace⁴⁹ demonstrated that MN-C1 exhibited stronger stemness and earlier pseudotime compared to MN-C2 cells (Supplementary Fig. 18c, h right panel). Moreover, multiple genes exhibited changes in splicing patterns ($\Delta\text{PSI} > 0.1$ and $FDR < 0.05$, Wilcoxon test) while no changes in expression levels ($\log_2 FC < 0.5$ and $P > 0.05$, Wilcoxon test) (Fig. 3e). For example, expression levels of *VPS29* did not show significant differences between MN-C1 and MN-C2 cells ($P = 0.41$, $\log_2 FC = -0.11$, Fig. 3f left panel), while MN-C1 cells expressed more isoform including 2nd exon (Fig. 3f right panel, Fig. 3g). *VPS29* has been demonstrated to have important functions in synaptic vesicle recycling and synaptic transmission⁵⁰. Collectively, these observations indicate the presence of splicing heterogeneity within MNs, and we suspect that MN-C1 cells are likely in the early stages of differentiation from iPSCs to MNs. Furthermore, we examined the splicing regulatory relationships between RBPs and splicing events. Previous publications have linked *PTBP1* expression to the mutual exclusivity of exon 9 or exon 10 in *PKM*⁵¹, and exon 4 exclusion of *SRSF3*⁵². The imputed PSI values estimated by SCSES for these events were highly correlated with *PTBP1* expression (Supplementary Fig. 18e–g). Additionally, we assessed whether the predicted PSI values could support infer pseudotime trajectories. The pseudotime inferred from original gene expression using monocle3⁴⁸ served as benchmarks (Supplementary Fig. 18h). In terms of individual cell, the pseudotime predicted by SCSES-imputed PSI presented a higher correlation with benchmarks across all datasets (Fig. 3h). From the



perspective of cell groups, the trajectories were highly consistent with the cell differentiation stages (Fig. 3i, Supplementary Fig. 18i).

SCSES reveals a cell group at primary diagnosis of multiple myeloma that is associated with potential drug-resistance

Growing evidence indicates that AS contributes to tumor relapse and therapeutic failure in multiple tumors⁵³. Here, we investigated the

effect of AS on drug-resistance in multiple myeloma (MM). 127 MM cells by Smart-seq2 sequencing were collected from the same patient before and post treatment with thalidomide and bortezomib (BTZ)⁵⁴. 62/127 cells were collected from bone marrow at diagnosis, and others were collected after metastasizing to ascites dissemination at disease relapse⁵⁴. The conventional gene expression profiles detected two clusters (EC1 and EC2), which matched the cell sampling from initial or

Fig. 3 | SCSES reproduces the splicing status across cell populations. **a** Bar plot of NMI across three test datasets. For each dataset, down-sampling is replicated three times. **b** left panel: UMAP plots of 174 cells from iPSC test dataset by PSI values estimated by SCSES-RC. The colors indicate the cell labels from the previous publication. right panel: UMAP plots show the projection and clusters of MNs using the original PSI profile (non-down sampling data). The shape of points represents subgroups of MN cells. **c** UMAP plots of iPSC test dataset by PSI values estimated by other algorithms. **d** Biological function enrichment analysis of the DEGs in MN-C1 compared to MN-C2. **e** top panel: Heatmap showing PSI profiles of DSEs between MN-C1 and MN-C2 groups, where the corresponding target genes do not exhibit differential expression. bottom panel: Heatmap displaying the expression profiles of the target genes corresponding to the splicing events shown in the top panel. **f** Box plot showing the gene expression (left panel) and PSI (right panel) of *VPS29* between MN-C1 and MN-C2. $N_{MN-C1} = 10$, $N_{MN-C2} = 60$. The *P*-values are calculated by

the Wilcoxon test (two-sided test) without any adjustments. **g** Read coverage showing the inclusion of *VPS29* exon 2 on the pseudobulk of MN-C1 and MN-C2. Alternative exons are highlighted in red. **h** Bar plot of PCC between pseudotime inferred by PSI values from test datasets and gene expression benchmark, with calculations on each dataset replicated three times. The color legend is the same with (a). **i** Box plot showing the consistency of pseudotime inferred by PSI values and real differentiation stages in iPSC (left panel) and hEE (right panel) dataset. $N_{iPSC} = 63$, $N_{NPC} = 41$, $N_{MN-C1} = 10$, $N_{MN-C2} = 60$. $N_{Oocyte} = 3$, $N_{Zygote} = 3$, $N_{2-cell} = 6$, $N_{4-cell} = 12$, $N_{8-cell} = 20$, $N_{Morulae} = 16$, $N_{Blastocyst} = 30$. Error bars in (a) and (h) represent the standard error of the mean from three independent replicates. The boxes in (f) and (i) indicate median (center), Q25, and Q75 (bounds of box), the smallest value within 1.5 times interquartile range below Q25 and the largest value within 1.5 times interquartile range above Q75 (whiskers). Source data of panels is provided as a Source Data file.

recurrent tumor (Supplementary Fig. 19a). In contrast, 12,468 AS events imputed by SCSES detected four clusters, which subdivided each gene expression cluster into two subgroups (EC1 into SC1 and SC2, EC2 into SC3 and SC4), respectively (Fig. 4a, b, Supplementary Fig. 19b). This grouping could not be detected by other splicing inference algorithms (Supplementary Fig. 19c). To investigate the biological relevance of these splicing-defined clusters, we predicted the cell evolution trajectory by pseudotime analysis with monocle³⁴⁸ and CytoTrace⁴⁹, and both methods reported a highly consistent order from SC1 to SC4 (Fig. 4c, Supplementary Fig. 19d). In addition, we observed that the most variable splicing events by SCSES exhibited a monotonically increasing or decreasing pattern from SC1 to SC4 (Fig. 4d). The previous publication has reported a linear tumor evolution pattern in these cells, with increased copy number variation upon tumor relapse⁵⁴. Similarly, we found increasing copy number changes from SC1 to SC4 (Fig. 4e). Together, these results indicate a continuous cell evolution path by splicing variations. Next, we wished to validate this cell evolution path from the perspective of RNA velocity, which models the transcriptional dynamics using spliced and unspliced reads. We hypothesize that, with the refined splicing information by SCSES, the junction reads supporting alternative intron retentions (IR) can be utilized to better model the RNA dynamics and increase the prediction accuracy. Therefore, we inferred the RNA velocity by scVelo⁵⁵ with the imputed IR junctions, and obtained an evolution direction from SC1 to SC4, which was consistent with the above results (Fig. 4f). However, the original scVelo prediction had a total reverse direction from SC4 to SC1, and scVelo using raw IR junctions without imputation also made a disordered prediction (Supplementary Fig. 19e).

Next, we performed marker gene analysis in SC1-SC4 and found that SC1 highly expressed *CD79A*, which is reported to be associated with favorable overall survival⁵⁶ (Supplementary Fig. 19f), while SC2-SC4 showed increased expression of genes associated with poor prognosis, including *SRRM2* in SC2⁵⁷, *ADAM10* in SC3⁵⁸, *AURKA* and *CHEK1* in SC4^{59,60} (Supplementary Fig. 19f). Interestingly, SC2-SC4 increased expression of genes associated with BTZ-resistance, including *NFE2L3* and *NOTCH2* in SC2^{61–63}, *FOXO3* and *MRPL20* in SC3⁶⁴, *CDC25B* in SC4⁶⁵ (Supplementary Fig. 19f). These results indicate that SC2 cells may already develop BTZ-resistance potential at diagnosis. Thus, we further focused on the transcriptomic changes between SC1 and SC2, and we found that exclusively expressed genes in tumor relapse already demonstrated the same regulation changes in SC2 cells than SC1 (Fig. 4g). Specifically, genes upregulated in SC2 cells were related to cell mitosis, heat shock response and L-glutamine (Gln) process (Supplementary Fig. 19g). Previous studies reported that BTZ resistance can be induced by high levels of heat shock proteins in MM cells⁶⁶. In addition, upregulated Gln metabolism regulators promote proteasome inhibitors resistance in plasma cell myeloma⁶⁷. We also found dysregulation of genes associated with these functions along the cell evolution path from SC1 to SC4 (Supplementary Fig. 19h).

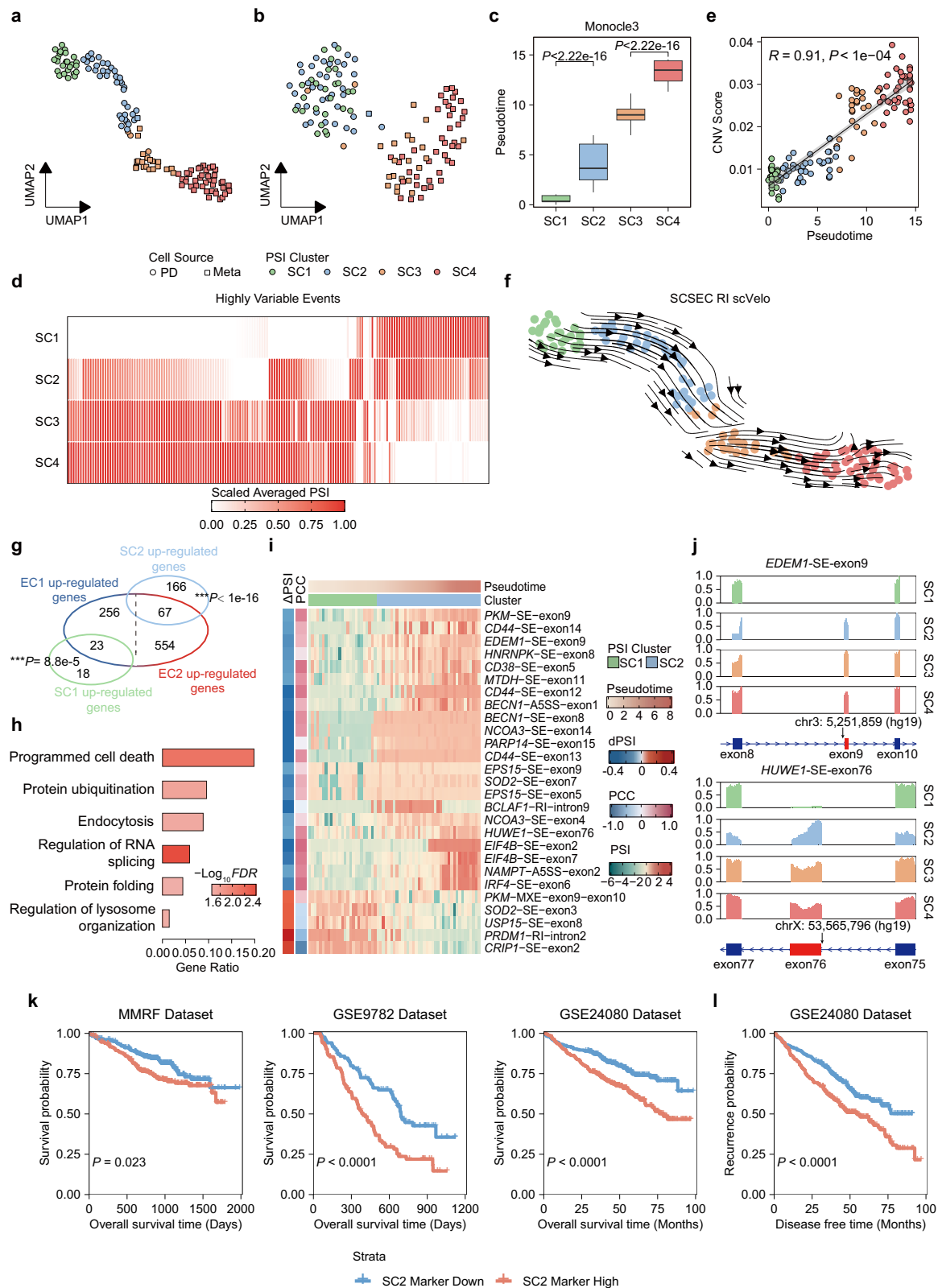
Finally, we analyzed the AS patterns between SC1 and SC2. In total, we detected 248 DSEs between SC1 and SC2, while the overall expression levels of the genes harboring these DSEs were not differentially changed (Supplementary Fig. 20a, Supplementary Data 6). The intensities of splicing changes were highly correlated with the pseudotime prediction from SC1 to SC4 (Supplementary Fig. 20a, b). Genes harboring splicing abnormalities were functionally enriched with protein metabolism and cell death pathways (Fig. 4h). Particularly, 13 genes associated with protein ubiquitination were alternatively spliced (Supplementary Data 6). The mechanism of BTZ treatment involves increasing the endoplasmic reticulum stress and cell apoptosis by inhibiting the activity of proteasome, which leads to the accumulation of ubiquitinated proteins⁶⁸. Here, the aberrant splicing changes in ubiquitination pathway genes may indirectly dysregulate the ubiquitin–proteasome system, which potentially decrease the sensitivity to BTZ. This observation requires experimental supports in further studies. We also found the exclusion of *CADMI* exon 10 and the inclusion of exon 8 and exon9 in SC2 (Supplementary Fig. 20b). *CADMI* isoforms can regulate cell survival and homotypic adhesion in human mast cells⁶⁹, but their functions in MM cells have not been investigated. In addition, 19 differentially spliced genes have been reported to be involved in BTZ resistance (Fig. 4i, Supplementary Data 7). For instance, expressions of *EDEM1* and *EPS15* are associated with varying outcomes of BTZ treatment^{70,71}. Phosphorylation of *EIF4B* partially leads to the acquisition of BTZ resistance⁷². And a treatment combination of *HUWE1* inhibitors with BTZ can increase the effect of BTZ⁷³. These DSEs from BTZ-associated genes persistently presented in SC2 to SC4 cells (Fig. 4i, j, Supplementary Fig. 21a, b), indicating that besides the altered gene expression, AS may also contribute to the BTZ resistance.

To validate the potential BTZ-resistance of SC2 cells, we also analyzed independent scRNA-seq data from a primary MM patient and confirmed that existence of SC2 cells (Supplementary Fig. 22a–d, Supplementary Note 4). Interestingly, by analyzing three additional cohorts of BTZ-treated MM patients, we found that increased activity of SC2 marker genes was associated with worse patient overall survival (Fig. 4k) and disease-free survival (Fig. 4l), and relapsed patients exhibited higher SC2 marker genes activity than non-relapsed group (Supplementary Fig. 22e).

Taken together, with an improved splicing profile by SCSES, we identified a cell subgroup with resistance potential to BTZ treatment at MM diagnosis, which was undetectable by conventional gene expression analysis. Splicing alterations could correctly track the cell evolution trajectory, and reveal the potential mechanism for BTZ resistance.

SCSES deciphers the splicing dynamics from mesendoderm to definitive endoderm during human embryo development

Previous studies have highlighted the profound impacts of AS during embryonic stem cell (ESC) differentiation with bulk data^{74,75}. Here, we used SCSES to evaluate the role of AS during transition from



pluripotency maintenance to lineage-specific differentiation on single-cell resolution. To this end, we collected public Smart-seq2 sequenced data on 758 cells from three major differentiation states of H9 hESC (pluripotent state: 0 h, mesendoderm/ME: 12 h and 24 h, definitive endoderm/DE: 36 h, 72 h, 96 h)⁷⁶. Most cells at different time points were clustered into independent groups by SCSES splicing profile, indicating clear splicing dynamics throughout the differentiation

course (Fig. 5a), while a small set of cells from 96 h overlapped with cells from 72 h (Fig. 5a). Further analysis demonstrated that the cells appeared to reach a relatively stable state by 72 h, consistent with previous studies⁷⁶, and the non-overlapped 96h-cells were in a more mature stage compared to overlapped 96h-cells (Supplementary Fig. 23, Supplementary Note 5). Enrichment analysis on differentially spliced genes (Supplementary Data 8) showed that splicing alterations

Fig. 4 | SCSES reveals drug resistance profiles in the treatment-refractory multiple myeloma patient. a, b UMAP plots showing clustering of 127 cells in MM34 patient estimated by splicing profile (a) and gene expression profile (b). Cell colors represent the clusters, and shapes represent the source of samplings. PD: primary diagnosis, Meta: metastasis. **c** Box plot showing the consistency between PSI clusters and pseudotime estimated by Monocle3. $N_{SC1} = 25$, $N_{SC2} = 38$, $N_{SC3} = 23$, $N_{SC4} = 41$. The P -values are calculated by the Wilcoxon test (two-sided test) without any adjustments. The boxes indicate median (center), Q25, and Q75 (bounds of box), the smallest value within 1.5 times interquartile range below Q25 and the largest value within 1.5 times interquartile range above Q75 (whiskers). **d** Heatmap showing the averaged splicing levels of highly variable events in SC1 to SC4. **e** Scatter plot showing the correlation between pseudotime and CNV accumulation. The correlation coefficient (R) and associated P -value are calculated by Pearson correlation test ($N = 127$). **f** RNA velocity estimated by imputed raw junction counts of RI events with SCSES. **g** Venn plot showing the overlap of SC1/SC2 DEGs and EC1/

EC2 DEGs. The P -values are calculated by the two-sided Fisher's exact test without any adjustments. **h** The gene function enrichment analysis of genes associated with 248 DSEs between SC1 and SC2. **i** Heatmap showing the splicing profiles of DSEs between SC1 and SC2. The associated genes harboring these DSEs were reported to be associated BTZ resistance. ΔPSI represents the difference of averaged splicing levels in SC1 and SC2. PCC represents the correlation between PSI and pseudotime in all cells. **j** Exon read coverage in two representative DSEs associated with BTZ-resistance genes by merging the reads from the same cell group. Alternative exons are highlighted in red in the genome annotation track. **k** Kaplan-Meier survival curves showing overall survival differences based on SC2 marker gene activities in MMRF, GSE9782, and GSE24080 datasets. Survival differences were evaluated using the two-sided log-rank test. **l** Kaplan-Meier survival curves showing disease-free survival differences based on SC2 marker gene activities in GSE24080 dataset. Survival differences were evaluated using the two-sided log-rank test. Source data of panels is provided as a Source Data file.

were involved in the differentiation progress, such as endodermal cell fate commitment, WNT signaling pathway, and epithelial to mesenchymal transition (EMT) processes (Fig. 5b). By further checking DSEs determining this clustering, we found events that exclusively appeared at unique time points, as well as changes that persistently presented in successive stages (Fig. 5c, d, Supplementary Fig. 24a). For example, *JMJD1C* intron 6 retention specifically appeared at 12 h, while the inclusion level of *DNMT3B* exon 21 underwent a progressive decrease from 0 h to 96 h (Fig. 5d). These results suggest different isoforms are generated at distinct stage of differentiation and their dynamic changes are informative to infer cell state transition. mRNA splicing is orchestrated by numerous RBPs with extensive and dynamic interactions⁷⁷. Next, we constructed a splicing regulatory network between uniquely expressed RBPs at specific time points and regulated DSEs (Fig. 5e, Supplementary Fig. 24b, Supplementary Data 9, Supplementary Note 5). In this network, several RBPs involved in embryonic development presented as hub nodes, which were associated with the large number of DSEs^{78–84}. For example, The RBP with the largest number of regulated network targets was *ESRPI*, a critical regulator in epithelial-mesenchymal transition (EMT)⁸⁵. EMT takes place during the ingression of pluripotent epiblast cells through the primitive streak, initiating their differentiation into the mesoderm and DE germ layers⁸⁶. Moreover, we found that *RBM24*, reported to regulate mesenchymal-like splicing patterns⁸⁷, was up-regulated in DE stage cells (Supplementary Fig. 24c).

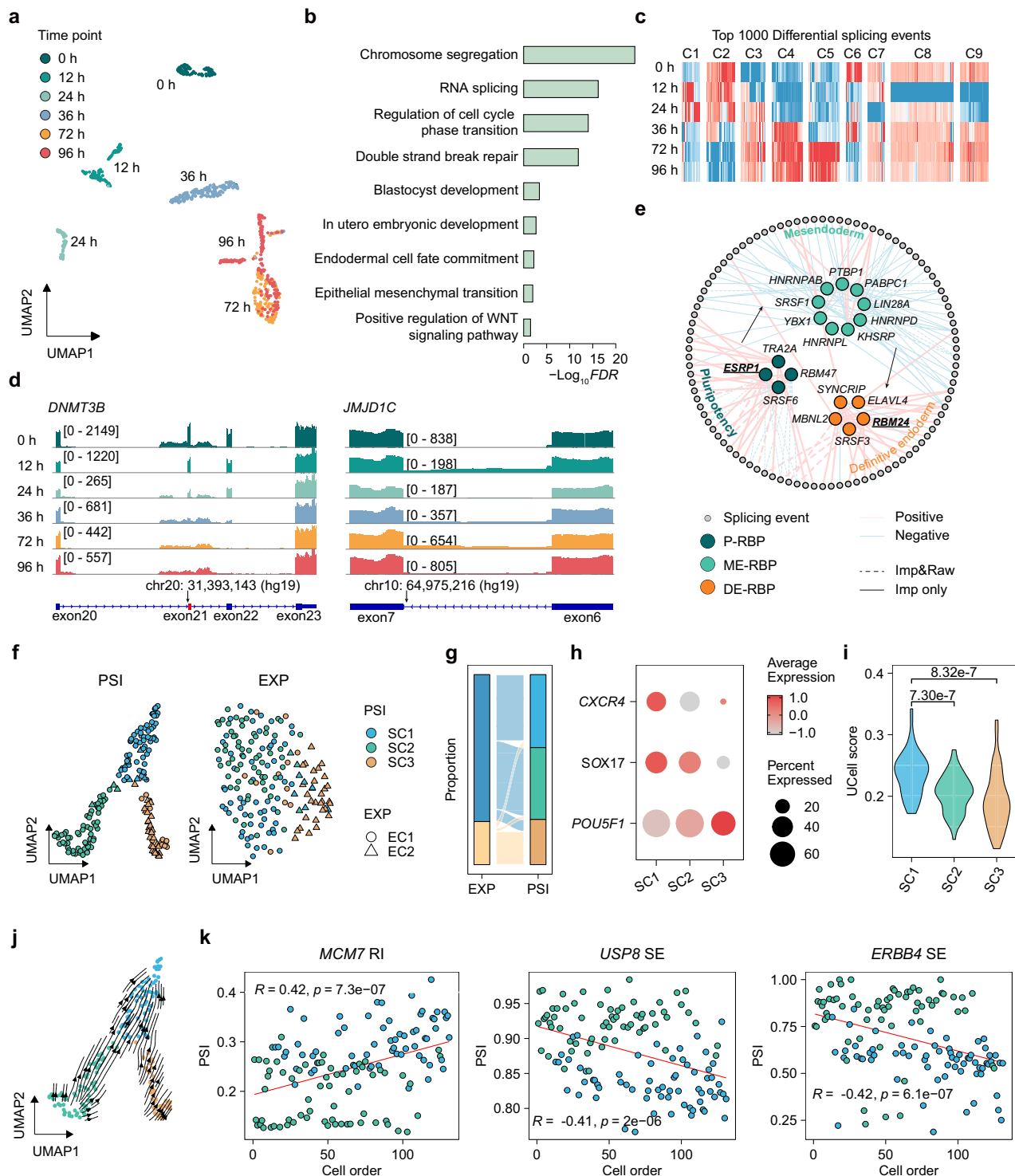
Finally, we focused on the critical transition from ME to DE at 36 h and identified two clusters (EC1, EC2) using overall gene expression without clear demarcation between them. While under the same clustering resolution, the SCSES splicing features yielded three clusters (SC1-SC3) by splitting one gene expression cluster into two subgroups (Fig. 5f, g), which could not be detected by other splicing inference algorithms (Supplementary Fig. 24d). SC1 cells highly expressed DE markers *CXCR4* and *SOX17*, while cells in SC2 and SC3 highly expressed the pluripotency marker *POU5F1*⁸⁶ (Fig. 5h), indicating SC1-SC3 corresponding to different differentiation phases. We next inferred the differentiation order of cells by measuring the activity score of differentiation signatures, whose expression monotonically increased from 0 h to 96 h⁸⁸ (Supplementary Note 5). In so doing, we found that cells in SC1 had higher differentiation scores than SC2 and SC3 cells (Fig. 5i), which was also confirmed by RNA velocity analysis (Fig. 5j). These results suggest the SC1 cells as a unique cell cluster presenting a late stage of ME to DE transition. The separation of SC1 from SC2 could only be detected by AS changes, rather than overall gene expression. Notably, the PSI changes of *MCM7* 4th intron inclusion, which introduced PTCs in *MCM7* transcripts, was positively correlated with the inferred differentiation order from SC2 to SC1 (Fig. 5k). *MCM7* is reported to be involved in EMT process⁸⁹ and exhibited lower expression in SC1 as well as monotonically decreased expression along the differentiation times (Supplementary Fig. 24e, f), suggesting that AS is likely to contribute to

EMT by affecting *MCM7* expression. Moreover, the inclusion levels of *USP8* exon 3 and *ERBB4* exon 26 exhibited a significant negative correlation with the differentiation order without altering gene overall expressions (Supplementary Fig. 24e, f). Based on previous studies, *USP8* preserves stemness of ESCs⁹⁰, and the *ERBB4* isoforms lacking exon 26 prevent activation of the PI3K signaling pathway⁹¹, facilitating ME to DE differentiation⁹². These results imply that AS may drive ME to DE transition by modulating transcript compositions, and SCSES helps to reveal critical splicing alterations during this transition.

SCSES identifies an activated monocyte subtype in induced HSC differentiation

Droplet-based protocols, such as inDrop²⁶ and 10x Chromium⁹³, are still the most commonly used technique for single-cell sequencing, with lower cost and higher throughput compared to full-length protocols. However, they introduce more challenges for detecting AS events due to the biased read coverage and low read depth, highlighting the need for SCSES for AS correction. To test SCSES on droplet-based data, we explored the splicing heterogeneity during hematopoietic stem cell (HSC) differentiation with lineage-tracing cells prepared by inDrop-v3⁹⁴. 31,542 mouse cells were collected for 2-, 4-, and 6 day culture in vitro, and 11,192 of them had definite clonal fates. To expand the splicing events pool in the inDrop-seq dataset, we added the alternative last exon events detected by MAJIQ into our analysis (Supplementary Note 6). Totally, 2978 valid AS events were detected and quantified. Compared to the SMART-seq dataset⁹⁵, the read coverage in inDrop data was enriched near the 3' end of genes (Supplementary Fig. 25a). This 3' biased read coverage influenced the detection of splicing events towards the 3' ends of genes (Supplementary Fig. 25b). Consequently, AS events in the 3' untranslated regions (3' UTRs) were more frequently detected in inDrop sequencing data compared to SMART-seq data (Supplementary Fig. 25c). 13 cell clusters were detected by combining gene expression profile and AS profile of 2,978 AS events (Supplementary Fig. 26a). Cell clusters were highly consistent with major cell types defined in the original publication (Supplementary Fig. 26b). Moreover, some cell types could be further distinguished into several subgroups by integrating AS profiles after SCSES imputation (Supplementary Fig. 26c), such as neutrophils (cluster1, 2, 3, 5, 7, 9) and monocytes (Mono1, Mono2, Mono3). The three subgroups of monocytes exhibited clearly different UMAP projections by SCSES-imputed splicing changes, which could not be detected by gene expression clustering (Fig. 6a). This analysis demonstrates the capability and application of SCSES in uncovering AS heterogeneity in droplet-based data.

We next sought to understand the difference between the three monocytes AS subgroups (Supplementary Data 10). Comparing the culture time of cells, Mono3 contained more cells cultured for 2 days, and fewer cells cultured for 6 days (Fig. 6b), suggesting cells in different subgroups may have distinct differentiation efficiency. The



inclusion level of exon 6 in *Runx1* was strikingly elevated in Mono3 (Fig. 6c). *Runx1* is reported to promote the maturity of granulocytic and monocytic cells^{96,97}. Moreover, *Runx1* isoform including exon 6 can promote myeloid lineages commitment of multipotent progenitor cells within the mouse fetal liver^{98,99}. These results suggest that *Runx1* isoforms with alternative exon 6 may also lead to different differentiation speed in monocytes. In addition, Mono3 exhibited higher expression of the primary granules genes, including *Ctsg*, *Elane* and *Prtn3*, indicating a “neutrophil-like” feature¹⁰⁰ (Supplementary Fig. 26d). It is reported that some monocytes share common progenitors with granulocytes¹⁰¹, which was validated by the prevalence of

neutrophils within the 6 day cell population derived from the Mono3 specific clones (Supplementary Note 6, Fig. 6d). Besides, it is known that autophagy is pivotal for monocytes survival and differentiation¹⁰². We identified splicing changes in 19 autophagy-related genes (Fig. 6e). For example, *Lamp2* generates 3 isoforms, *Lamp2a*, *Lamp2b* and *Lamp2c*, by altering the last exon (Supplementary Fig. 26e). Within Mono2, there was a notable enrichment of the *Lamp2a* (Fig. 6f left panel). *Lamp2a* plays an important role in lysosomal docking of HSC70-involved complexes and substrate translocation into the lumen involved in chaperone-mediated autophagy¹⁰³. *Sh3glb1* displayed a preference towards the transcripts missing exon 11 in Mono2,

Fig. 5 | SCSES improves the inference of cell state transition in human embryo development. **a** UMAP of 758 cells by SCSES splicing profile. Colors indicate cell sampling time points along the differentiation from pluripotent state through mesoderm to definitive endoderm. **b** Biological function enrichment analysis of the differentially spliced genes. **c** Heatmaps showing the mean PSI value of top 1000 DSEs in each time point. The color shows the z-score of the PSI value. Clustering at the event level resulted in nine clusters. **d** Pseudobulk read coverage showing the inclusion of *DNMT3B* exon 21, *JMJD1C* intron 6 at each time point. Alternative exons are highlighted in red in the genome annotation track. **e** Splicing regulatory network between differentially expressed RBPs and DSEs. The RBPs annotated as splicing factors and regulating >70 DSEs are selected. Edges whose correlation coefficient is >0.5 are kept. For each RBP, 10 DSEs with strongest correlation are shown. Colors for RBPs indicate different differentiation stages. Edge Colors indicate a positive or negative correlation between RBPs and splicing events.

Line shapes indicate whether the correlation can be detected by raw PSI value. **f** UMAP of 172 cells at 36 h based on transcript splicing (left panel) or gene expression features (right panel), colored by cell clusters using splicing features, shaped by cell clusters using gene expression features. **g** Sankey diagram showing the changes of Seurat clusters from gene expression to transcript splicing. **h** Bubble heatmap showing the expression of *POU5F1*, *SOX17*, and *CXCR4* in the clusters. **i** Violin plot showing the differentiation scores by a gene signature, which exhibits a continuous increase in expression from 0 h to 96 h. The *P*-values are calculated by the Wilcoxon test (two-sided test) without any adjustments. $N_{SC1} = 66$, $N_{SC2} = 65$, $N_{SC3} = 41$. **j** The velocity fields estimated by scVelo. **k** Scatter plots showing the order of differentiation scores (x-axis) versus the PSI of *MCM7*, *USP8*, and *ERBB4* of cells in SC1 and SC2. The correlation coefficient (*R*) and associated *P*-value are calculated by Spearman's correlation analysis ($N = 131$). Source data of panels is provided as a Source Data file.

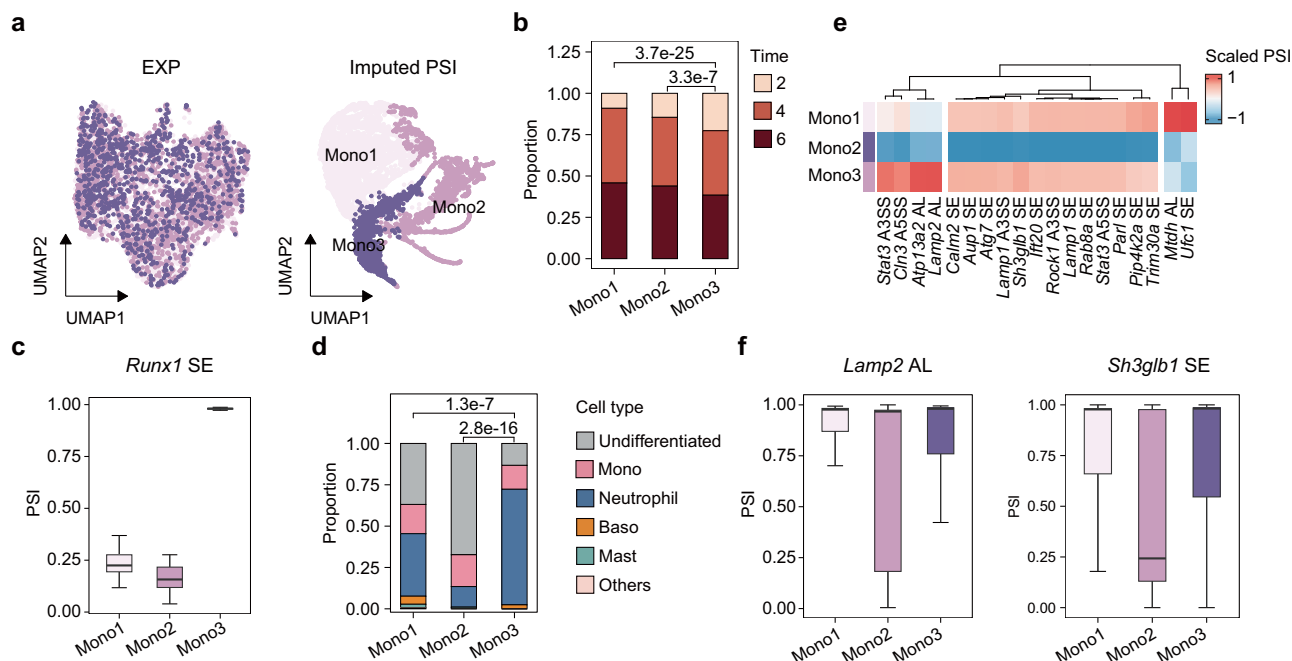


Fig. 6 | SCSES identifies an activated monocyte subtype in induced HSPC differentiation. **a** UMAP plot of 5256 cells from the three monocyte subclasses based on gene expression (left panel) and transcript splicing features (right panel), showing three subclasses: Mono1, Mono2, and Mono3 corresponding to cluster 4, 8, and 10 from Supplementary Fig. 26a. **b** Proportions of the culture time points at which cells were collected in the three monocyte subclasses. The *P*-values are calculated by the one-sided Fisher's exact test without any adjustments. **c** Boxplot of *Runx1* splicing change in the cells of each monocyte subgroup. $N_{Mono1} = 2828$, $N_{Mono2} = 1477$, $N_{Mono3} = 951$. **d** Proportions of the late-stage cell types in each

monocyte subgroup. The *P*-values are calculated by the one-sided Fisher's exact test without any adjustments. **e** Heatmap showing scaled mean PSI among each monocyte subgroup for splicing events associated with autophagy-related genes. AL: alternative last exon event. **f** Boxplot of splicing events *Lamp2* (left panel) and *Sh3glb1* (right panel) in the cells of each monocyte subgroup. $N_{Mono1} = 2828$, $N_{Mono2} = 1477$, $N_{Mono3} = 951$. The boxes in (c) and (f) indicate median (center), Q25, and Q75 (bounds of box), the smallest value within 1.5 times interquartile range below Q25 and the largest value within 1.5 times interquartile range above Q75 (whiskers). Source data of panels is provided as a Source Data file.

leading to an incomplete carboxy-terminal Src-homology 3 (SH3) domain (Fig. 6f right panel). The broken SH3 domain may inhibit the autophagosome formation by interrupting the interaction between *Sh3glb1*, *UVRA*G, and *Beclin1*-*PI3KC3* complex^{104,105}. However, the overall expression levels of *Sh3glb1* and *Lamp2* did not change significantly among the monocyte subgroups (Supplementary Fig. 26f). These results imply that monocytes may regulate the autophagy process via generating splicing variants in associated genes. In summary, these results suggest SCSES can be leveraged to analyze droplet-based scRNA-seq datasets and decipher the transcriptional heterogeneity across large populations of cells.

Discussion

In this study, we introduced a computational method called SCSES to improve the identification and qualification of splicing alterations at

single-cell level. By simulation studies, SCSES outperformed existing methods in recapitulating splicing changes of individual cells and between cell groups. More intriguingly, SCSES demonstrated the ability to reveal new cell clusters with unique splicing pattern that are biologically relevant, which could not be detected by single-cell gene expressions or current methods on splicing analysis.

Notably, SCSES works on data from various single-cell sequencing platforms including droplet-based protocols, which are incompatible with the majority of current methods. These good performances of SCSES could be explained by the innovative data imputation strategies, which consider the specificity of each cell-event pair based on the abundance of alternative junctions in the target cell and neighbor cells. SCSES firstly aggregates the splicing status from both similar cells and similar events, and then estimates the splicing changes of all cell-event pairs, whether this pair is associated with “dropout” or not. The recent

method SCASL only works on events where both inclusion and exclusion junctions are missed. However, such events may take up only <20% of the total splicing events. Raw PSI values of the other events (>80%) could also be incorrect due to the limited junction reads counts in each single-cell (Supplementary Fig. 27). Finally, SCSES integrates imputation results on each cell-event pair with different biological scenarios, under which alternative choices of similarity and imputation methods have distinct performances (Supplementary Figs. 7, 8). Through systematic comparison of the integrated strategy versus individual strategies, we demonstrated that the integrated approach improved PSI estimation accuracy and enhanced clustering consistency in most cases (Supplementary Fig. 28). Additionally, several carefully designed computational procedures, such as similarity diffusion and fine-tune procedure of the scenario decision model, provided additional performance enhancements, validated by ablation studies (Supplementary Figs. 29–30, Supplementary Note 7). Furthermore, it is well recognized that different splicing event detection tools often produce markedly different outputs, which can lead to conflicting interpretations (Supplementary Fig. 14e–h, 16a)¹⁰⁶. To address this limitation, SCSES has been designed as an open framework that allows users to incorporate event lists generated by their preferred detection tools. Besides defining novel cell clusters, SCSES may also contribute to infer cell evolution trajectory. We would use SCSES to construct the relationships between gene expression, AS events and cell pseudotime, as well as to detect critical RBPs and splicing events that drive cell evolution.

There are multiple strategies to improve SCSES. In the current SCSES Strategy 1, raw PSI values are used directly. However, the reliability of a PSI value largely depends on its supporting junction read counts. PSI values with higher reliability provide more accurate information for imputing splicing profiles in similar cells. Therefore, incorporating PSI confidence scores into SCSES could enhance the accuracy of the estimated PSI values. SCSES highly relies on the cell splicing similarity network, of which the accuracy may determine the final performance. In SCSES, we used the gene expressions of RBPs, raw PSI, or raw junction read counts to calculate cell splicing similarity in current version. RBPs are crucial regulators of RNA splicing process, which bind directly to specific RNA sequences and control key steps such as splice site selection and spliceosome assembly. However, transcript-level RBP expression may only indirectly reflect the actual functional splicing activity within individual cells. In this context, junction read counts and PSI values provide a more direct representation of cellular splicing states, while their quantitative accuracy is constrained by the inherent sparsity of junction reads in single-cell data. Moreover, RBP expression is only one of many layers involved in splicing regulation, other factors, such as post-transcriptional and post-translational modifications of RBPs, RNA modifications (e.g., m⁶A), and splicing-associated noncoding RNAs, can also affect the splicing process. Therefore, integrating these diverse regulatory factors into our modeling framework would improve the depiction of cellular splicing heterogeneity. Another way is to implant the cell splicing similarity from some optimization problems instead of defining it directly. For instance, gene expression imputation can be achieved by non-negative matrix factorization (NMF)¹⁰⁷, which inspires us to use multiview NMF algorithms to impute gene expression and splicing changes simultaneously. Similarly, gene expression and splicing similarities can be modeled by searching the optimal cell embeddings that give stable cell clusters in the latent spaces. Here, a number of deep learning algorithms, like autoencoder, can also be applied. Furthermore, despite the good performance of SCSES on droplet-based data, it is still difficult to estimate the intensities of splicing events that are away from the 3'-end of transcripts. To address this, generative artificial intelligence models can be used. For example, we can use large-scale full-length single-cell data or bulk data to pre-train a variational autoencoder, which captures the association between gene expression and splicing events. Then, droplet-based data could be used to fine-tune the model to generate the complete splicing alterations.

Methods

SCSES pipeline

Data processing. The input data of SCSES is sequence alignment files (SAM/BAM) of scRNA-seq data. For full-length data, gene expressions are extracted by featureCounts¹⁰⁸ and normalized by TPM (Transcripts Per Million). For UMI-base data, gene expressions are extracted by UMI-tools and normalized by library size¹⁰⁹. Junction read counts are calculated by a JAVA program based on HTSJDK packages (<https://github.com/samtools/htsjdk>). Junction read counts are normalized by library size to account for sequencing depth variability.

Splicing event detection. SCSES employs three widely adopted algorithms to identify different AS types: rMATS¹⁹ for skipped exon (SE) events and mutually exclusive exons (MXE) events, MAJIQ²⁰ for alternative 3'/5' splice site (A3SS/A5SS) events, and IRFinder²¹ for retained intron (RI) events.

Cell splicing similarity. SCSES collects the splicing information from cells with similar splicing patterns for imputation. RNA-binding proteins are critical regulators of RNA splicing process, and the gene expression of RBPs can reflect the cell splicing states. Moreover, the junction read counts of events, as well as PSI values, can directly represent cellular splicing intensity. Hence, SCSES uses these three features to quantify cell-cell splicing similarities. For input feature matrix $\mathbf{D}_{M \times N}$ with M cells and N features, SCSES performs dimension reduction with PCA. It keeps the principal components (PC) until the cumulative explained variance exceeds 0.9, resulting in a reduced matrix $\hat{\mathbf{D}}_{M \times P}$, where P is the number of retained PCs. The distance between cell c_1 and cell c_2 is calculated by Euclidean distance:

$$\text{Dis}(c_1, c_2) = \sqrt{\sum_i (\hat{\mathbf{D}}(c_1, i) - \hat{\mathbf{D}}(c_2, i))^2} \quad (1)$$

To identify cells with the most similar splicing pattern, SCSES selects the K-nearest neighbor (KNN) cells based on Dis . Since cell counts are usually variable for different cell types, a dynamic- K strategy is used to adaptively find the appropriate k -value for each cell. Specifically, for a cell c , the distances to other cells are sorted in ascending order, denoted as $\text{dis}_c^{\text{asc}} = (d_1, d_2, \dots, d_{M-1})$. Then SCSES computes the differences between successive distances, and generate a difference vector $\mathbf{df}_c = (d_2 - d_1, d_3 - d_2, \dots, d_{M-1} - d_{M-2}) = (\Delta d_1, \Delta d_2, \dots, \Delta d_{M-2})$. Next, SCSES calculates the cumulative average of each position in $\mathbf{df}_c$, denoted as $\mathbf{cdf}_c = (\sum_{i=1}^1 \Delta d_i, \frac{1}{2} \sum_{i=1}^2 \Delta d_i, \dots, \frac{1}{M-2} \sum_{i=1}^{M-2} \Delta d_i)$. SCSES computes the average value of positive difference between $\mathbf{df}_c$ and $\mathbf{cdf}_c$ across all offset pairs, as $\bar{\Delta}_c = \frac{1}{R} \sum_i (\text{relu}(\mathbf{df}_c(i+1) - \mathbf{cdf}_c(i)))$, where relu is the linear rectification function and R represents the number activated by relu function. Finally, SCSES searches the elements in $\mathbf{cdf}_c$ from beginning to find the first position k_c that makes $\mathbf{df}_c(k_c) - \mathbf{cdf}_c(k_c - 1) > \bar{\Delta}_c$, where k_c is the k -value for cell c . SCSES constructs the KNN distance matrix Dis_{KNN} by retaining the distances to the k nearest neighbors for each cell. Dis_{KNN} is transferred to a similarity matrix by Gaussian kernel function:

$$\mathbf{S}_{\text{cell}}(\mathbf{c}_1, \mathbf{c}_2) = e^{-\left(\frac{\text{Dis}_{\text{KNN}}(\mathbf{c}_1, \mathbf{c}_2)}{\sigma_{c_1}}\right)^2} \quad (2)$$

where σ_{c_1} is a zoom factor for cell c_1 , set to be $d_{[k_{c_1}/3]+1}$ ³¹. Due to the asymmetry of $\mathbf{S}_{\text{cell}}$, SCSES symmetrizes the cell splicing similarity matrix by $\mathbf{S}_{\text{cell}} \leftarrow \mathbf{S}_{\text{cell}} + \mathbf{S}_{\text{cell}}'$.

The RBP list of human was manually curated from RBPbase (<https://apps.ebi.ac.uk/rbpbase/>) and a well-used resource¹¹⁰, while the RBPs for mouse were collected from SFMetaDB¹¹¹ and RBPDB¹¹². SCSES supports user-defined RBP list for different species.

Event similarity. SCSES integrates the event sequence similarity and splicing regulation information to measure the global event-event similarities. Sequence features for SE events are collected from BRIE²³, and similar features for other event types are designed by imitating the characteristics of SE events (Supplementary Note 1, Supplementary Data 11). For each event type, SCSES trains an auto-encoder model to extract features and obtains the latent embedding matrix for further processing. Splicing regulation features are defined by the correlation between RBP expression and events PSI. The Pearson correlation coefficients (PCC) between RBPs and events are computed for further processing.

For an event type et with E events, set $Emb_{E \times L}$ as the sequence feature embedding matrix, and $PER_{E \times B}$ as the PCC matrix between events and RBP, where L is the embedding dimension, and B is the count of RBPs. The distances between two events e_1 and e_2 in Emb and PER are calculated by the normalized Euclidean distance (NED) respectively, defined as:

$$NED_{Emb}(e_1, e_2) = \sqrt{\sum_i \frac{(Emb(e_1, i) - Emb(e_2, i))^2}{Var(Emb(\cdot, i))}} \quad (3)$$

$$NED_{PER}(e_1, e_2) = \sqrt{\sum_i \frac{(PER(e_1, i) - PER(e_2, i))^2}{Var(PER(\cdot, i))}} \quad (4)$$

where $Var(\cdot)$ is the variance of the feature. To merge NED_{Emb} and NED_{PCC} , a combination coefficient ω is calculated by:

$$\omega(e_1, e_2) = \frac{1}{1 + e^{a \times p(e_1, e_2) + b}} \quad (5)$$

where $p(e_1, e_2)$ is the P -value of PCC between $PER(e_1, \cdot)$ and $PER(e_2, \cdot)$, $a = -2\ln(9999)$ and $b = \ln(9999)$ in SCSES. The integrated distance between events is merged by

$$Dis_{event}(e_1, e_2) = (1 - \omega(e_1, e_2)) \times NED_{Emb}(e_1, e_2) + \omega(e_1, e_2) \times NED_{PER}(e_1, e_2) \quad (6)$$

SCSES keeps distances of nearest events with fixed k_{event} (10 by default, Supplementary Fig. 31, Supplementary Note 1), and set others to be infinite, obtaining matrix Dis_{event_KNN} . Dis_{event_KNN} is transferred to a similarity matrix by Gaussian kernel function:

$$S_{event}(e_1, e_2) = e^{-\left(\frac{Dis_{event_KNN}(e_1, e_2)}{\sigma_e}\right)^2} \quad (7)$$

where σ_e is a zoom factor, set to be the $\lceil k_{event}/3 \rceil + 1$ th smallest distance with e_1 . S_{event} is also symmetrize by $S_{event} \leftarrow S_{event} + S_{event}^T$.

Network Diffusion. To obtain the network global similarity, we perform the random walk with restart (RWR) algorithm on S_{cell} and S_{event} , respectively. S_{cell} and S_{event} are normalized by row to obtain transition probability matrix $\hat{S}_{cell}$ and $\hat{S}_{event}$. For a cell c (or an event e), we define the similarity vector at step $t=0$ to be $v_c^0 = \hat{S}_{cell}(c, \cdot)$ ($v_e^0 = \hat{S}_{event}(e, \cdot)$). The similarity information is diffused in the whole network with:

$$v_c^{t+1} \leftarrow (1 - \lambda)v_c^t + \lambda \hat{S}_{cell} v_c^0 \quad (8)$$

$$v_e^{t+1} \leftarrow (1 - \lambda)v_e^t + \lambda \hat{S}_{event} v_e^0 \quad (9)$$

where λ is the restart probability (0.2 by default, Supplementary Fig. 32, Supplementary Note 7). The random walk is carried out for all cells (and events), and we can obtain the similarity matrix $\tilde{S}_{cell}^t$ and $\tilde{S}_{event}^t$ at step t . After each iteration, we calculate the change in the

similarity matrix by

$$\Delta_{cell}(t) = \frac{SSE(\tilde{S}_{cell}^t, \tilde{S}_{cell}^{t-1})}{SST(\tilde{S}_{cell}^t, \tilde{S}_{cell}^{t-1})} \quad (10)$$

$$\Delta_{event}(t) = \frac{SSE(\tilde{S}_{event}^t, \tilde{S}_{event}^{t-1})}{SST(\tilde{S}_{event}^t, \tilde{S}_{event}^{t-1})} \quad (11)$$

where SSE is the sum of squares error and SST is the total sum of squares. The walk is stopped when $\Delta(t)$ is less than a certain threshold, which is 0.05 in SCSES by default (Supplementary Fig. 33, Supplementary Note 7). After obtaining the diffused similarity $\tilde{S}_{cell}$ and $\tilde{S}_{event}$, we still keep the k most similar elements for each row, and others are set to be 0. The k values here are the same as those in S_{cell} and S_{event} , respectively.

Imputation. The PSI value is defined as the ratio between inclusion read counts and total read counts of the event (Supplementary Fig. 1). SCSES employs multiple strategies for PSI imputation, combining different similarities. For an event e consisted of $IJC = \{ijc_1, ijc_2, \dots, ijc_l\}$ (for l inclusion junctions) and $EJC = \{ejc_1, ejc_2, \dots, ejc_j\}$ (for j exclusion junctions), three strategies are used for PSI imputation.

Strategy 1: given RC_{ijc}^{raw} and RC_{ejc}^{raw} as the raw normalized read count matrices of IJC and EJC , respectively, the raw PSI of event e in cell c , $PSI^{raw}(c, e)$, is defined as:

$$PSI^{raw}(c, e) = \frac{\frac{1}{l} \sum_{ijc \in IJC} RC_{ijc}^{raw}(c, ijc)}{\frac{1}{l} \sum_{ijc \in IJC} RC_{ijc}^{raw}(c, ijc) + \frac{1}{j} \sum_{ejc \in EJC} RC_{ejc}^{raw}(c, ejc)} \quad (12)$$

We then impute the raw PSI with cell similarities as follow:

$$PSI_{PSI}^{cell} = \hat{S}_{cell} \times PSI^{raw} \quad (13)$$

Strategy 2: we first impute raw inclusion and exclusion read counts RC_{ijc}^{cell} and RC_{ejc}^{cell} , respectively, and calculate the imputed PSI of event e in cell c by cell similarities, which is formalized as:

$$RC_{ijc}^{cell} = \hat{S}_{cell} \times RC_{ijc}^{raw} \quad (14)$$

$$RC_{ejc}^{cell} = \hat{S}_{cell} \times RC_{ejc}^{raw} \quad (15)$$

$$PSI_{RC}^{cell}(c, e) = \frac{\frac{1}{l} \sum_{ijc \in IJC} RC_{ijc}^{cell}(c, ijc)}{\frac{1}{l} \sum_{ijc \in IJC} RC_{ijc}^{cell}(c, ijc) + \frac{1}{j} \sum_{ejc \in EJC} RC_{ejc}^{cell}(c, ejc)} \quad (16)$$

Strategy 3: we combine cell similarities and event similarities for imputation. Specifically, we first impute the PSI using strategy 2, then further impute the results using event similarities:

$$PSI^{both} = \left(\left(\hat{S}_{event} \right)^T \times \left(PSI_{RC}^{cell} \right)^T \right)^T \quad (17)$$

For each imputation strategy, the imputation procedure is performed iteratively until the change between the imputed PSI matrices of two consecutive steps meets a convergence threshold. The loss function is identical to that used in network diffusion, with a default threshold value of 0.05.

Prediction of event-cell pair scenarios. To determine the groups that event-cell pairs belong to, we use cascade decision models, including a well-defined model (Model0) and two logistical regression models (Model1, Model2) for the prediction. Model0 is used to determine if

pairs belong to ND (non dropout) or WD (with dropout). Pairs with PSI values not equal to 0 or 1 are classified as ND, and others are in WD. In SCSES, the PSI values of pairs without any supporting junction reads (neither inclusion nor exclusion junction) are also set to 0. Model1 is used to predict the probabilities of BD (biological dropout) and TD (technical dropout), while Model2 is used to predict probabilities of TD-Info and TD+Info. In Model1, 17 features (Supplementary Data 12, Supplementary Fig. 34) are collected to reflect the distribution of junction read counts and PSI values of an event in the target cell and its neighbor cells. In Model2, two more features are added to reflect the effect of cell-similarity imputation (Supplementary Data 12, Supplementary Fig. 34). All features are normalized by min-max normalization before training.

To train these models, we collected scRNA-seq data from five cell lines^{35,36} and corresponding bulk RNA-seq data from NCBI Gene Expression Omnibus (GEO) and CCLE. By comparing PSI values between scRNA-seq data and corresponding bulk data, event-cell pairs in scRNA-seq data are separated into four groups. Two logistical models are pre-trained with corresponding features extracted from these single-cell datasets (Supplementary Fig. 35).

To improve the model adaptability for the new dataset, we also provide a procedure to fine-tune these models. We collected a set of splicing events with conserved splicing levels in different human tissues. First, we collected the housekeeping genes¹¹³, essential genes¹¹⁴, and genes without tissue specificity (<https://zenodo.org/records/6408906>) from published database. We selected genes presented in all three databases for further processing. Then, we removed genes associated with development or differentiation in GO:BP terms. Next, we collected all annotated splicing events in these genes from GTF file and kept the events whose variation of PSI values among all tissues in GTEx Splicemap were $<10^{-4}$. Finally, we obtained 344 conserved events from 92 genes. We recorded the splicing levels of these 344 events as a reference. For a new dataset, we compare the splicing status of the conserved events in new data with the reference records and give the group definition to each event-cell pair, which is used to fine-tune the pre-trained model above.

Estimation. SCSES employs different imputation strategies based on the event-cell pair scenarios: strategy 1 for pairs in ND are imputed with strategy 1, strategy 2 for pairs in BD and TD+Info, and strategy 3 for pairs in TD-Info. For each event-cell pair, the imputation results from these different strategies are combined, using the probabilities of the pair belonging to different groups as coefficients. The final imputation of PSI values is formalized as:

$$PSI_{impute} = P_{ND} \odot PSI_{PSI}^{cell} + (1 - P_{ND}) \odot (P_{BD} \odot PSI_{RC}^{cell} + (1 - P_{BD}) \odot (P_{TD+Info} \odot PSI_{RC}^{cell} + (1 - P_{TD+Info}) \odot PSI^{both})) \quad (18)$$

where $\odot$ is the element-wise product of two matrices, P_{ND} is an indicator matrix for ND group, $\mathbf{1}$ is a matrix with all elements equal to 1, P_{BD} and $P_{TD+Info}$ are the probability matrices for group BD and TD+Info, respectively.

Evaluation

Dataset preparation. We prepared multiple datasets for the evaluation of SCSES. Bulk sequencing data for four cell lines (HCT116, HCC1954, HepG2, and HL-60) was downloaded from CCLE to serve used as benchmarking datasets. We generated two datasets as test data for evaluation. (1) the real scRNA-seq data of the four corresponding cell lines from GEO (Supplementary Fig. 9a), (2) an artificial synthetic dataset by spanki simulator (Supplementary Fig. 9b). For the latter, we proposed a pipeline to simulate the isoform expression in individual cells from four cell lines by referring isoform expressions of corresponding bulk sequencing data and integrating the noise and low read coverage of

scRNA-seq data. The Spanki simulator³⁷ was used to generate bam files of different cells according to the isoform expression profiles (Supplementary Note 2). To assess the performance of SCSES in the real heterogeneous single-cell environment, we collected two datasets where paired short-read and long-read sequencing libraries were constructed from individual cells, one of which was derived from high-grade serous ovarian carcinoma tissues³⁸ (sample P1, 8 cell types), and the other was from adult human hippocampus³⁹ (sample f1, 16 cell types).

To assess the biological significance, we downloaded three additional scRNA-seq datasets (nPSC, hEE, and iPSC) from GEO. These datasets had average library sizes ranging from 5.6 to 25 million reads per cell (Supplementary Fig. 36a). Down-sampling was performed on these datasets to generate synthetic data with different ratios: 0.15 for the nPSC, 0.03 for the hEE, and 0.1 for the iPSC (Supplementary Fig. 9d). The down-sampling process ensured the resulting data had similar sequencing abundance with standard scRNA-seq data¹¹⁵ (Supplementary Fig. 36a).

Evaluation of PSI accuracy. In the real scRNA-seq dataset, the splicing events detected by different methods were significantly varied, and most methods only considered SE events. Therefore, for a fair comparison, we only used the events identified by SCSES and each individual compared method. Then, the events with high confidence in bulk datasets were selected for evaluation based on the following procedure: (1) the events should be detected by pairwise-comparison algorithms (e.g., SCSES and rMATS) in both single-cell dataset and matched bulk dataset; (2) the junction read counts supporting either inclusion or exclusion exons should exceed 20 in bulk dataset; (3) when read counts of either upstream or downstream inclusion junction have fewer than 5 reads in bulk dataset, the read count fold change between them should be <10 ; (4) the genes harboring events should be expressed in both single-cell dataset and matched bulk dataset; (5) removing splicing events exhibiting cell-cycle dependence. Specifically, we curated eight cell cycle-synchronized RNA-seq datasets (GSE123958, GSE81485, GSE143275, GSE216497, GSE97774, GSE94479, GSE116131 and PRJEB7566) from GEO and European Nucleotide Archive (ENA). Among these datasets, five contained cell cycle phase annotations in their original publications, while the “CellCycleScoring” function from the “Seurat” package was used to assign cell cycle phases for the remaining three datasets (Supplementary Data 13). Differential splicing analysis was then conducted through pairwise comparisons between all cell cycle phases. Cell cycle-regulated splicing events were defined based on the following criteria: $\Delta PSI > 0.2$ and $P < 0.05$ (Wilcoxon test), with consistent direction of splicing changes across all samples for each phase pair. Subsequently, AS events associated with cell cycle were excluded from evaluation. The event counts for each comparison and each cell are listed in Supplementary Data 2. We computed two metrics: Spearman correlation coefficient (SCC) in cells and root mean squared error (RMSE) in events, to assess the discrepancy between estimated PSI values and the benchmarks. The SCC between the imputed PSI and benchmark PSI in cell c is defined as:

$$SCC_c = 1 - \frac{6 \sum_{i=1}^n (IPR_c(i) - BPR_c(i))^2}{n(n^2 - 1)} \quad (19)$$

where $IPR_c(i)$ and $BPR_c(i)$ are the imputed PSI rank and benchmark PSI rank of event i in cell c , respectively, n is the event count. And the RMSE between imputation PSI IP_e and benchmark PSI BP_e in event e is defined as:

$$RMSE_e = \sqrt{\frac{1}{M} \sum_i (IP_e(i) - BP_e(i))^2} \quad (20)$$

where $IP_e(i)$ and $BP_e(i)$ are the imputed and benchmark PSI values of event e in cell i , respectively.

In the synthetic dataset, which had known SE events (detected by SUPPA¹⁶) as ground truths (Supplementary Note 2), the union of events detected by different methods are used to calculate the accuracy score, which is defined to assessing the accuracy of event identification and PSI estimation. The accuracy score of algorithm *alg* in cell *c* is defined as:

$$ACS_{c,alg} = \begin{cases} SCC_{c,alg} \times rec_{alg} & \text{if } SCC_{c,alg} > 0 \\ SCC_{c,alg} \times (0.5 - rec_{alg}) & \text{if } SCC_{c,alg} \leq 0 \end{cases} \quad (21)$$

where rec_{alg} is the events recall rate of algorithm *alg*, and $SCC_{c,alg}$ is the SCC value between *alg* imputed PSI values and benchmarks in cell *c*. The accuracy score of algorithm *alg* is calculated by the summing accuracy scores of all cells.

In paired long-read and short-read datasets, to ensure a fair comparison, we used only the SE events identified by both SCSES and each individual method under comparison. High-confidence event–cell pairs from the long-read data were used as the benchmark; these pairs were defined as splicing events supported by >10 long-read molecules in a given cell. The reference PSI values were computed based on the ratio of long-read molecules including the alternative exon versus the total number of long-read molecules including/excluding the exon.

Evaluation of differentially spliced event detection capacity. Only the events detected by both SCSES and the compared algorithm were used for assessing the capacity of differentially spliced events (DSEs). The events with $\Delta PSI > 0.2$ between two cell lines in bulk sequencing data were designated as the ground-truth DSEs. In the single-cell data, we constructed the receiver operating characteristic (ROC) curves using the *P*-values from the Wilcoxon test as thresholds. The area under the ROC curve (AUC) was used to evaluate the performance of differential splicing event detection. Additionally, we assessed the consistence of dysregulation direction by calculating the SCC of *APSI* between imputation results and benchmarks.

Evaluation of cell clustering and cell trajectory. The cell clustering was executed in Seurat pipeline. For all compared methods, top 30% highly variable splicing events were selected to execute principal component analysis (PCA), and the PCs with cumulated variance >0.9 were obtained for clustering. *k*-means algorithm was used to detect cell clusters, with *k* set to the number of cell types annotated in the original paper ($K_{npsc} = 6$, $K_{HE} = 7$, and $K_{IPSC} = 3$). UMAP was used to visualize cell clusters in each dataset. For SCASL, we used its built-in functions to generate the cell clustering results and UMAP embedding. The normalized mutual information (NMI) was used to evaluate the consistence between splicing-based clusters and reference cell types.

Monocle 3 (v1.3.5) was used to infer cell developmental trajectories based on both gene expression profiles and splicing profiles. The starting cell types for trajectory inference were designated based on biological knowledge, i.e. inner cell mass for the nPSC data, oocyte for the hEE data, and induced pluripotent stem cells for the iPSC data.

Implementation of SCSES application. The analysis of MM, hESC and HSC datasets was conducted using the Seurat package in R. In each dataset, dimension reduction and cluster identification were performed with the same parameters for gene expression and splicing profiles (Supplementary Note 4–6). Enrichment analysis was performed by gProfiler (v0.2.3)¹⁷ and clusterProfiler (v4.6.2)¹⁸ packages in R. DEGs and DSEs were identified with Wilcoxon test. For DEGs, the fold change was calculated by the ratio of mean expressions in two group cells. For DSEs, the ΔPSI was calculated by the difference of mean PSI in two group cells. RNA velocity was performed by scVelo package in Python. Other details are shown in Supplementary Note 4–6.

Statistics and reproducibility. The statistical tests have been described in the sections above, Supplementary Note and in the figure legends. No statistical method was used to predetermine sample size. In multiple myeloma dataset, to focus on the cells with BTZ resistance, cells from patient MM16 were excluded in analysis. In inDrop dataset, to remove cells with low quality, cells with <500 expressed genes or library size <1000 were removed. No additional data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Code availability

The open-source SCSES R package and tutorial are available at GitHub (<https://github.com/lvxuan12/SCSES>), <https://doi.org/10.5281/zenodo.17087025>.

Data availability

All datasets analyzed in this study are publicly available. The bulk RNA-seq data of cell lines were downloaded from CCLE Project from NCBI Sequence Read Archive (SRA) under accession codes SRP186687. The single-cell RNA-seq data HCC1954 and HL60 were downloaded from NCBI SRA under accession codes SRP041736. The single-cell RNA-seq data of HCT116 and HepG2 were downloaded from Gene Expression Omnibus (GEO) under accession codes GSE150993. The paired long-read and short-read of ovarian cancer dataset was downloaded from SRA under accession PRJNA993664. The paired long-read and short-read of human hippocampus dataset was downloaded from The Neuroscience Multi-omic Data Archive (NeMO) under identifier nemo:dat-ho986e6 and nemo:dat-unjyo0u, respectively. The single-cell RNA-seq data of human naïve pluripotent stem cells, human early embryos, and induced human pluripotent stem cells were downloaded from GEO under accession codes GSE171820, GSE36552 and GSE85908, respectively. The cell-cycle synchronized datasets used for removing cell-cycle-dependent splicing events were downloaded from GEO under accession codes GSE123958, GSE81485, GSE143275, GSE216497, GSE97774, GSE94479, GSE116131 and PRJEB7566. The single-cell RNA-seq data of multiple myeloma patients were downloaded from GEO under accession codes GSE110499. The single-cell RNA-seq dataset for validation of multiple myeloma were downloaded from GEO under accession codes GSE118900, GSE9782, GSE24080 and TCGA MMRF project. The single-cell RNA-seq data of H9 human ESC differentiation were downloaded from GEO under accession codes GSE75748. The inDrop-v3 RNA-seq data of mouse hematopoietic stem cell were provided by Caleb Weinreb from Allon Klein Lab in Department of Systems Biology, Harvard Medical School, which can be accessed by Dropbox repository. Source data are provided with this paper.

References

- Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
- Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
- Winkler, E. A. et al. A single-cell atlas of the normal and malformed human brain vasculature. *Science* **375**, eabi7377 (2022).
- Nieto, P. et al. A single-cell tumor immune atlas for precision oncology. *Genome Res* **31**, 1913–1926 (2021).
- Wagner, J. et al. A single-cell Atlas of the tumor and immune ecosystem of human breast cancer. *Cell* **177**, 1330–1345 e1318 (2019).

6. Braun, E. et al. Comprehensive cell atlas of the first-trimester developing human brain. *Science* **382**, eadf1226 (2023).
7. Chao, Y. et al. Regulatory roles and mechanisms of alternative RNA splicing in adipogenesis and human metabolic health. *Cell Biosci.* **11**, 66 (2021).
8. Bonnal, S. C., Lopez-Oreja, I. & Valcarcel, J. Roles and mechanisms of alternative splicing in cancer - implications for care. *Nat. Rev. Clin. Oncol.* **17**, 457–474 (2020).
9. Hu, Y., Wang, K. & Li, M. Detecting differential alternative splicing events in scRNA-seq with or without unique molecular identifiers. *PLoS Comput Biol.* **16**, e1007925 (2020).
10. Wang, E. & Aifantis, I. RNA splicing and cancer. *Trends Cancer* **6**, 631–644 (2020).
11. Keren, H., Lev-Maor, G. & Ast, G. Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* **11**, 345–355 (2010).
12. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
13. Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
14. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev. Biochem.* **72**, 291–336 (2003).
15. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nat. Genet.* **30**, 13–19 (2002).
16. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
17. Deveson, I. W. et al. Universal Alternative Splicing of Noncoding Exons. *Cell Syst.* **6**, 245–255.e245 (2018).
18. Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
19. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–E5601 (2014).
20. Vaquero-Garcia, J. et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**, e11752 (2016).
21. Middleton, R. et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.* **18**, 51 (2017).
22. Brennecke, P. et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
23. Huang, Y. & Sanguinetti, G. BRIE: transcriptome-wide splicing quantification in single cells. *Genome Biol.* **18**, 123 (2017).
24. Wen, W. X., Mead, A. J. & Thongjuea, S. MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data. *Nucleic Acids Res.* **51**, e29 (2023).
25. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
26. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
27. Song, Y. et al. Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. *Mol. Cell* **67**, 148–161.e145 (2017).
28. Huang, Y. & Sanguinetti, G. BRIE2: computational identification of splicing phenotypes from single-cell transcriptomic experiments. *Genome Biol.* **22**, 251 (2021).
29. Buen Abad Najar, C. F., Burra, P., Yosef, N. & Lareau, L. F. Identifying cell state-associated alternative splicing events and their coregulation. *Genome Res.* **32**, 1385–1397 (2022).
30. Xiang, X., He, Y., Zhang, Z. & Yang, X. Interrogations of single-cell RNA splicing landscapes with SCASL define new cell identities with physiological relevance. *Nat. Commun.* **15**, 2164 (2024).
31. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e727 (2018).
32. Wei, J., Zhou, T., Zhang, X. & Tian, T. DTFLOW: Inference and visualization of single-cell pseudotime trajectory using diffusion propagation. *Genomics Proteom. Bioinform.* **19**, 306–318 (2021).
33. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
34. Ghandi, M. et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508 (2019).
35. Wang, X., Yu, L. & Wu, A. R. The effect of methanol fixation on single-cell RNA sequencing data. *BMC Genomics* **22**, 420 (2021).
36. Pollen, A. A. et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).
37. Sturgill, D. et al. Design of RNA splicing analysis null models for post hoc filtering of Drosophila head RNA-Seq data with the splicing analysis kit (Spanki). *BMC Bioinform.* **14**, 320 (2013).
38. Byrne, A. et al. Single-cell long-read targeted sequencing reveals transcriptional variation in ovarian cancer. *Nat. Commun.* **15**, 6916 (2024).
39. Joglekar, A. et al. Single-cell long-read sequencing-based mapping reveals specialized splicing patterns in developing and adult mouse and human brain. *Nat. Neurosci.* **27**, 1051–1063 (2024).
40. Mochizuki, Y. et al. Alternative microexon splicing by RBFOX2 and PTBP1 is associated with metastasis in colorectal cancer. *Int J. Cancer* **149**, 1787–1800 (2021).
41. Xu, T. et al. Alternative splicing downstream of EMT enhances phenotypic plasticity and malignant behavior in colon cancer. *Elife* **11**, e82006 (2022).
42. He, Z. et al. PTBP1 promotes hepatocellular carcinoma progression by regulating the skipping of exon 9 in NUMB pre-mRNA. *Heliyon* **9**, e17387 (2023).
43. Bechara, E. G., Sebestyen, E., Bernardis, I., Eyra, E. & Valcarcel, J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell* **52**, 720–733 (2013).
44. Yanagida, A. et al. Naive stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell* **28**, 1016–1022.e1014 (2021).
45. Yan, L. et al. Single-cell RNA-Seq profiling of human pre-implantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
46. Liu, H. L. et al. The role of RNA splicing factor PTBP1 in neuronal development. *Biochim. Biophys. Acta Mol. Cell Res.* **1870**, 119506 (2023).
47. Vuong, C. K., Black, D. L. & Zheng, S. The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016).
48. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
49. Gulati, G. S. et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405–411 (2020).
50. Ye, H. et al. Retromer subunit, VPS29, regulates synaptic transmission and is required for endolysosomal function in the aging brain. *Elife* **9**, e51977 (2020).
51. Clower, C. V. et al. The alternative splicing repressors hnRNP A1/A2 and PTB influence pyruvate kinase isoform expression and cell metabolism. *Proc. Natl. Acad. Sci. USA* **107**, 1894–1899 (2010).
52. Guo, J., Jia, J. & Jia, R. PTBP1 and PTBP2 impaired autoregulation of SRSF3 in cancer cells. *Sci. Rep.* **5**, 14548 (2015).
53. Deng, K., Yao, J., Huang, J., Ding, Y. & Zuo, J. Abnormal alternative splicing promotes tumor resistance in targeted therapy and immunotherapy. *Transl. Oncol.* **14**, 101077 (2021).
54. Fan, J. et al. Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* **28**, 1217–1227 (2018).

55. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
56. Ngo, N. T. et al. The significance of tumour cell immunophenotype in myeloma and its impact on clinical outcome. *J. Clin. Pathol.* **62**, 1009–1015 (2009).
57. Guo, J. et al. SRRM2 may be a potential biomarker and immunotherapy target for multiple myeloma: a real-world study based on flow cytometry detection. *Clin. Exp. Med.* **24**, 28 (2024).
58. Zingoni, A. et al. Genotoxic stress induces senescence-associated ADAM 10-dependent release of NKG2d MIC ligands in multiple myeloma cells. *J. Immunol.* **195**, 736–748 (2015).
59. Shuaishuai, Z. et al. AURKA regulates the mechanism of ATR mediated DNA damage response about drug resistance in multiple myeloma. *Blood* **130**, 5354 (2017).
60. Gu, C. et al. CHEK1 and circCHEK1_246aa evoke chromosomal instability and induce bone lesion formation in multiple myeloma. *Mol. Cancer* **20**, 84 (2021).
61. Waku, T. et al. NRF3-POMP-20S proteasome assembly axis promotes cancer development via ubiquitin-independent proteolysis of p53 and retinoblastoma protein. *Mol. Cell Biol.* **40**, e00597-19 (2020).
62. Duechler, M. et al. Induction of apoptosis by proteasome inhibitors in B-CLL cells is associated with downregulation of CD23 and inactivation of Notch2. *Leukemia* **19**, 260–267 (2005).
63. Xu, D. et al. Dll1/Notch activation contributes to bortezomib resistance by upregulating CYP1A1 in multiple myeloma. *Biochem. Biophys. Res. Commun.* **428**, 518–524 (2012).
64. Ortiz-Ruiz, A. et al. Increase of mitochondrial activity contributes to the bortezomib-relapsed in multiple myeloma, a novel therapeutic opportunity. *Blood* **134**, 4408 (2019).
65. Bonvini, P., Zorzi, E., Basso, G. & Rosolen, A. Bortezomib-mediated 26S proteasome inhibition causes cell-cycle arrest and induces apoptosis in CD-30+ anaplastic large cell lymphoma. *Leukemia* **21**, 838–842 (2007).
66. Farrell, M. L. & Reagan, M. R. Soluble and cell-cell-mediated drivers of proteasome inhibitor resistance in multiple myeloma. *Front Endocrinol. (Lausanne)* **9**, 218 (2018).
67. Prelowska, M. K. et al. Inhibition of the L-glutamine transporter ASCT2 sensitizes plasma cell myeloma cells to proteasome inhibitors. *Cancer Lett.* **507**, 13–25 (2021).
68. Ruckrich, T. et al. Characterization of the ubiquitin-proteasome system in bortezomib-adapted cells. *Leukemia* **23**, 1098–1105 (2009).
69. Moiseeva, E. P., Leyland, M. L. & Bradding, P. CADM1 isoforms differentially regulate human mast cell survival and homotypic adhesion. *Cell Mol. Life Sci.* **69**, 2751–2764 (2012).
70. Otsuka, M., Mizuki, M., Fujita, J., Kang, S. & Kanakura, Y. Constitutively active FGFR3 with Lys650Glu mutation enhances bortezomib sensitivity in plasma cell malignancy. *Anticancer Res.* **31**, 113–122 (2011).
71. Kowalski, J. et al. Integrative, multi-platform, whole-genome analyses identify clinically relevant common- and cell-specific signatures in multiple myeloma. *Blood* **120**, 3974–3974 (2012).
72. Nogami, A. et al. Inhibition of the STAT5/pim kinase axis enhances cytotoxic effects of proteasome inhibitors on FLT3-ITD-positive AML cells by cooperatively inhibiting the mTORC1/4EBP1/S6K/Mcl-1 pathway. *Transl. Oncol.* **12**, 336–349 (2019).
73. Morgan, J. J., Kennedy, B. G., Williams, R., Mills, K. I. & Crawford, L. J. Targeting HUWE1 in multiple myeloma leads to impaired replicative stress response and enhanced sensitivity to bortezomib. *Blood* **142**, 2774 (2023).
74. Gabut, M. et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147**, 132–146 (2011).
75. Tahmasebi, S. et al. Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc. Natl. Acad. Sci. USA* **113**, 12360–12367 (2016).
76. Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173 (2016).
77. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
78. Bebee, T. W. et al. The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *Elife* **4**, e08954 (2015).
79. Wang, J. et al. MicroRNA-125b/Lin28 pathway contributes to the mesendodermal fate decision of embryonic stem cells. *Stem Cells Dev.* **21**, 1524–1537 (2012).
80. Guo, C. et al. Nanog RNA-binding proteins YBX1 and ILF3 affect pluripotency of embryonic stem cells. *Cell Biol. Int.* **40**, 847–860 (2016).
81. Suckale, J. et al. PTBP1 is required for embryonic development before gastrulation. *PLoS ONE* **6**, e16992 (2011).
82. Xie, W. et al. Crucial roles of different RNA-binding hnRNP proteins in stem cells. *Int. J. Biol. Sci.* **17**, 807–817 (2021).
83. Han, H. et al. MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**, 241–245 (2013).
84. Grifone, R., Shao, M., Saquet, A. & Shi, D.-L. RNA-binding protein Rbm24 as a multifaceted post-transcriptional regulator of embryonic lineage differentiation and cellular homeostasis. *Cells* **9**, 1891 (2020).
85. Warzecha, C. C. et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J.* **29**, 3286–3300 (2010).
86. Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
87. Jun, Y. et al. Comprehensive analysis of alternative splicing in gastric cancer identifies epithelial-mesenchymal transition subtypes associated with survival. *Cancer Res.* **82**, 543–555 (2022).
88. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
89. Zhang, Y. et al. Phenotypic transition of tumor cells between epithelial- and mesenchymal-like state during adaptation to acidosis. *Cell Cycle* **18**, 1938–1947 (2019).
90. Gu, H. et al. USP8 maintains embryonic stem cell stemness via deubiquitination of EPG5. *Nat. Commun.* **10**, 1465 (2019).
91. Elenius, K. et al. Characterization of a naturally occurring ErbB4 isoform that does not bind or activate phosphatidylinositol 3-kinase. *Oncogene* **18**, 2607–2615 (1999).
92. McLean, A. B. et al. Activin efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem Cells* **25**, 29–38 (2007).
93. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
94. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381 (2020).
95. Sala Frigerio, C. et al. The major risk factors for Alzheimer's disease: age, sex, and genes modulate the microglia response to Aβ plaques. *Cell Rep.* **27**, 1293–1306 e1296 (2019).
96. Behrens, K. et al. Runx1 downregulates stem cell and megakaryocytic transcription programs that support niche interactions. *Blood, J. Am. Soc. Hematol.* **127**, 3369–3381 (2016).
97. Huang, G. et al. PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat. Genet.* **40**, 51–60 (2008).

98. Komeno, Y. et al. Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood* **123**, 3760–3769 (2014).
99. Ghanem, L. R. et al. Poly(C)-binding protein pcbb2 enables differentiation of definitive erythropoiesis by directing functional splicing of the runx1 transcript. *Mol. Cell Biol.* **38**, e00175–00118 (2018).
100. Xie, X. et al. Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. *Nat. Immunol.* **21**, 1119–1133 (2020).
101. Yáñez, A. et al. Granulocyte-monocyte progenitors and monocyte-dendritic cell progenitors independently produce functionally distinct monocytes. *Immunity* **47**, 890–902.e894 (2017).
102. Zhang, Y., Morgan, M. J., Chen, K., Choksi, S. & Liu, Z. -g Induction of autophagy is essential for monocyte-macrophage differentiation. *Blood J. Am. Soc. Hematol.* **119**, 2895–2905 (2012).
103. Kaushik, S. & Cuervo, A. M. The coming of age of chaperone-mediated autophagy. *Nat. Rev. Mol. Cell Biol.* **19**, 365–381 (2018).
104. Takahashi, Y. et al. Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis. *Nat. cell Biol.* **9**, 1142–1151 (2007).
105. Takahashi, Y., Meyerkord, C. L. & Wang, H. G. Bif-1/endophilin B1: a candidate for crescent driving force in autophagy. *Cell Death Differ.* **16**, 947–955 (2009).
106. Liu, Z. & Rabadan, R. Computing the role of alternative splicing in cancer. *Trends Cancer* **7**, 347–358 (2021).
107. Elyanow, R., Dumitrascu, B., Engelhardt, B. E. & Raphael, B. J. netNMF-sc: leveraging gene-gene interactions for imputation and dimensionality reduction in single-cell expression analysis. *Genome Res* **30**, 195–204 (2020).
108. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
109. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**, 491–499 (2017).
110. Sebestyen, E. et al. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res* **26**, 732–744 (2016).
111. Li, J. et al. SFMetaDB: a comprehensive annotation of mouse RNA splicing factor RNA-Seq datasets. *Database (Oxford)* **2017**, bax071 (2017).
112. Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* **39**, D301–308 (2011).
113. Hounkpe, B. W., Chenou, F., de Lima, F. & De Paula, E. V. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res.* **49**, D947–D955 (2021).
114. Luo, H. et al. DEG 15, an update of the database of essential genes that includes built-in analysis tools. *Nucleic Acids Res.* **49**, D677–D686 (2021).
115. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris: The Tabula Muris Consortium. *Nature* **562**, 367 (2018).
116. Alamancos, G. P., Pages, A., Trincado, J. L., Bellora, N. & Eyra, E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* **21**, 1521–1531 (2015).
117. Kolberg, L. et al. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207–W212 (2023).
118. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

Acknowledgements

This work is supported by the National Key R&D Program of China (grant 2023YFF0725400, grant 2025YFF1207901 and 2022YFC2704202 to Z.L.), the National Natural Science Foundation of China (grant 32300558 to X.W., 62403390 to P.Z.W.), the Beijing Natural Science Foundation (grant Z220012 to Z.L.) and Chinese Academy of Sciences Hundred Talents Program (to Z.L.). We acknowledge all participants involved in the study.

Author contributions

Z.L. conceived and supervised the entire project. X.W. and X.L. designed and implemented the method. P.Z.W. contributed to the method development. X.L. performed the evaluation on real datasets. N.H. performed the evaluation on synthetic datasets. X.L. and W.X. performed the application of MM, hESC, and mHSC. D.G. and L.Z. downloaded and prepared datasets for evaluation and application. X.W. and X.L. wrote the manuscript with the feedback from Z.L. P.Z.W. modified the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-64517-5>.

Correspondence and requests for materials should be addressed to Zhaoqi Liu.

Peer review information *Nature Communications* thanks Sika Zheng, who co-reviewed with Naoto Kubota, Emma Busarello, who co-reviewed with Toma Tebaldi and Gabriele Tornè; and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025