Cell Systems

Synthesis

Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data

Highlights

- Sixteen cell-potency methods categorized as average, entropy, and knowledge based
- Unified notation reveals conceptual and implementation differences across methods
- Analyses show effects of preprocessing and protein-protein interaction priors
- Current benchmarks lack single-cell-resolution ground-truth potency labels

Authors

Qingyang Wang, Zhiqian Zhai, Qiuyu Lian, Dongyuan Song, Jingyi Jessica Li

Correspondence

lijy03@fredhutch.org

In brief

Wang et al. summarize 16 methods for estimating cell potency from scRNA-seq and categorize them as average, entropy, and knowledge based within a unified mathematical framework. They show how assumptions, preprocessing, and protein-protein interaction priors affect estimates and highlight challenges in single-cell-resolution benchmarking and comparisons with pseudotime and RNA velocity methods.

Wang et al., 2026, Cell Systems *17*, 101762
December 16, 2026 © 2026 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.cels.2026.101762

CellPress

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Cell Systems

CellPress
OPEN ACCESS

Synthesis

Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data

Qingyang Wang,[1] Zhiqian Zhai,[1] Qiuyu Lian,[2] Dongyuan Song,[3] and Jingyi Jessica Li[1,3,4,5,6,7,8,9,*]
[1]Department of Statistics and Data Science, University of California, Los Angeles, Los Angeles, CA 90095-1554, USA
[2]Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, Cambridgeshire, UK
[3]Bioinformatics Interdepartmental PhD Program, University of California, Los Angeles, Los Angeles, CA 90095-7246, USA
[4]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095-7088, USA
[5]Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095-1766, USA
[6]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90095-1772, USA
[7]Radcliffe Institute for Advanced Study, Harvard University, Cambridge, MA 02138, USA
[8]Present address: Biostatistics Program, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA
[9]Lead contact
*Correspondence: lijy03@fredhutch.org
https://doi.org/10.1016/j.cels.2026.101762

SUMMARY

Inferring hidden cellular dynamics from static single-cell RNA sequencing (scRNA-seq) data remains a critical challenge. Beyond pseudotime trajectories and RNA velocities, cell potency provides another perspective by estimating cellular differentiation potential from gene expression levels. We review 16 computational methods that define cell potency under different, sometimes contrasting, assumptions. On the basis of how they summarize gene expression into a potency measure, we categorize them into average-based, entropy-based, and knowledge-based methods. Using a unified set of mathematical notations, we detail their methodologies, explain key similarities and differences within and between categories, and summarize usage complexities. Given the diversity of methods, community benchmark datasets at single-cell resolution and careful sensitivity analyses are crucial for evaluating methods from multiple angles and potentially improving stability and generalizability. This article provides a basis for method benchmarking and highlights the remaining challenge of comparing methods for inferring pseudotime trajectories, RNA velocities, and cell potency.

INTRODUCTION

The development of single-cell RNA sequencing (scRNA-seq) enabled researchers to profile gene expression at unprecedented single-cell resolution. However, scRNA-seq provides solely a static snapshot of individual cells, rather than the underlying dynamic process of cell differentiation (i.e., the process during which young, immature (unspecialized) cells take on individual characteristics and reach their mature (specialized) form and function[1]) or cell reprogramming (i.e., the conversion of cells from one type to another type[2]), if one exists. To infer the dynamic process from static scRNA-seq data, many computational methods have been developed, including the methods that infer cell pseudotime (also known as trajectory inference),[3] RNA velocity,[4] and cell potency—three concepts that carry related but distinct meanings. Specifically, pseudotime is a quantitative measure of progress through a biological process[3]; RNA velocity is the time derivative of the gene expression state[4]; cell potency, also known as stemness, is a quantitative measure of a cell's capacity to adopt different fates and its degree of commitment to particular lineages. Under this definition, potency primarily reflects fate diversity (i.e., the number of distinct cell types a cell can give rise to) and lineage commitment, rather than proliferative capacity per se, since the methods reviewed here are designed to rank cells along a differentiation hierarchy. Potency is not limited to normal differentiation; it also encompasses exogenous fate changes, such as cell reprogramming, and pathological fate changes, such as the acquisition of stem-like properties by cancer stem cells[5] or the abnormal dedifferentiation and transdifferentiation of cells in disease.[6,7] For a statistical mechanics perspective on pseudotime, RNA velocity, and cell potency, see Teschendorff and Feinberg.[8]

Pseudotime vs. RNA velocity vs. potency

In addition to the differences in the biological information that pseudotime, RNA velocity, and potency can reveal, the computational methods to obtain such information also differ in the following four aspects. First, while the inference of cell pseudotime and RNA velocity is suitable for many transcriptional dynamic processes such as immune responses and development,

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

cell potency estimation is specific to the cell differentiation process. Second, while pseudotime inference typically requires users to specify the starting point of a dynamic cellular process, RNA velocity inference and cell potency estimation do not have this requirement due to their inherent directional nature. RNA velocity indicates the direction of cell state change, and a higher potency value indicates a more stem-like cell. However, because of this difference, discrepancies may exist between pseudotime and the other two concepts. In Figures 1B–1D, pseudotime inference uses the specification that colonic stem cell (CSC) is the root cell type, while RNA velocities and cell potency indicate a direction from proliferative colonic stem cell (proCSC) to CSC, which differs from the established knowledge. Third, the methods for estimating pseudotime, RNA velocity, and cell potency have distinct inputs and outputs (Table 1). Pseudotime inference methods (e.g., Monocle[3] and Slingshot[9]) require users to input (1) cells' low-dimensional embeddings (e.g., embeddings by the principal component analysis [PCA], t-distributed stochastic neighbor embedding [t-SNE],[10] or uniform manifold approximation and projection [UMAP][11]) projected from a gene-by-cell expression matrix, (2) cells' cluster or type labels, and (3) the root cell of a trajectory (or multiple trajectories) on the basis of users' prior knowledge. The output of pseudotime inference methods is a cell-by-trajectory pseudotime matrix where each entry marks a cell's pseudotime in a trajectory. RNA velocity inference methods (e.g., velocyto[4] and scVelo[12]) require two inputted gene-by-cell count matrices (of the same dimensions), which correspond to the genes' spliced and unspliced read counts, respectively, in the cells. The output of RNA velocity inference methods is a cell-by-gene velocity matrix, in which every cell has a velocity vector with a length equal to the number of genes. Typically, cells' velocity vectors are visualized together with the cells in a two-dimensional t-SNE or UMAP plot. Cell potency estimation methods, the focus of this article, all require a gene-by-cell expression matrix, and some methods also need a protein-protein interaction (PPI) network and/or a gene-gene similarity matrix. The output of cell potency estimation methods is a cell potency vector, whose length is equal to the number of cells. Figures 1B–1D illustrates the different outputs of the three types of methods. In particular, pseudotime inference methods and cell potency estimation methods have the same output format when a single trajectory is specified for pseudotime inference; that is, each cell receives one pseudotime value and one potency value. However, it remains an open question whether any correspondence exists between cells' pseudotime values and potency values. Fourth, the three types of methods rely on different core algorithms. Most pseudotime inference methods use the minimum spanning tree algorithm to connect user-inputted cell clusters into a trajectory or trajectories. Most RNA velocity inference methods fit ordinary differential equations to estimate cells' velocity vectors. Unlike the first two types, computational methods that estimate cell potency employ diverse algorithms, ranging from summary statistic calculations to supervised learning algorithms.

Experimental and computational approaches to study cell potency

Knowing cell differentiation states helps researchers discover biological insights, such as the developmental relationships among various cell types,[14,15] the effects of a nanozyme-based eye drop on corneal epithelium regeneration,[16] and changes in cell differentiation potential as mice age.[17] Therefore, both biological and computational researchers have employed various ways to identify pluripotent stem cells (PSCs) and quantify the potency of individual cells.

Experimentally, the approaches[18–20] to evaluate cell potency include (1) measuring the expression levels of the transcription factors that are known to play a key role in helping cells maintain pluripotent, (2) inducing PSCs *in vitro* and checking the cells' ability to differentiate into three germ layers (*in vitro* differentiation assay), (3) injecting test cells into various host environments (e.g., immunocompromised mice for teratoma formation, blastocysts, tetraploid embryos, viable embryos, diploid or tetraploid blastocysts) and assessing the contribution of the injected cells to tissues from all three germ layers or the viability of the resulting chimera, and (4) conducting lineage-tracing experiments to track the progeny of individual cells over time. Earlier International Stem Cell Initiative guidelines suggest that *in vitro* differentiation and molecular assays can assess pluripotent differentiation potential, whereas teratoma assays can also reveal malignant potential.[21] More recent International Society of Stem Cell Research standards recommend evaluating pluripotency using multiple criteria, including morphology, lineage-specific mRNAs or proteins, cell-surface markers, functional properties, and downregulation of undifferentiated markers, and no longer require teratoma formation as a routine assay.[22]

While these experimental approaches provide important functional evidence for cell potency, they are typically applied at the cell line or population level rather than the single-cell level. Moreover, these assays are costly, time consuming, and difficult to scale. Therefore, leveraging scRNA-seq data, numerous computational methods have been developed to quantify individual cells' potency. Ideally, cell potency estimation methods should assign the highest potency to stem cells and the lowest potency to fully differentiated cells. Accurate estimation of cell potency should reveal the hidden differentiation status of individual cells, thus expanding the capacity of scRNA-seq data analysis. From the literature, we studied 16 cell potency estimation methods, including CytoTRACE,[23] ORIGINS,[24] NCG,[25] stemID,[26] SLICE,[27] dpath,[28] scEnergy,[29] cmEntropy,[30] SCENT,[31] Markov chain entropy (MCE),[32] SPIDE,[33] stemFinder,[34] mRNAsi,[35] CCAT,[36] FitDevo,[37] and CytoTRACE2.[38] On the basis of the assumptions and approaches of these methods, we divided them into three categories, including three *average-based methods* (CytoTRACE, ORIGINS, and NCG), nine *entropy-based methods* (stemID, SLICE, dpath, scEnergy, cmEntropy, SCENT, MCE, SPIDE, and stemFinder), and four *knowledge-based methods* (mRNAsi, CCAT, FitDevo, and CytoTRACE2). As an overview, average-based methods summarize a subset of genes' expression levels into an average-like statistic to represent a cell's potency; entropy-based methods define an entropy-like measure to describe the expression-level distribution of genes or gene groups within a cell or its surrounding neighborhood; and knowledge-based methods relate potency to how strongly a cell's transcriptome aligns with predefined stemness patterns. Since the single-cell field lacks a summary and comparison of cell potency methods, we are motivated to categorize the existing methods and help users understand them.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

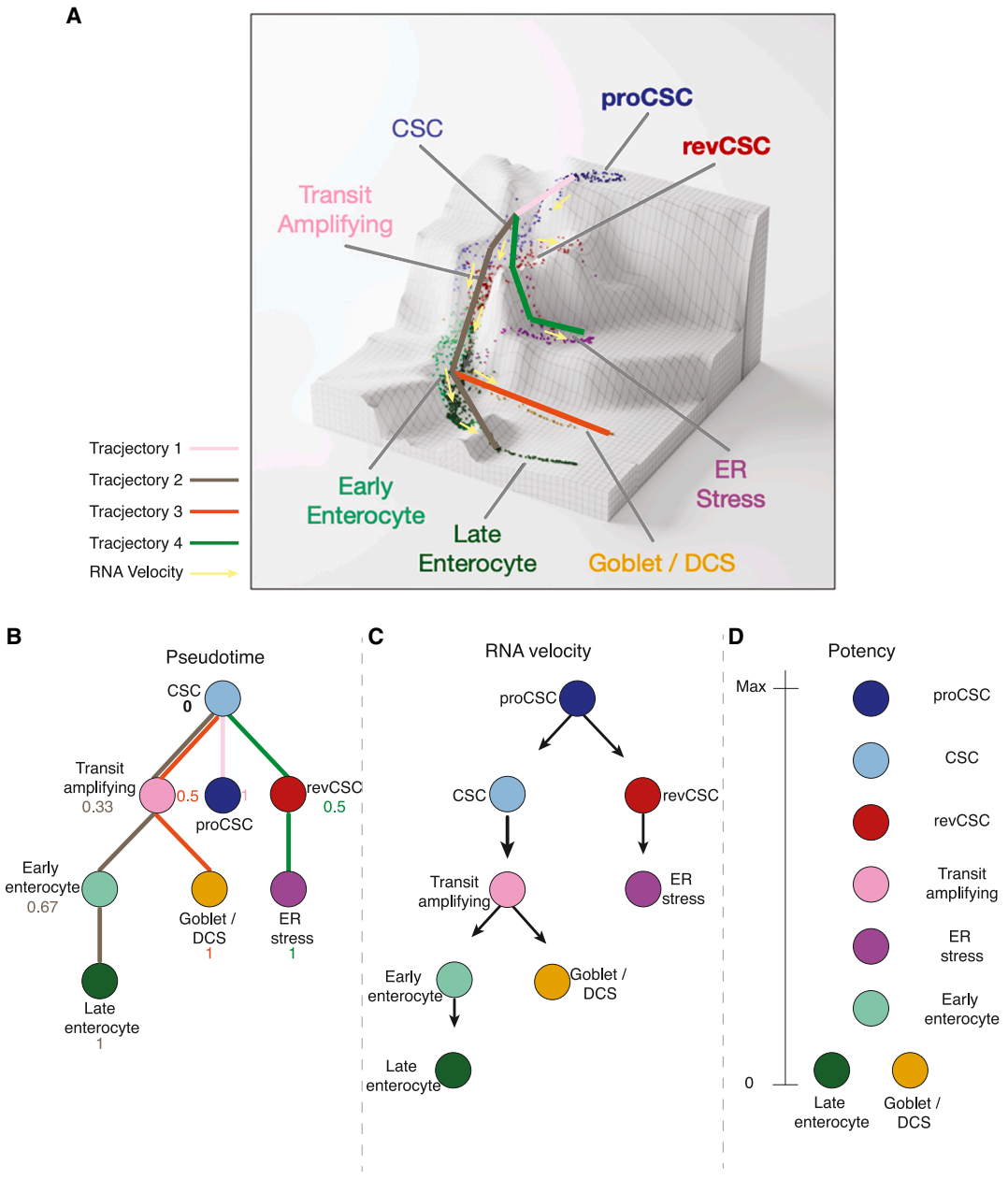


Figure 1. Different information provided by pseudotime, RNA velocity, and potency

(A) Combined illustration of pseudotime, RNA velocity, and cell potency. This figure is recreated on the basis of a part of Figure 7A from Qin et al.'s open access publication,[13] and this recreation is permitted under the Creative Commons CC-BY license. The dots with different colors represent cells from different cell types, which include colonic stem cells (CSCs), oncogene-driven LRIG1+ hyper-proliferative CSCs (proCSCs), fibroblast-induced clusterin (CLU)+ revival CSCs (revCSCs), transit amplifying cells (transit amplifying), cells under endoplasmic reticulum stress (ER stress), early enterocytes, late enterocytes, and goblet and deep crypt secretory cells (goblet/DCS). The brown, pink, green, and orange lines represent four differentiation trajectories, which can be inferred using pseudotime inference methods with CSC specified as the root cell type. The yellow arrows represent RNA velocity vectors, which can be inferred using RNA velocity inference methods. The altitude of this three-dimensional plot reflects cell potency, which can be inferred using cell potency estimation methods.

(B) Illustration of pseudotime inference results, containing lineages (represented by lines of different colors) and pseudotime values within each lineage (represented by numerical values from 0 to 1, with 0 indicating the root cell). Colored circles represent cell types.

(C) Illustration of RNA velocity inference results, with RNA velocities represented by arrows indicating direction and varying lengths indicating velocity magnitudes.

(D) Illustration of cell potency estimation results, where potency values are numeric and can be used to rank the cells or cell types. Note that there is a discrepancy between the pseudotime results in (B) and the RNA velocity and cell potency results in (C) and (D). In (B), CSC is specified as the root cell type based on domain knowledge, while in (C) and (D), proCSC is an ancestor of CSC in terms of RNA velocity and cell potency. According to Qin et al.'s publication,[13] proCSC is oncogene-driven and emerges subsequent to CSC. Therefore, the pseudotime results based on domain knowledge are considered more reasonable, suggesting that RNA velocity and cell potency results that do not incorporate domain knowledge might be problematic in some cases.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Table 1. Comparisons of the inputs and outputs for pseudotime inference, RNA velocity inference, and cell potency estimation methods

Category	Input	Output
Pseudotime inference	low-dimensional cell embeddings of the scRNA-seq data; cells' cluster or type labels; the root cell	cell-by-trajectory pseudotime matrix
RNA velocity inference	spliced and unspliced gene expression matrices	cell-by-gene velocity matrix
Potency estimation	gene-by-cell expression matrix; (optional: PPI network; gene-gene similarity matrix; cell cluster labels)	cell potency vector

This review

Here, we review 16 published computational methods mentioned above for estimating cell potency from scRNA-seq data using multiple perspectives, including the definition, resolution, and calculation of cell potency, as well as the input data and the external information such as Gene Ontology (GO) and a PPI network. We include only the methods that do not require users to have prior knowledge of the root cell, i.e., the starting point of cell differentiation, which is required for pseudotime inference. For this reason, we do not include Palantir,[39] a method that first infers pseudotime and then estimates cell potency from the inferred cell pseudotime values. Tables 2 and 3 summarize the 16 methods' characteristics and high-level intuitions, respectively.

During our detailed review of these methods, we spot differences between several methods' descriptions in their publications and their implementations examined in the corresponding software packages/scripts, whose version numbers or latest update dates when accessed (if the version number is unavailable) are listed in Table 2. We summarize these in Table S1 for user awareness. We also list the number of hyperparameters of the 16 methods in Table 4 to reflect each method's complexity. More specifically, Table S2 summarizes the hyperparameters tunable by users and the hyperparameters without explicitly evidenced defaults. In Figure 3, we use radar plots to compare the 16 methods' complexities in three dimensions: total number of hyperparameters, number of inputs, and number of differences (between publications and implementations we examined). Nevertheless, we have no empirical evidence linking these differences or defaults to performance, and benign explanations likely exist.

In the following sections, we will introduce mathematical notations to unify the notations used in the 16 methods, summarize these methods into the three categories: average-based methods, entropy-based methods, and knowledge-based methods (illustrated in Figure 2), and provide a detailed explanation for each method.

RESULTS

Unified mathematical notations

Before describing the 16 methods, we introduce the following mathematical notations to unify the notations used in different methods and facilitate our discussion.

- $X \in \mathbb{Z}_{\geq 0}^{p\times n}$: a gene-by-cell count matrix, assuming we have p genes and n cells. The (i,j)th entry X_{ij} indicates the read or unique molecular identifier (UMI) count, which represents the measured expression level, of gene i in cell j. The jth column $X_{\cdot j}$ is a p-dimensional column vector consisting of all p genes' counts in cell j.
- $S \in \mathbb{R}_{\geq 0}^{p\times n}$: a gene-by-cell normalized expression matrix, in which the cell library size is normalized out. The (i,j)th entry S_{ij} is the normalized expression level of gene i in cell j. Note that the purpose of normalization is to make gene expression levels comparable across cells, and different methods may use different normalization approaches.
- $Z \in \mathbb{R}_{\geq 0}^{p\times n}$: a gene-by-cell log-normalized expression matrix (i.e., applying log transformation on S). The (i,j)th entry Z_{ij} is the log-normalized expression level of gene i in cell j. Note that different methods may use different log-normalizing approaches. The jth column $Z_{\cdot j}$ is a p-dimensional column vector consisting of all p genes' log-normalized expression levels in cell j.
- $A \in \{0,1\}^{p\times p}$: a gene-by-gene binary adjacency matrix from a user-inputted PPI network. A is symmetric. For $i \neq i'$, the (i,i')th entry $A_{ii'} = 1$ means that an edge exists between genes i and i' in the PPI network, and $A_{ii'} = 0$ means otherwise. Note that $A_{ii} = 0$ in SCENT and SPIDE, but $A_{ii} = 1$ in MCE.
- $N(i)$: a set of gene i's neighboring genes in the PPI network (i.e., the genes that have edges connecting with gene i). The size of $N(i)$ is $\sum_{i'=1}^{p} A_{ii'}$.
- $\Pi \in [0,1]^{p\times n}$: a gene-by-cell steady-state probability matrix. The (i,j)th entry Π_{ij}, defined on the basis of X_{ij}, S_{ij}, or Z_{ij} (different methods may have different definitions), represents the steady-state probability of gene i in cell j. Intuitively, Π_{ij} serves as a weight for gene i in cell j. For SCENT and SPIDE, if we fix the total expression level of gene i's neighboring genes in cell j, i.e., $\sum_{i'\in N(i)} Z_{i'j}$, then Π_{ij} increases with Z_{ij}. For MCE, $\Pi_{ij} = S_{ij} := X_{ij}/\left(\sum_{i'=1}^{p} X_{i'j}\right)$ and thus increases with S_{ij}.
- $R_j \in [-1,1]^{p\times p}$: a gene-by-gene correlation matrix of cell j. The (i,i')th entry $R_{ii'j}$ is the Pearson correlation calculated on the basis of

$$\left\{\left(Z_{ij}, Z_{i'j}\right) : \text{cell } j' \text{ is among the } K \text{ nearest neighbors of cell } j\right\}.$$

  Specifically, each cell j's k-nearest neighbors are defined on the basis of the Pearson correlation applied to Z, i.e., the distance between cells j and j' is $1 - \text{cor}\left(Z_{\cdot j}, Z_{\cdot j'}\right)$.
- $P_j \in [0,1]^{p\times p}$: a gene-by-gene transition probability matrix of cell j. The (i,i')th entry $P_{ii'j}$ represents the transition probability from gene i to gene $i' \in N(i)$ in cell j. This transition probability is defined as an aggregation of the gene expression and PPI information (A), including Z_{ij},

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Table 2. Summary of cell potency estimation methods

Method	Potency measure	Input	Cell potency resolution	Implementation (version or update date when accessed)
CytoTRACE[23]	average[a]	count matrix[b]	single-cell level	R package CytoTRACE (v.0.3.3)
ORIGINS[24]	average[a]	count matrix[b]; PPI network	single-cell level	R package ORIGINS (v.0.1.0)
NCG[25]	average[a]	count matrix[b]; PPI network; gene-gene similarity matrix	single-cell level	R script (September 1, 2021)
stemID[26]	entropy	count matrix[b]; output from RaceID2	cell-cluster level	R script (March 3, 2017)
SLICE[27]	entropy	normalized expression matrix[c]; gene-gene similarity matrix	single-cell level	R package SLICE (v.0.99.0)
dpath[28]	entropy	TPM matrix	single-cell level	R package dpath (in Supplementary Software 1; v.2.0.1)
scEnergy[29]	entropy	normalized expression matrix[c]	single-cell level	MATLAB script (February 16, 2021)
cmEntropy[30]	entropy	count matrix[b]	single-cell level	R script (December 30, 2020)
SCENT[31]	entropy	log-normalized expression matrix[d]; PPI network	single-cell level	R package SCENT (v.1.0.3)
MCE[32]	entropy	count matrix[b]; PPI network	single-cell level	MATLAB script (in supplementary data; January 2020)
SPIDE[33]	entropy	count matrix[b]; PPI network	single-cell level	Python package SPIDE (June 22, 2022)
stemFinder[34]	entropy	log-normalized expression matrix[d]; cell-cycle genes	single-cell level	R package stemFinder (v.0.1.0)
mRNAsi[35]	knowledge	count matrix[b]	single-cell level	R script (December 2, 2021)
CCAT[36]	knowledge	log-normalized expression matrix[d]; PPI network	single-cell level	R package SCENT (v.1.0.3)
FitDevo[37]	knowledge	count matrix[b]	single-cell level	R script (April 6, 2024)
CytoTRACE2[38]	knowledge	count matrix[b]	single-cell level	R package CytoTRACE2 (v.1.1.0) and Python package CytoTRACE2 (v.1.1.0)

[a]Subject to weighting and smoothing.
[b]A gene-by-cell count matrix processed from scRNA-seq reads, denoted by X.
[c]A gene-by-cell normalized expression matrix, denoted by S.
[d]A gene-by-cell log-normalized expression matrix, denoted by Z.

$\{Z_{i'j} : i' \in N(i)\}$, as well as $R_{ii'j}$. Different methods may aggregate these sources of information in different ways.

- $W_j \in \mathbb{R}_{\geq 0}^{p \times p}$: a gene-by-gene edge weight matrix of cell j. The (i,i')th entry $W_{ii'j}$ represents the edge weight between gene i and gene $i' \in N(i)$. $W_{ii'j}$ is defined on the basis of Z_{ij}, $Z_{i'j}$ and $R_{ii'j}$, and different methods may have different definitions.
- E_j: the entropy for cell j calculated by an entropy-based method.
- NE_j: the normalized entropy for cell j (E_j is normalized by a constant C; the definition of C differs across entropy-based methods).

Category 1: Average-based methods

Referred to as the average-based methods, CytoTRACE,[23] ORIGINS,[24] and NCG[25] summarize the gene expression levels in a cell into an average-like summary statistics to represent the potency of the cell. Among the three methods, CytoTRACE requires the fewest user inputs: only the gene-by-cell count matrix. In short, CytoTRACE first selects potency-related genes and then assigns a higher potency to the cells exhibiting higher expression levels of the selected genes. In addition to the gene expression data, ORIGINS requires a PPI network. Briefly, ORIGINS defines the potency of a cell as the sum of the cell-specific PPI edge weights, where each edge weight is the product of the expression levels of the two genes connected by the edge. Besides the gene expression data and the PPI network, NCG further relies on gene-gene similarities pre-calculated on the basis of genes' GO terms. First, NCG assigns weights to genes on the basis of the PPI network and gene-gene similarities. Then, NCG defines the potency of a cell as the weighted sum of gene expression levels in the cell.

CytoTRACE

CytoTRACE[23] relies on the assumption that a cell's *gene count* (i.e., the number of expressed genes) reflects the cell's differentiation status. Accordingly, CytoTRACE first selects the genes whose expression levels highly correlate with the gene count across cells. Then CytoTRACE defines the *gene count signature* (*GCS*) as the geometric mean of the selected genes' expression levels in each cell. Finally, CytoTRACE performs smoothing on the cells' GCSs and then converts the smoothed GCSs into ranks, which serve as the final estimates of the cells' potency.

Technically, CytoTRACE first transforms $X \in \mathbb{Z}_{\geq 0}^{p \times n}$ to count per million (CPM) $S \in \mathbb{R}_{\geq 0}^{p \times n}$ ($S_{ij} = X_{ij} / \sum_{i'=1}^{p} X_{i'j} \times 10^6$) and then to $Z \in \mathbb{R}_{\geq 0}^{p \times n}$, with $Z_{ij} = \log_2\left(C_j \frac{S_{ij}}{\sum_{i'=1}^{p} S_{i'j}} + 1\right)$, where

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Table 3. Intuitions underlying the 16 cell potency estimation methods

Method	Potency measure	Intuition (what cells are more stem-like)
CytoTRACE	average	cells with more genes expressed
ORIGINS	average	cells where a larger number of highly expressed genes are connected in a PPI network
NCG	average	cells with higher expression levels of a specific subset of genes, which are closely connected in a PPI network and share similar GO terms
stemID	entropy	cell clusters that have more neighboring clusters and more cells with genes exhibiting similar expression levels
SLICE	entropy	cells where gene groups have more similar proportions of expressed genes
dpath	entropy	cells where metagenes exhibit more similar expression levels
scEnergy	entropy	cells where genes exhibit more similar expression levels
cmEntropy	entropy	cells where genes exhibit more similar expression levels
SCENT	entropy	cells whose influential genes (highly expressed genes that are densely connected with other highly expressed genes in a PPI network) have neighboring genes expressed at more similar levels
MCE	entropy	cells with more similar weighted gene-gene transition probabilities
SPIDE	entropy	cells whose influential genes (highly expressed genes that are highly correlated and densely connected with other highly expressed genes in a PPI network) have neighboring genes expressed at more similar levels
stemFinder	entropy	cells whose cell-cycle genes exhibit more heterogeneous expression levels in their surrounding neighborhood
mRNAsi	knowledge	cells that have a more consistent ranking between gene expression levels and pre-trained gene importance scores
CCAT	knowledge	cells that have a stronger positive correlation between genes' expression levels and their degrees in a PPI network
FitDevo	knowledge	cells whose gene expression levels have a stronger correlation with predefined gene importance scores
CytoTRACE2	knowledge	cells with more activated high-potency-specific gene sets

$C_j = \sum_{i'=1}^{p} \mathrm{I}(X_{i'j} > 0)$ is cell j's gene count. Second, CytoTRACE calculates the Pearson correlation between all n cells' gene counts $(C_1, C_2, \ldots, C_n)^T$ and gene i's log-normalized expression levels in the n cells $(Z_{i1}, Z_{i2}, \ldots, Z_{in})^T$, for each gene $i = 1, \ldots, p$ (The number of cells in the code implementation differs from what is described in the publication. Instead of n cells, it becomes $\tilde{n}$ selected cells, in which at least one of the $1,000$ genes selected in the substep (1) of the fourth step [the step that filters genes] is expressed; because CytoTRACE includes a cell-filtering step, some cells do not receive potency estimates; we note this behavior for user awareness, and the choice of filtering threshold for reliable estimation remains to be determined.). Third, CytoTRACE selects the top 200 genes that have the highest correlation values to compute the GCS for each cell, obtaining $G = (G_1, G_2, \ldots, G_n)^T \in \mathbb{R}^n_{\geq 0}$, where G_j is cell j's GCS, defined as the geometric mean of the 200 genes' log-normalized expression levels in cell j. In the implementation we examined, the GCS is computed as the arithmetic mean of the selected genes' log-normalized expression levels, whereas the original article describes it as a geometric mean (see STAR Methods for details). Fourth, CytoTRACE performs smoothing on G by taking the following sub-steps.

(1) Filter the rows of Z to retain only the genes that are expressed in at least 5% of the n cells. For each retained gene i, compute a *dispersion index* as the gene's variance over its mean across the n cells, i.e., $\frac{1}{n-1}\sum_{j=1}^{n}(Z_{ij} - \overline{Z}_i)^2/\overline{Z}_i$, where $\overline{Z}_i = \frac{1}{n}\sum_{j=1}^{n} Z_{ij}$. Further, filter the rows of Z to retain only the $1,000$ genes with the largest dispersion indices. Hence, the dimension of Z becomes $1,000 \times n$.

We note that this gene-selection substep is performed before the second step above in the code implementation, and the selected $1,000$ genes are used to select $\tilde{n}$ cells (see explanation of number of cells above), leading to a slightly different gene-selection procedure in the implementation compared with the description in the publication. Specifically, in the second step above, each Pearson correlation is calculated across all n cells in the publication but only the $\tilde{n}$ cells in the code implementation. Hence, the top 200 genes selected in the third step above differ between the code implementation and the publication. Also, the dimension of Z in the code becomes $1,000 \times \tilde{n}$. Thus, in the following sub-steps, the cell number n becomes $\tilde{n}$ in the code.

(2) Obtain a cell-by-cell Markov matrix $B \in [-1, 1]^{n \times n}$ in two steps.
   (a) Compute a cell-by-cell similarity matrix $D \in [-1, 1]^{n \times n}$, in which $D_{jj'}$ is the Pearson correlation between cell j and cell j' based on the retained $1,000$ genes' expression levels in Z, i.e., the Pearson correlation between the two columns $Z_{\cdot j}$ and $Z_{\cdot j'}$.
   (b) Transform D to B as

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Table 4. Hyperparameters of the 16 methods

Method	Hyperparameters	Total number of parameters
CytoTRACE	number of genes used to calculate GCS (200); number of genes with the largest dispersion indices (1,000); minimum percentage of cells in which a gene must be expressed to be included in the similarity matrix calculation (5%); α in the diffusion process (0.9)	4
ORIGINS	none	0
NCG	offset used in log transformation (1.1); cutoff value for s_{ij} (1)	2
stemID	significant level for identifying significant links (0.01)	1
SLICE	gene expression levels threshold for filtering genes (1); size of the subsample G_b (p_s); number of gene groups (M); number of subsamples (B)	4
dpath	number of metagenes (M); mean parameter for the Poisson distribution for the dropout (0.1); π_{ij} in the first round of NMF for $Z_{ij} = 0$ (0.1)	3
scEnergy	correlation threshold when constructing the gene network (0.1)	1
cmEntropy	none	0
SCENT	offsets in log transformation (1.1)	1
MCE	none	0
SPIDE	offset in log transformation (1.1); k in finding a cell's k-nearest neighbors (10 or 25)	2
stemFinder	k in finding a cell's k-nearest neighbors; threshold for binarization (τ)	2
mRNAsi	none	0
CCAT	none	0
FitDevo	number of PCs (L)	1
CytoTRACE2	four hyperparameters from CytoTRACE; number of gene sets (M); number of bins for selecting background genes (24); Gaussian mean for sampling background gene number (100); Gaussian variance ($100 \times \frac{n_{\text{bin}} - 100}{n_{\text{bin}}}$)	8

$$B_{jj'} = \begin{cases} 0 & \text{if } j = j' \text{ or } D_{jj'} < \tau \\ D_{jj'} & \text{otherwise} \end{cases},$$

where $\tau = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j'=1}^{n} D_{jj'}$ is the mean of D.

(3) Transform G to G_0 using non-negative least squares regression, which finds an n-dimensional vector $\beta_0 = \text{argmin}_{\beta \geq 0} \| G - B\beta \|_2^2$, and defining $G_0 = B\beta_0$.

(4) Simulate a diffusion process for $T = 10,000$ iterations or until convergence:

$$G_{t+1} = \alpha \cdot BG_t + (1 - \alpha) \cdot G_0; t = 0, \ldots, T - 1,$$

where the degree of diffusion $\alpha = 0.9$.

Finally, CytoTRACE estimates the *potency* by converting the n values in G_T from the last iteration into ranks n to 1 (with n indicating the maximum value in G_T).

The intuition behind CytoTRACE is that a cell has a higher potency if it has more genes expressed.

The hyperparameters in CytoTRACE include the two gene numbers (200 and 1,000) used to calculate G and D, respectively, the percentage threshold (5%) used to filter out genes for the cell-cell similarity matrix calculation, as well as the degree of diffusion α in the diffusion process. In total, we count CytoTRACE to have four hyperparameters.

CytoTRACE is a widely used tool with diverse applications in investigating disease mechanisms. In cancer research, it has helped identify the potential origins of *LAMP3* + conventional dendritic cells across various cancer types.[40] In the context of Crohn's disease, it has been used to infer the differentiation trajectory from stem cells to LND cells, a distinct epithelial cell type characterized by high expression of LCN2, NOS2, DUOX2, and genes related to antimicrobial response and immunoregulation.[41] Additionally, in immuno-oncology, CytoTRACE has been applied to analyze scRNA-seq data from cancer patients treated with immune checkpoint inhibitors, revealing negative correlations between CytoTRACE-estimated cell potency and anti-tumor immunity.[42]

ORIGINS

ORIGINS[24] defines cell potency from a gene-by-cell count matrix $Z \in \mathbb{R}_{\geq 0}^{p \times n}$(preprocessed externally, as the package does not have a built-in normalization step) and a PPI network $A \in \{0,1\}^{p \times p}$. ORIGINS calculates a weighted adjacency matrix $W_j \in \mathbb{R}_{\geq 0}^{p \times p}$ for each cell j, with the (i, i')th entry set as $W_{ii'j} = A_{ii'}Z_{ij}Z_{i'j}$. Then, ORIGINS sums over the off-diagonal entries of W_j to obtain $L_j = \sum_{i=1}^{p}\sum_{i' \neq i} W_{ii'j}$, which is assumed to represent cell j's activity level associated with the differentiation process. ORIGINS scales L_j to obtain the *potency* $L_j^* \in [0,1]$ for cell j as

$$L_j^* = \frac{L_j - \min_{j'' \in \{1,\cdots,n\}} L_{j''}}{\max_{j' \in \{1,\cdots,n\}} L_{j'} - \min_{j'' \in \{1,\cdots,n\}} L_{j''}}.$$

The intuition behind ORIGINS is that a cell has a higher potency if the genes connected in the PPI network are more highly expressed.

We count ORIGINS as having zero hyperparameters.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

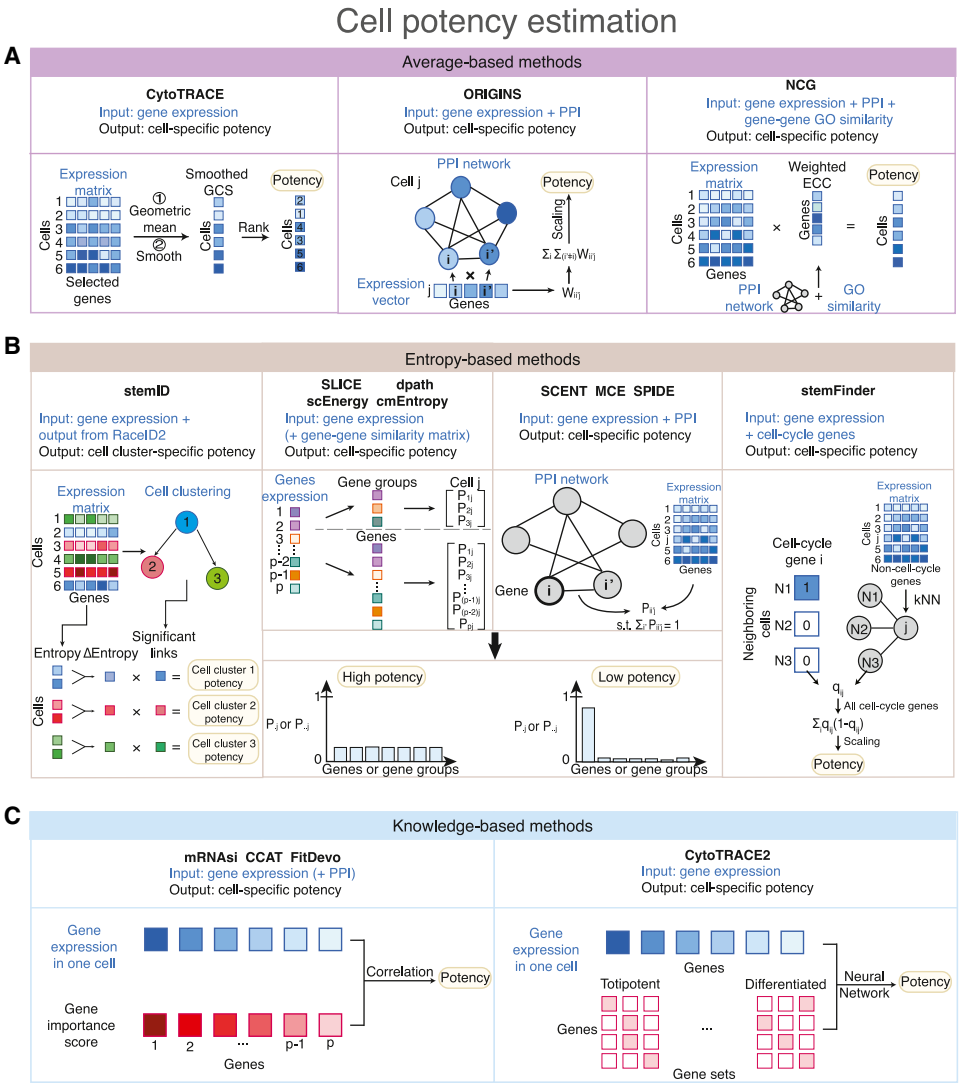


Figure 2. A conceptual categorization of the 16 methods into 3 categories: Average-based methods, entropy-based methods, and knowledge-based methods

All methods except stemID calculate the potency at the single-cell resolution, while stemID offers the cell-cluster resolution only.

(A) Average-based methods. The CytoTRACE method (left) computes a gene count signature (GCS) for each cell using the selected genes only. Then CytoTRACE defines cells' potency as the ranks of the cells' smoothed GCS values. The ORIGINS method (middle) assigns every edge in the PPI network a cell-specific weight, calculated as the product of the cell-specific expression levels of the two genes connected by the edge. ORIGINS defines a cell's potency as a scaled sum of these cell-specific edge weights. The NCG method (right) defines a weight for each gene by aggregating the gene's similarities to other genes. These gene-gene similarities are based on the genes' edge weights in a PPI network and their GO terms. NCG then uses the defined gene weights and a gene-by-cell log-normalized expression matrix to estimate each cell's potency.

(B) Entropy-based methods. These nine methods use entropy to define the potency and can be divided into four sub-categories. In the first sub-category, stemID defines the potency of each cell cluster, not each cell, as the product of the cell cluster's relative entropy (ΔEntropy) and links (similarities) to other cell clusters. In the second sub-category, SLICE, dpath, scEnergy, and cmEntropy first calculate a proportion vector $P_{\cdot j}$ for each cell j, where P_{ij} is the proportion (i.e., the relative expression level) of gene group i (SLICE and dpath) or gene i (scEnergy and cmEntropy) in cell j, and then define cell j's potency as the entropy of $P_{\cdot j}$.

Among the four methods in the second sub-category, SLICE requires users to input a gene-gene similarity matrix (the Kappa statistics for all gene pairs) while the other three methods do not. In the third sub-category, SCENT, MCE, and SPIDE define cell j's potency as the entropy of a transition probability matrix (or extended vector) $P_{\cdot\cdot j}$, where $P_{ii'j}$ is the transition probability (i.e., the normalized edge weight) from gene i to gene i' in cell j. In the fourth sub-category, stemFinder uses non-cell-cycle genes to identify the k-nearest neighbors of each cell j and defines its potency on the basis of the summed Gini impurity of binarized cell-cycle gene expression within the local neighborhood. Here, q_{ij} is the proportion of cell j's k neighbors that share the same binarized state as cell j for gene i.

(C) Knowledge-based methods. The mRNAsi, CCAT, and FitDevo methods define a cell's potency as the correlation between the cell's gene expression vector and a vector of predefined, non-cell-specific gene important scores. In mRNAsi, genes' importance scores are the genes' coefficients in a pre-trained OCLR model for predicting the probability that a given cell is a stem cell. In CCAT, genes' importance scores are the genes' degrees in a PPI network. In FitDevo, a gene's importance score is the sum of a pre-trained gene weight and a sample-specific gene weight. CCAT requires users to input a PPI network while mRNAsi and FitDevo do not. In contrast, CytoTRACE2 uses predefined gene sets, rather than individual genes, with a neural network to predict each cell's potency category and compute the corresponding potency score.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

NCG

NCG[25] defines cell potency from scRNA-seq data, a PPI network, and GO annotations(note that the required inputs of NCG include a count matrix, a PPI network, and a gene-gene similarity matrix defined on the basis of GO annotations). First, NCG performs quantile normalization on the inputted count matrix $X = (X_{ij})$ to obtain $S = (S_{ij})$. Then, NCG log transforms S to obtain $Z = (Z_{ij})$. In the publication, if $S_{ij} < 1$, $Z_{ij} = \log_2(S_{ij} + 1.1)$; otherwise, $Z_{ij} = \log_2(S_{ij} + 0.1)$ (note that in NCG's code implementation, Z_{ij} is always $\log_2(S_{ij} + 0.1)$; if $S_{ij} < 1$, then S_{ij} is set to 1, and Z_{ij} becomes a constant $\log_2 1.1$). Next, NCG defines a *weighted edge clustering coefficient of the gene* (*ECG*) for each gene in both the scRNA-seq data and the PPI network. The ECG for gene i is $ECG_i = \sum_{i'=1}^{p} w_{ii'} \times ECC_{ii'}$, in which p is the number of genes in both the scRNA-seq data and the PPI network, $w_{ii'}$ is the similarity between gene i and gene i' defined as the Kappa statistic applied to the two genes' GO annotations (detailed below), and $ECC_{ii'}$ is the *edge clustering coefficient* (*ECC*) (defined as $|N(i) \cap N(i')|/(\min(N(i), N(i')) - 1)$, the number of common neighbors between gene i and gene i' divide by the minimum degree of the two genes minus 1; it is set to 0 when the denominator is 0). Note that the ECC measures the degree of closeness of two genes in a PPI network. Intuitively, if a gene has a higher ECG value, it is similar to more genes based on the PPI network and GO annotations.

Next, NCG computes the *potency* for cell j as

$$NCG_j = \frac{1}{\max_{j' \in \{1,\cdots,n\}} NCG_{j'}} \sum_{i=1}^{p} Z_{ij} \times ECG_i.$$

A high value of NCG_j indicates a high potency value of cell j.

The Kappa statistic for each gene pair is a required input that users must specify. Since the NCG paper[25] does not provide the full formula for the Kappa statistic, we describe it here.[43,44] The Kappa statistic $w_{ii'}$ between gene i and gene i' is defined as follows. Given N GO terms in total, the following confusion matrix can be constructed from the two genes' GO annotations.

	Gene i'		
Gene i	1	0	Row total
1	$N_{ii'}^{11}$	$N_{ii'}^{10}$	$N_{i\cdot}^{1*}$
0	$N_{ii'}^{01}$	$N_{ii'}^{00}$	$N_{i\cdot}^{0*}$
Column total	$N_{\cdot i'}^{*1}$	$N_{\cdot i'}^{*0}$	N

In the confusion matrix, the N GO terms are partitioned by whether they are associated with gene i and/or gene i', with the row or column label of 1 indicating belonging and 0 otherwise. Then, for example, $N_{ii'}^{11}$ indicates the number of GO terms shared by genes i and i', and $N_{ii'}^{10}$ means the number of GO terms belonging to gene i but not to gene i'. The Kappa statistic is then defined as $w_{ii'} = \frac{O_{ii'} - E_{ii'}}{1 - E_{ii'}}$ where $O_{ii'} = \frac{N_{ii'}^{11} + N_{ii'}^{00}}{N}$ and $E_{ii'} = \frac{N_{i\cdot}^{1*} N_{\cdot i'}^{*1} + N_{i\cdot}^{0*} N_{\cdot i'}^{*0}}{N^2}$.

The intuition behind NCG is that, with a subset of genes identified as closely connected in the PPI network and sharing similar GO terms, a cell has a higher potency if the identified genes have higher expression in it.

The hyperparameters in NCG include the offset 1.1 used in the log transformation ($Z_{ij} = \log_2(S_{ij} + \text{offset})$) and the cutoff value of 1 for S_{ij} to decide which offset value (1.1 or 0.1) to use. We do not count the other offset 0.1 used for $S_{ij} \geq 1$ as a hyperparameter because we interpret this offset in the implementation as ensuring continuity at $S_{ij} = 1$; the published formula uses a slightly different expression, so the two are not exactly identical. In total, we count NCG to have two hyperparameters.

Category 2: Entropy-based methods

Out of the 16 methods, 9 methods (stemID,[26] SLICE,[27] dpath,[28] scEnergy,[29] cmEntropy,[30] SCENT,[31] MCE,[32] SPIDE,[33] and stemFinder[34]) use various ways to define and calculate the entropy or related measures of cellular heterogeneity, which are then used as potency scores for individual cells or cell clusters. These methods are based on the rationale that stem cells should have higher entropy-like values (i.e., genes have similar expression levels in stem cells), while mature cells should have lower entropy-like values (i.e., genes exhibit different expression levels in mature cells.).

On the basis of the methodological details, we further separate the nine methods into four sub-categories, depending on the resolution of cell potency (at the cell-cluster or single-cell level), the use of a PPI network, and the metric used to quantify potency. The first sub-category includes stemID, the only method that calculates potency at the cell-cluster level. stemID first finds cell clusters and then uses clusters' similarities and relative gene expression levels to calculate entropy for each cluster. The second sub-category includes SLICE, dpath, scEnergy, and cmEntropy, which use the relative or binarized expression levels of either gene groups (SLICE and dpath) or individual genes (scEnergy and cmEntropy) to define the potency of each cell. The third sub-category requires a PPI network and includes SCENT, MCE, and SPIDE, which use different ways to define genes' steady-state probabilities and gene-to-gene transition probabilities based on the PPI network and gene expression levels. Then, they aggregate these probabilities into an entropy-like measure for each cell. Finally, they normalize the measure as a potency estimate for each cell. The fourth sub-category includes stemFinder, which uses Gini impurity rather than entropy to define potency.

stemID

Among the nine entropy-based methods, stemID[26] is the only one that estimates the potency of cell clusters instead of cells. Hence, stemID requires users to run the clustering method RaceID2,[26] which was developed by the same authors. The RaceID2 result contains each cell's cluster label and each cluster's medoid.

First, stemID links all pairs of cluster medoids and determines the significance of each link as follows: every cell is projected to the closest link (by vertical distance in the embedding space by multi-dimensional scaling, which preserves the pairwise cell distances; the distance between cells j and j' is defined as one minus the Pearson correlation of the two cells' count vectors $X_{\cdot j}$ and $X_{\cdot j'}$, and the dimension of the multi-dimensional scaling embedding space is the number of positive eigenvalues of the cell distance matrix) among the links that

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

connect the cell's cluster medoid to other cluster medoids; after all cells are projected, each link has its projected cells, whose number would then be converted to a p value (considered significant if under a default significance level of 0.01). Given the significant links, stemID computes the number of significant links involving each cluster k's medoid and denotes the number as l_k.

Next, stemID computes the *entropy value* of cell j as $E_j = -\sum_{i=1}^{p} S_{ij} \log_p S_{ij}$ (here we include the negative sign, which was omitted in the original article), where $S_{ij} := X_{ij} / \sum_{i'=1}^{p} X_{i'j}$ is the proportion of gene i's expression level in cell j.

Finally, stemID computes a *stemness score* s_k as the estimated *potency* for each cell cluster k as

$$s_k = l_k \times \Delta E_k,$$

where ΔE_k is the difference between cluster k's median cell entropy value (i.e., the median of the entropy values of the cells in cluster k) and the minimum of all clusters' median cell entropy values.

The intuition behind stemID is that a cluster has a higher potency if it has more clusters in proximity and contains more cells with genes exhibiting similar expression levels.

The hyperparameter in stemID is the significance level of 0.01 used for identifying significant links. In total, we count stemID to have one hyperparameter.

Unlike stemID, the other eight entropy-based methods estimate the potency of each cell.

SLICE

SLICE[27] defines the potency measure, called scEntropy, for each cell on the basis of the inputted gene expression matrix (which can be Seurat's normalized data or be on the scale of fragments per kilobase of transcript per million mapped reads [FPKM], reads per kilobase per million mapped reads [RPKM], or transcripts per million [TPM]) and GO annotations. Specifically, SLICE calculates a cell-specific entropy based on gene groups instead of genes, unlike stemID; SLICE clusters genes into groups by performing k-means clustering on a gene-gene dissimilarity matrix, in which the gene-gene dissimilarities are measured by one minus the Kappa statistic applied to the genes' GO terms (same as the Kappa statistic used in NCG), and this Kappa statistic matrix is a required user input. SLICE's R package provides the similarity matrices measured by Kappa statistics for human genes and mouse genes, respectively.

Note that SLICE applies k-means clustering directly to a symmetric gene-gene dissimilarity matrix, whereas k-means is more commonly applied to an instance-by-feature matrix (genes by features).

In detail, for each cell j, SLICE computes the scEntropy, which is defined as the *potency* for cell j, as

$$E_j = \frac{1}{B} \sum_{b=1}^{B} \sum_{m=1}^{M} -p_j^m(G_b, F_b) \log p_j^m(G_b, F_b),$$

where G_b is a random subsample of $p_s = 1{,}000$ genes with expression levels greater than 1 in at least one cell (in the original article, G_b is referred to as a "bootstrap sample," although technically it is a random subsample of genes), F_b denotes the $M = \sqrt{p_s/2} = 22$ gene groups clustered from the genes in G_b, and $p_j^m(G_b, F_b) = \frac{c_j^m(G_b, F_b) + \alpha^m}{C_j(G_b, F_b) + A}$, in which $c_j^m(G_b, F_b)$ denotes the number of genes in G_b expressed in cell j and belonging to the mth group in F_b, α^m denotes the proportion of genes in G_b belonging to the mth group in F_b (i.e. the number of genes in G_b belonging to the mth group in F_b divided by p_s), $C_j(G_b, F_b) = \sum_{m=1}^{M} c_j^m(G_b, F_b)$, and $A = \sum_{m=1}^{M} \alpha^m$. Hence, $p_j^m(G_b, F_b)$ measures the relative degree of activeness (i.e., proportion) of the mth gene group in F_b in cell j, and E_j denotes cell j's average entropy calculated on the proportions of M gene groups across $B = 100$ subsamples. Here, the values of M and B are from SLICE's tutorials.[45]

Note that SLICE treats gene expression counts as binary because it only considers if a gene is expressed or not in a cell in its scEntropy calculation.

The intuition behind SLICE is that a cell has a higher potency if the M gene groups have similar proportions of expressed genes. The M gene groups are defined separately within each gene subsample, so the group assignment can vary across subsamples.

The hyperparameters in SLICE include the expression level threshold of 1 used to filter out genes, the number of genes in each subsample (p_s), the number of gene groups (M), and the number of subsamples (B). In total, we count SLICE to have four hyperparameters.

dpath

dpath[28] is similar to SLICE but uses a different way to partition genes into groups: dpath first log transforms the inputted gene-by-cell TPM matrix to obtain Z ($Z = \log(\text{TPM} + 1)$). Then, dpath performs a weighted non-negative factorization (NMF) on Z to cluster genes into M (default 5) groups (called "metagenes"). The factorization is in the form of $Z \approx UV$, where U is a gene-by-metagene non-negative matrix, and V is a metagene-by-cell non-negative matrix satisfying $\sum_{m=1}^{M} V_{mj} = 1$ for all cell j (i.e., every column of V sums up to 1, so V_{mj} indicates the proportion of metagene m in cell j). The dpath authors denoted $\mu = (\mu_{ij}) = UV$ as the expected gene-by-cell logTPM matrix. Given the observed gene-by-cell logTPM matrix Z, the dpath authors wrote the log likelihood of U, V, and π (a gene-by-cell weight matrix as a parameter to be estimated) as

$$\ell(U, V, \pi) = \sum_{i=1}^{p} \sum_{j=1}^{n} \left[\pi_{ij} \cdot \text{logPoisson}(Z_{ij} | \mu_{ij}) + (1 - \pi_{ij}) \cdot \text{logPoisson}(Z_{ij} | 0.1)\right],$$

where $\text{Poisson}(\cdot | \lambda)$ represents the probability mass function of the Poisson distribution with mean λ; π_{ij} is the weight for gene i in cell j; μ_{ij} is the expected logTPM of gene i in cell j; and 0.1 is the mean parameter of the Poisson distribution modeling the dropout events. However, the Poisson distribution is used for count data, so applying it directly to logTPM values, which are no longer counts, represents a non-standard modeling choice.

Then dpath estimates U, V, and π through an iterative optimization process to maximize the log likelihood $\ell(U, V, \pi)$, and this process is referred to as the "weighted NMF." In detail, dpath performs two rounds of iterations in the optimization process. First, dpath iteratively solves U and V by fixing π across all iterations: $\pi_{ij} = 1$ if $Z_{ij} > 0$, and $\pi_{ij} = 0.1$ otherwise (in the implementation we used, π_{ij} is initialized at 0 rather than 0.1 for $Z_{ij} = 0$, which differs slightly from the default value reported in the dpath paper); hence, non-zero Z_{ij}'s have larger weights than zero Z_{ij}'s.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Second, dpath uses the U and V found by the first round of iterations as the initialization and iteratively solves U, V and π; in the tth iteration of the second round, π_{ij} is updated as $\pi_{ij}^t = \frac{\text{Poisson}\left(Z_{ij}\middle|\mu_{ij}^t\right)}{\text{Poisson}\left(Z_{ij}\middle|\mu_{ij}^t\right)+\text{Poisson}\left(Z_{ij}\middle|0.1\right)}$, where $\mu_{ij}^t = \sum_{m=1}^{M} U_{im}^t V_{mj}^t$.

Using V outputted by the second round, dpath estimates the *potency* for cell j as

$$E_j = -\sum_{m=1}^{M} V_{mj} \log V_{mj}.$$

The intuition behind dpath is that a cell has a higher potency if its metagenes have similar expression levels.

The hyperparameters in dpath include the number of metagenes (M), the mean parameter for the Poisson distribution modeling the dropout event (default hyperparameter value 0.1), and the value of $\pi_{ij} = 0.1$ used in the first round of the weighted NMF for $Z_{ij} = 0$. In total, we count dpath to have three hyperparameters.

scEnergy

scEnergy[29] is different from SLICE and dpath because it does not calculate the cell-specific entropy based on gene groups. Instead, scEnergy considers every gene and the neighboring genes in a gene network constructed from the inputted gene-by-cell expression matrix (which can be on the scale of TPM, FPKM, or UMI count). Specifically, scEnergy first constructs a gene network by connecting every two genes that have absolute Spearman correlations greater than a threshold (default 0.1); the Spearman correlation is defined between two genes' log-transformed counts, i.e., between $\left(S_{i1}^*,\ldots,S_{in}^*\right)^T$ for gene i and $\left(S_{i'1}^*,\ldots,S_{i'n}^*\right)^T$ for gene i', where $S_{ij}^* = \log_2\left(X_{ij}+1\right)$. Then for cell j, scEnergy defines the *energy* as

$$E_j = -\sum_{i=1}^{p} Z_{ij}^* \log \frac{Z_{ij}^*}{\sum_{i'\in N(i)} Z_{i'j}^*},$$

where $Z_{ij}^* = \frac{S_{ij}^* - \min_{i''\in\{1,\ldots,p\}} S_{i''j}^*}{\max_{i'\in\{1,\ldots,p\}} S_{i'j}^* - \min_{i''\in\{1,\ldots,p\}} S_{i''j}^*}$.

Finally, scEnergy estimates the *potency* for cell j by rescaling the energy E_j as

$$NE_j = \frac{\left(E_j/\overline{E}\right)^2}{1+\left(E_j/\overline{E}\right)^2},$$

where $\overline{E}$ is the average energy of all cells, and NE_j is outputted as the estimated *potency* of cell j. While the intuition behind the E_j definition was described as being based on maximum entropy and statistical thermodynamics, the precise connection to these principles, and the rationale behind the rescaling were not explicit.[29]

The rough intuition behind scEnergy is that a cell has a higher potency if genes have similar expression levels in it.

The hyperparameter in scEnergy is the threshold of 0.1 for the absolute Spearman correlations of gene pairs when constructing the gene network. In total, we count scEnergy to have one hyperparameter.

cmEntropy

cmEntropy[30] was proposed to quantify the cardiomyocyte maturation status at the single-cell level using the Shannon entropy. The cmEntropy authors also found cmEntropy applicable to other tissues for quantifying cell potency. Given the inputted gene-by-cell count matrix X, cmEntropy defines the *potency* for cell j as

$$E_j = -\sum_{i=1}^{p} S_{ij} \log S_{ij},$$

where $S_{ij} = X_{ij}/\left(\sum_{i'=1}^{p} X_{i'j}\right)$.

The intuition behind cmEntropy is that a cell has higher potency if its genes have more similar expression levels.

We count cmEntropy as having zero hyperparameters.

SCENT

SCENT[31] differs from stemID, SLICE, dpath, and scEnergy in the probabilities used for entropy calculation. Using the signaling entropy definition in the SCENT authors' previous work,[46,47] SCENT takes as inputs (1) a log-normalized gene-by-cell expression matrix Z, with the log-normalized expression of gene i in cell j as $Z_{ij} = \log_2\left(S_{ij}+1.1\right)$, where S_{ij} is X_{ij} (the count of gene i in cell j) multiplied by cell j's specific scaling factor calculated externally and (2) a PPI network that specifies the gene-by-gene binary adjacency matrix $A \in \{0,1\}^{p\times p}$. Then for each cell j, SCENT calculates the *signaling entropy* as

$$E_j = -\sum_{i=1}^{p} \Pi_{ij} \sum_{i'\in N(i)} P_{ii'j} \log P_{ii'j}.$$

The E_j formula has two key components. First, the steady-state probability of gene i in cell j is

$$\Pi_{ij} = \frac{Z_{ij}\left(A\cdot Z_{\cdot j}\right)_i}{Z_{\cdot j}^{\intercal}\cdot\left(A\cdot Z_{\cdot j}\right)},$$

where $Z_{\cdot j} \in \mathbb{R}_{\geq 0}^{p}$ is the log-normalized gene expression vector of cell j, and $\left(A\cdot Z_{\cdot j}\right)_i$ is the ith element of the vector $\left(A\cdot Z_{\cdot j}\right) \in \mathbb{R}_{\geq 0}^{p}$ and is equivalent to $\sum_{i'\in N(i)} Z_{i'j}$. Second, the transition probability from gene i to gene i' in cell j is

$$P_{ii'j} = \frac{W_{ii'j}}{\sum_{i''\in N(i)} W_{ii''j}},$$

where $W_{ii'j} = Z_{ij}Z_{i'j}$, so $P_{ii'j} = \frac{Z_{i'j}}{\sum_{i''\in N(i)} Z_{i''j}}$, is irrelevant to Z_{ij} and only uses gene i's neighbors in the PPI network.

SCENT estimates the *potency* for each cell j as

$$NE_j = E_j/C,$$

in which the normalization constant C is the maximum signaling entropy defined as

$$C = \max_{P=(P_{ii'})}\left(-\sum_{i=1}^{p}\Pi_{ij}\sum_{i'\in N(i)} P_{ii'}\log P_{ii'}\right) =$$

$$-\sum_{i=1}^{p}\Pi_{ij}\sum_{i'\in N(i)}\left(\frac{A_{ii'}\nu_{i'}}{\lambda\nu_i}\right)\log\left(\frac{A_{ii'}\nu_{i'}}{\lambda\nu_i}\right) = \log\lambda \text{ for any cell } j,$$

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

where λ is the largest eigenvalue of the adjacency matrix A, and ν_i is the ith element in the eigenvector ν corresponding to the largest eigenvalue. The last equality $C = \log \lambda$ was proven in Demetrius and Manke.[48]

As the number of cells increases, SCENT becomes computationally intensive. Therefore, CCAT,[36] a faster proxy of SCENT, was proposed. Since CCAT's approach does not involve entropy calculation, it belongs to another category we define as "knowledge-based methods," and we will discuss CCAT in the next section. Later, the same research group proposed LandSCENT,[49] a direct extension of SCENT. However, since the step in LandSCENT that computes cell potency is identical to that of SCENT, we have not included LandSCENT in this review.

The intuition behind SCENT is that a cell j has a higher potency if its "influential" genes (i.e., gene i is influential in cell j if it has a large Π_{ij}, indicating gene i is highly expressed in cell j or has neighboring genes highly expressed in cell j) have neighboring genes expressed at similar levels.

The hyperparameter in SCENT is the offset value of 1.1 used in log transformation $Z_{ij} = \log_2(S_{ij} + \text{offset})$. In total, we count SCENT to have one hyperparameter.

The last two entropy-based methods, MCE[32] and SPIDE,[33] are similar to SCENT in that they all use genes' expression levels and genes' relationships in a PPI network to compute an entropy for each cell.

MCE

MCE[32] defines the *entropy* for each cell j as

$$E_j = \max_{P_j = (P_{ii'j}) \text{ s.t. } P_j \Pi_{\cdot j} = \Pi_{\cdot j}} \left(- \sum_{(i,i') \in \hat{G}} \Pi_{ij} P_{ii'j} \log(\Pi_{ij} P_{ii'j}) \right),$$

where (i, i') indicates an edge between genes i and i' in a PPI network, and $\hat{G}$ is a set including all edges in the PPI network, including all self-loop edges. In the above definition of E_j, $\Pi_{ij} = X_{ij} / (\sum_{i'=1}^{p} X_{i'j})$ is the steady-state probability of gene i in cell j, and $\Pi_{\cdot j}$ is cell j's steady-state probability vector; $P_{ii'j}$ is the (i, i')th element in cell j's transition matrix P_j that satisfies $P_j^T \Pi_{\cdot j} = \Pi_{\cdot j}$ and $P_j 1 = 1$. The computation of E_j is solved as a convex optimization problem.

MCE's final estimate of cell j's *potency* is $NE_j = E_j / C$, in which the normalization constant $C = \log(\sum_{i=1}^{p} d_i)$, with d_i representing the degree of gene i in the PPI network.

Two major differences between MCE and SCENT are as follows. First, the inputted PPI network for MCE can have all ones on the diagonal of the adjacency matrix, while for SCENT the diagonal must have all zeros. Second, the transition probability matrix P_j of cell j is optimized given the steady-state probability Π_j for MCE, while P_j is computed on the basis of the log-normalized gene expression levels for SCENT.

The intuition behind MCE is that a cell has a higher potency if its weighted gene-to-gene transition probabilities are similar.

We count MCE as having zero hyperparameters.

SPIDE

Similar to SCENT and MCE, SPIDE[33] also defines an entropy for each cell using both gene expression levels and a PPI network. SPIDE uses a more elaborate procedure to process gene expression levels. Because the published notation for SPIDE is internally inconsistent and cannot be used to implement the method as written, we base our description on a reconciliation of the publication with the accompanying code.

Specifically, SPIDE first performs imputation to reduce data sparsity. In the implementation we examined, for each cell j, SPIDE finds its k-nearest neighbors ($k = 10$ for datasets with $1,000$ cells or fewer; $k = 25$ for datasets with more than $1,000$ cells) as those cells with the highest Pearson correlations with cell j based on the $\log_2(\text{count} + 1.1)$ transformed expression levels (the count of gene i in cell j is X_{ij}). However, as illustrated in Figure 1 of SPIDE's publication, SPIDE finds the k-nearest neighbors on the basis of the cell-cell correlations calculated using log-transformed and quantile-normalized data, which is differ slightly from the implementation we examined. For each gene, if its raw count in cell j is zero, then SPIDE computes two averages of the gene's raw count—one across cell j's k-nearest neighbors and the other across all cells. If the first average is no less than the second (in the publication; in the code, "no less than" becomes "greater than"), SPIDE replaces the gene's zero expression level in cell j with the first average. Then, SPIDE further processes the imputed data by quantile normalization and log transformation.

With the imputed and log-normalized gene expression levels and the PPI network, SPIDE defines the *entropy* for cell j as

$$E_j = - \sum_{i=1}^{p} \Pi_{ij} \sum_{i' \in N(i)} P_{ii'j} \log P_{ii'j},$$

which is exactly the same formula as SCENT's signaling entropy.

Despite the same entropy formula, there are three differences between SPIDE and SCENT. First, the two methods use different definitions of Z_{ij}, the expression level of gene i in cell j, which is used to define $P_{ii'j}$, the transition probability from gene i to gene i' in cell j. Recall that SCENT defines Z_{ij} as a log-normalized expression level, while SPIDE defines Z_{ij} by imputation followed by quantile normalization and log transformation. Second, SPIDE changes the definition of $W_{ii'j}$ in $P_{ii'j} = \frac{W_{ii'j}}{\sum_{i'' \in \mathbb{N}(i)} Wii''j}$. Recall that in SCENT, $W_{ii'j} = Z_{ij} Z_{i'j}$, so $P_{ii'j} = \frac{Z_{i'j}}{\sum_{i'' \in \mathbb{N}(i)} Z_{i''j}}$. In contrast, SPIDE defines $W_{ii'j} = |R_{ii'j}| Z_{ij} Z_{i'j}$, where $R_{ii'j}$ the Pearson correlation between genes i and i' on the basis of the imputed, normalized, and log-transformed expression levels (in publication; in the code, $R_{ii'j}$ is calculated using the $\log_2(\text{count} + 1.1)$ transformed expression levels) in cell j's k-nearest neighbors. Hence, $P_{ii'j}$ in SPIDE becomes

$$P_{ii'j} = \frac{|R_{ii'j}| Z_{i'j}}{\sum_{i'' \in N_i} |R_{ii''j}| Z_{i''j}}.$$

Third, SPIDE changes the definition of A in $\Pi_{ij} = \frac{Z_{ij} (A \cdot Z_{\cdot j})_i}{Z_{\cdot j}^1 \cdot (A \cdot Z_{\cdot j})}$, the steady-state probability of gene i in cell j. Recall that in SCENT, A is the binary gene-gene adjacency matrix from the PPI network. In contrast, SPIDE changes A to A_j, which becomes specific to cell j, and defines the (i, i')th element of A_j as

$$A_{ii'j} = |R_{ii'j}| A_{ii'}.$$

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

Finally, SPIDE estimates the *potency* for cell j as

$$NE_j = E_j/C,$$

where C is exactly the same C as in SCENT, i.e., $C = \log \lambda$, and λ is the largest eigenvalue for the binarized adjacency matrix A. Because SPIDE modifies the definition of E_j, the constant C from SCENT no longer provides the stated normalization under the published formulation.

As SPIDE is conceptually similar to SCENT, they share the same intuition.

The hyperparameters in SPIDE include the offset value of 1.1 used in the log transformation $Z_{ij} = \log_2(S_{ij} + \text{offset})$ and the number of nearest neighbors (k) of each cell. In total, we count SPIDE to have two hyperparameters.

stemFinder

stemFinder is inspired by the observation that cell-cycle gene expression correlates with cell potency, and it quantifies a cell's potency using the heterogeneity of cell-cycle gene expression among its neighboring cells.

By default, stemFinder expects a Seurat-processed scRNA-seq object with log-normalized and scaled gene expression values, selected highly variable genes, and pre-computed PCA and k-nearest neighbor (kNN) graphs, corresponding to the default Seurat pipeline. Specifically, the log-normalized expression of gene i in cell j is

$$Z_{ij} = \log\left(\frac{X_{ij}}{\sum_{i'}^{p} X_{i'j}} \times 10,000 + 1\right),$$

which is then scaled (denoted as $\tilde{Z}$ by subtracting, for each gene i, its mean across cells and dividing by its standard deviation across cells. Let $\mathcal{C}$ denote the set of cell-cycle genes, and let $\mathcal{V}$ denote the set of highly variable genes. PCA is performed on the scaled expression of non-cell-cycle highly variable genes in the set $\mathcal{V}\backslash\mathcal{C}$, and a kNN graph is then constructed in this principal component (PC) space with a user-defined number of PCs and a default $k = \sqrt{n}$, where n is the number of cells.

To measure the local heterogeneity in cell-cycle gene expression for a cell j, stemFinder first binarizes each cell-cycle gene $i \in \mathcal{C}$ using a threshold τ (zero by default):

$$B_{ij} = \begin{cases} 1, & \tilde{Z}_{ij} > \tau, \\ 0, & \tilde{Z}_{ij} \le \tau, \end{cases} \quad j = 1, \ldots, n.$$

For a given cell j, the heterogeneity of each cell-cycle gene i is then quantified over the cell's k neighbors (excluding itself) using the Gini impurity:

$$h_{ij} = q_{ij}(1 - q_{ij}),$$

where q_{ij} is the proportion of cell j's k neighbors that share the same binarized state as cell j for gene i. As an alternative to binarization-based Gini impurity, one may also quantify heterogeneity using the standard deviation or variance of the log-normalized expression values across the k neighbors of cell j for gene i. We consider this heterogeneity an entropy-like measure. The overall raw heterogeneity score for cell j is then defined as the sum of the heterogeneity of all cell-cycle genes,

$$H_j = \sum_{i \in \mathcal{C}} h_{ij},$$

where higher values of H_j indicate a more stem cell like (less differentiated) state.

Finally, stemFinder defines the *potency* for each cell j as

$$\text{SF}_j = 1 - \frac{H_j}{\max_{j' \in \{1,\cdots,n\}} H_{j'}},$$

so that each score equals one minus the cell's raw heterogeneity score scaled by the maximum across all cells and lies between 0 (most stem cell like) and 1 (most differentiated). Note that this direction is opposite to that of the other potency estimation methods in this review, because the final stemFinder score is designed to reflect the cell's position along the differentiation trajectory (from stem cell like to differentiated) rather than its differentiation capacity.

The intuition behind stemFinder is that a cell is more stem cell like when the expression of cell-cycle genes is more heterogeneous across similar cells.

The hyperparameters in stemFinder include the number of nearest neighbors k and the threshold τ used to binarize the expression values of cell-cycle genes.

Category 3: Knowledge-based methods

Among the four potency estimation methods in this category, mRNAsi,[35] CCAT,[36] and FitDevo[37] explicitly estimate cell potency as the correlation between each cell's gene expression profile and a predefined vector of gene importance scores that is not cell specific. Whereas CytoTRACE2[38] uses predefined potency-related gene sets and a neural network model to predict each cell's potency category and derive a corresponding potency score.

mRNAsi

As a method that applies to both bulk RNA-seq and scRNA-seq data, mRNAsi[35] trains a one-class logistic regression (OCLR) model, where genes are features (with gene expression levels as feature values), cells (or bulk samples) are observations (also known as instances), and the only class in training data contains bulk stem-cell samples only. From the trained model, mRNAsi extracts genes' coefficients as genes' importance scores. Finally, mRNAsi defines the *potency* for each cell (or sample) in a new dataset as the Spearman correlation between the same genes' expression levels in the cell (or sample) and the importance scores. mRNAsi chooses the Spearman correlation instead of the Pearson correlation for better robustness to batch effects that differ between the training data and the new data.

The intuition behind mRNAsi is that a cell (or sample) has a higher potency if the rank of genes' expression levels in it is consistent with the genes' importance scores obtained from the pre-trained OCLR model.

There are no apparent hyperparameters in mRNAsi.

mRNAsi is a tool applicable to both bulk RNA-seq and scRNA-seq data, and it has gained popularity in cancer research. For instance, Chen et al.[50] utilized cell potency predicted by mRNAsi as a prognostic factor for lung adenocarcinoma. Similarly, Xavier

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

et al.[51] reported elevated cell potency predicted by mRNAsi in simple canine mammary tumors compared to other subtypes, highlighting its relevance in oncology.

CCAT

Unlike mRNAsi, CCAT[36] uses the Pearson correlation instead of the Spearman correlation. Developed by the same group of authors, CCAT is a fast proxy of SCENT that remedies SCENT's scalability issue. Instead of calculating cell-specific gene steady-state probabilities and gene-gene transition probabilities as SCENT does, CCAT estimates the *potency* for each cell using the Pearson correlation between the log-normalized gene expression levels in the cell and the genes' degrees in the inputted PPI network(In CCAT, the log-normalized expression of gene i in cell j is defined as $Z_{ij} = \log_2(S_{ij} + 1)$, where S_{ij} is X_{ij} [the count of gene i in cell j] multiplied by cell j's specific scaling factor calculated externally. Note that the definition of Z_{ij} is similar in CCAT and SCENT, except that the constant inside $\log_2$ is 1 in CCAT and 1.1 in SCENT.). However, the authors showed that CCAT approximates SCENT under the global mean-field assumption that every gene's neighboring genes have a constant expression level equal to the average expression level of all genes in each cell.

The intuition behind CCAT is that a cell has a higher potency if its gene expression levels has a stronger correlation with the genes' degrees in the PPI network; that is, the cell should have hub genes more highly expressed and non-hub genes more lowly expressed. While CCAT was shown to be robust to the choice of PPI network, an open question is whether the biological specificity of the network matters, or whether only general structural properties shared across PPI networks under different conditions are important (see discussion).

There are no apparent hyperparameters in CCAT.

FitDevo

Similar to CCAT, FitDevo[37] defines a cell's potency as the Pearson correlation between genes' importance scores, which are dataset-specific unlike in CCAT, and the expression levels of the same genes in the cell. FitDevo defines each gene's importance score through a complex procedure, calculating it as the sum of two binary scores: *binarized gene weight* (*BGW*) and *predicted BGW* (*pBGW*). Note that the BGWs are precomputed on the basis of 17 existing datasets and irrelevant to user-inputted scRNA-seq data, while the pBGWs are calculated on the basis of the BGWs and user-inputted scRNA-seq data. The calculation of BGWs and pBGWs is detailed below.

(1) To calculate the BGWs, one per gene, FitDevo uses the 17 datasets from CytoTRACE's publication,[23] where cells have discrete time-point labels. FitDevo calculates BGWs for $p = 14,717$ genes, which are expressed in at least 9 out of the 17 datasets. These BGWs are stored and made available in FitDevo. The detailed calculation is described in the following sub-steps.

(1.1) For each dataset $k = 1,\ldots,17$ with $n^{(k)}$ cells, we use $X_{ij}^{(k)}$ to indicate the expression level of gene $i = 1,\ldots,p$ in dataset k's cell $j = 1,\ldots,n^{(k)}$. (The FitDevo publication[37] does not specify the unit of the expression level, so it is unclear whether the expression level is count, normalized value, or log-normalized value.) Moreover, we use $T_j^{(k)}$ to denote the reverted time-point label for cell $j = 1,\ldots,n^{(k)}$ in dataset k. The time points are reverted so that a potency-related gene, whose expression decreases from early to late time points, can have a large, positive correlation with the reverted time points. Then, for each gene $i = 1,\ldots,p$ and dataset $k = 1,\ldots,17$, FitDevo computes C_{ik}, the Pearson correlation between gene i's expression levels in dataset k's $n^{(k)}$ cells $\left(X_{i1}^{(k)}, X_{i2}^{(k)}, \ldots, X_{in^{(k)}}^{(k)}\right)^T$ and the $n^{(k)}$ cells' reverted time-point labels $\left(T_1^{(k)}, T_2^{(k)}, \ldots, T_{n^{(k)}}^{(k)}\right)^T$.

(1.2) For each gene i, FitDevo standardizes each C_{ik} as $C'_{ik} = C_{ik}/\sqrt{\sum_{k-1}^{17}\frac{(C_{ik'})^2}{17-1}}$.

(1.3) FitDevo defines the gene weight (GW) for gene i by averaging the 17 standardized Pearson correlations as $GW_i = \sum_{k=1}^{17} C'_{ik}/17$.

(1.4) For each gene i, FitDevo binarizes the GW by the sign to obtain the BGW: $BGW_i = 1$ if $GW_i > 0$; $BGW_i = 0$ otherwise.

(2) To calculate the pBGWs, one per gene, FitDevo uses the pre-computed BGWs for the p genes and the user-inputted scRNA-seq data with n cells. The detailed calculation is described in the following sub-steps.

(2.1) By default, FitDevo follows the Seurat V4[52] pipeline (now also compatible with the Seurat V5[53] pipeline) to obtain the top $L = 50$ PCs from Z, i.e., the user-inputted scRNA-seq data after normalization, log transformation, highly variable gene selection, and scaling in Seurat. Specifically, $Z_{ij} = \log\left(\left(X_{ij}/\sum_{i'=1}^{P} X_{i'j}\right) \times 10,000 + 1\right)$, the (i,j)th entry of Z, denotes the log-normalized expression level of gene $i = 1,\ldots,p$ in cell $j = 1,\ldots,n$ of the user-inputted scRNA-seq data. We use M_{lj} to note the lth PC score of cell $j = 1,\ldots,n$ in the user-inputted scRNA-seq data, $l = 1,\ldots,L$.

(2.2) FitDevo calculates O_{il} as the Pearson correlation between gene i's log-normalized expression levels $(Z_{i1}, Z_{i2}, \ldots, Z_{in})^T$ and the lth PC $(M_{l1}, M_{l2}, \ldots, M_{ln})$, $i = 1,\ldots,p$; $l = 1,\ldots,L$.

(2.3) FitDevo fits a generalized linear model using the p genes' BGWs as the response and the genes' L correlations with the top L PCs as the covariates:

$$\begin{cases} BGW_i & \sim \text{Binomial}(\mu_i); \\ \text{logit}(\mu_i) & = \alpha + \sum_{l=1}^{L} \beta_l O_{il}. \end{cases}$$

(2.4) The parameters $\alpha, \beta_1, \ldots, \beta_L$ are estimated as $\widehat{\alpha}, \widehat{\beta}_1, \ldots, \widehat{\beta}_L$. Then, the predicted $\text{logit}(\mu_i)$ for gene i is $\text{logit}(\widehat{\mu}_i) = \widehat{\alpha} + \sum_{l=1}^{L} \widehat{\beta}_l O_{il}$. FitDevo defines the pBGW as

$$pBGW_i = \begin{cases} 1, & \text{if } \text{logit}(\widehat{\mu}_i) > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

With the p genes' BGWs and pBGWs, FitDevo defines the sample-specific GW (SSGW) for gene $i = 1, \ldots, p$ by adding up gene i's BGW and pBGW: $SSGW_i = BGW_i + pBGW_i$.

Finally, FitDevo defines the *potency* for a particular cell j in the user-inputted scRNA-seq data as the Pearson correlation between all p genes' log-normalized expression levels in this cell $(Z_{1j}, Z_{2j}, \ldots, Z_{pj})^T$ and their corresponding SSGWs $(SSGW_1, SSGW_2, \ldots, SSGW_p)^T$.

The only hyperparameter in FitDevo is the number of PCs used (L) when calculating pBGWs.

Additionally, although FitDevo depends on the Seurat pipeline,[52] it includes a "no normalization" option, allowing users to input already-normalized data that do not require further normalization.

CytoTRACE2

CytoTRACE2 differs from the original CytoTRACE in that it uses a neural network that learns potency-associated gene sets, defines discrete potency categories, and outputs calibrated potency scores that are comparable across datasets. CytoTRACE2 is implemented as a pipeline comprising three core steps: data preprocessing, modeling, and result postprocessing. The modeling step uses parameters learned from training datasets, and these parameters are kept fixed when the model is applied to a new dataset.

For clarity, we describe CytoTRACE2 by first introducing the model structure, then outlining the training procedure, and finally detailing how to apply the trained model to a new dataset.

Model structure. The core model of CytoTRACE2 is a neural network that takes a gene-by-cell log-normalized expression matrix Z and a gene-by-cell rank matrix L as inputs and outputs both a predicted potency category and a raw potency score for each cell. To construct Z, the input scRNA-seq gene-by-cell count matrix X is first restricted to the 14,271 genes in a curated human/mouse gene dictionary, assigning zero expression to any genes in this reference that are absent from the input dataset in all cells. Let $X' \in \mathbb{Z}_{\geq 0}^{p \times n}$ denote this restricted gene-by-cell count matrix with $p = 14,271$, then Z is defined as

$$Z_{ij} = \log_2\left(\frac{X'_{ij}}{\sum_{i'=1}^{p} X'_{i'j}} \times 10^6 + 1\right).$$

The gene-by-cell rank matrix L is defined as $L \in \mathbb{Z}_{\geq 1}^{p \times n}$, where

$$L_{ij} = \text{rank}\left(X'_{ij}; X'_{\cdot j}\right)$$

denotes the within-cell rank of gene i's expression in cell j among all genes, with rank 1 corresponding to the gene with the highest expression. Using the information from six pre-trained gene set binary networks (GSBNs), $B^{(g)} \in \{0,1\}^{p \times M}$ for $g \in \{1, \ldots, 6\}$, which encode the membership of genes in M gene sets for each potency category, CytoTRACE2 then transforms Z and L into expression-based and rank-based enrichment scores. In this formulation, each GSBN corresponds to one of six potency categories, so gene memberships in the M gene sets are category specific. For cell $j = 1, \ldots, n$, potency category $g = 1, \ldots, 6$, and gene set $m = 1, \ldots, M$, the expression-based enrichment score $V_{jm}^{(g)}$ is defined as

$$V_{jm}^{(g)} = \text{FG}_{jm}^{(g)} - \text{BG}_{jm}^{(g)},$$

where

$$\text{FG}_{jm}^{(g)} = \frac{\sum_{i=1}^{p} Z_{ij} B_{im}^{(g)}}{n_m^{(g)}}, \text{ with } n_m^{(g)} = \sum_{i=1}^{p} B_{im}^{(g)} \qquad \text{(Equation 1)}$$

is the average log-normalized expression of genes in potency category g's gene set m in cell j, and

$$\text{BG}_{jm}^{(g)} = \frac{\sum_{i=1}^{p} \sum_{i'=1}^{p} Z_{ij} H_{ii'} B_{i'm}^{(g)}}{\sum_{i=1}^{p} \sum_{i'=1}^{p} H_{ii'} B_{i'm}^{(g)}} \qquad \text{(Equation 2)}$$

is the average log-normalized expression of these genes' corresponding background genes. Here, the gene-to-background-gene mapping is encoded by $H \in \{0,1\}^{p \times p}$, which is learned from the training datasets and kept fixed when applying CytoTRACE2 to new data. For cell j, potency category g, and gene set m, the rank-based enrichment score $U_{jm}^{(g)}$ is defined as

$$U_{jm}^{(g)} = 1 + \left[\frac{n_m^{(g)}\left(n_m^{(g)} + 1\right) - 2\sum_{i=1}^{p} T_{ij}^{(g)} B_m^{(g)}}{2\tau^{(g)} n_m^{(g)}}\right], \qquad \text{(Equation 3)}$$

$$T_{ij}^{(g)} = \min\left(L_{ij}, \tau^{(g)}\right). \qquad \text{(Equation 4)}$$

$T_{ij}^{(g)}$ is the potency-category-specific trimmed rank, and $\tau^{(g)}$ is a truncation parameter learned from the training datasets. Then, for each cell j and potency category g, the expression-based and rank-based enrichment scores are concatenated into a vector

$$K_j^{(g)} = \left(U_{j1}^{(g)}, \ldots, U_{jM}^{(g)}, V_{j1}^{(g)}, \ldots, V_{jM}^{(g)}\right)^{\cdot} \in \mathbb{R}^{2M},$$

and the standardized $K_j^{(g)}$ (obtained by subtracting the mean and dividing by the standard deviation across all cells, denoted $\tilde{K}_j^{(g)}$) is passed to a feedforward layer that maps $\tilde{K}_j^{(g)}$ to a category-specific logit:

$$q_{jg} = \tilde{K}_j^{(g)} \cdot \beta^{(g)} + b^{(g)}, \qquad \text{(Equation 5)}$$

where the slope $\beta^{(g)}$ and intercept $b^{(g)}$ are learned from the training datasets. These logits are converted to probabilities via a softmax layer,

$$\pi_{jg} = \frac{\exp(q_{jg})}{\sum_{g'=1}^{6} \exp(q_{jg'})},$$

and the predicted potency category for cell j is given by

$$\hat{y}_j = \arg\max_{g \in \{1,\ldots,6\}} \pi_{jg}.$$

The raw potency score for cell j is then defined as

$$\text{RPS}_j = 0 \cdot \pi_{j1} + 0.2 \cdot \pi_{j2} + 0.4 \cdot \pi_{j3} + 0.6 \cdot \pi_{j4} + 0.8 \cdot \pi_{j5} + 1.0 \cdot \pi_{j6}.$$

Training procedure. Using 19 labeled scRNA-seq datasets, in which each cell is labeled with one of the 6 potency categories,

CytoTRACE2 jointly estimates the 6 GSBNs and the parameters of the feedforward (logistic regression) layer, while the gene-to-background-gene mapping used in the enrichment score computation is derived directly from the same training data. In training, each cell j has an known potency category label y_j and a predicted potency category label $\widehat{y}_j$ (as defined in "model structure" above), and CytoTRACE2 minimizes the objective:

$$J\left(\left\{W^{(g)}\right\}_{g=1}^{6},\left\{\alpha^{(g)}\right\}_{g=1}^{6}\right)=J_S\left(\left\{W^{(g)}\right\}_{g=1}^{6},\left\{\alpha^{(g)}\right\}_{g=1}^{6}\right)+\sum_j J_P\left(\widehat{y}_j,y_j\right).$$

The gene-set-size penalty is

$$J_S\left(\left\{W^{(g)}\right\}_{g=1}^{6},\left\{\alpha^{(g)}\right\}_{g=1}^{6}\right)=a\lambda\sum_{g=1}^{G}\left\|\frac{1}{N}\left(B^{(g)}\right)^{\top}B^{(g)}\odot I_M\right\|_F,$$

$$a=\frac{\sqrt{12}}{\sqrt{M}},B^{(g)}=1\left(W^{(g)}>\alpha^{(g)}\right).$$

Here, I_M is the $M\times M$ identity matrix, $\odot$ denotes element-wise multiplication, and $\alpha^{(g)}$ is a threshold used to binarize $W^{(g)}$ into $B^{(g)}$. The prediction loss for cell j is

$$J_P\left(\widehat{y}_j,y_j\right)=\omega_j\mathrm{CE}\left(\widehat{y}_j,y_j\right),$$

where ω_j is the cell-specific loss weight that depends on y_j (to prevent the loss from being dominated by major categories), $\widehat{y}_j$ is the model-predicted label given the current $B^{(1)},\ldots,B^{(6)}$, and $\mathrm{CE}\left(\widehat{y}_j,y_j\right)$ is the cross-entropy loss. Among the parameters involved in this model, $\beta^{(g)}$ and $b^{(g)}$ in Equation 5 are directly learnable feedforward-layer parameters. In contrast, $B^{(g)}$ in Equations 1, 2, and 3 and $\tau^{(g)}$ in Equations 3 and 4 are not learned directly; instead, they are transformations of other learnable parameters. Specifically, CytoTRACE2 directly optimizes the objective function to obtain the continuous GSBN weight matrix $\widehat{W}^{(g)}$, the binarization threshold $\widehat{\alpha}^{(g)}$, and the auxiliary rank-cutoff parameter $\widehat{\eta}^{(g)}$ used to compute the category-specific rank cutoff. After training, each binary GSBN matrix is obtained by thresholding the learned continuous weights,

$$\widehat{B}^{(g)}=1\left(\widehat{W}^{(g)}>\widehat{\alpha}^{(g)}\right),$$

and each category-specific rank cutoff is computed as

$$\widehat{\tau}^{(g)}=10+\max_{1\le m\le M}\widehat{n}_m^{(g)}+1,000\times\max\left(0,\widehat{\eta}^{(g)}\right),$$

with

$$\widehat{n}_m^{(g)}=\sum_{i=1}^{p}\widehat{B}_{im}^{(g)}.$$

In addition to these optimized parameters, CytoTRACE2 also constructs a gene-to-background-gene mapping matrix $\widehat{H}$ in Equation 2 from the training data. All genes are first ranked by their average expression in the training data and then partitioned into 24 bins, each of size n_{bin}. For each gene i, a background set of genes with similar average expression is sampled from the same bin, with the set size drawn from a Gaussian distribution with mean 100 and variance $100\times\frac{n_{\mathrm{bin}}-100}{n_{\mathrm{bin}}}$. The entry $\widehat{H}_{ii'}$ is set to 1 if gene i' is selected as a background gene for gene i, and to 0 otherwise.

Application to new data. Given this neural network with pretrained parameters $\widehat{B}^{(g)}$, $\widehat{\tau}^{(g)}$, $\widehat{\beta}^{(g)}$, $\widehat{b}^{(g)}$, and $\widehat{H}$, when a new scRNA-seq gene-by-cell count matrix X is supplied, CytoTRACE2 outputs, for each cell j, a predicted potency category $\widehat{y}_j$ and the raw potency score RPS_j. Then, CytoTRACE2 applies three smoothing steps to RPS_j to obtain the final estimated potency score, described as follows:

(1) The same smoothing procedure based on a diffusion process as the fourth step in CytoTRACE's description to obtained the smoothed potency score denoted by SPS, one per cell.
(2) Within each predicted potency category, cells are ranked from 1 to the total number of cells assigned to that category, with rank 1 assigned to the cell with the lowest SPS within the category. These within-category ranks are then rescaled to a fixed interval, yielding the binned potency score SPS^B, one per cell. The interval assigned to each potency category is listed below:

Predicted potency category	Assigned interval
Differentiated	[0,1/6]
Unipotent	[1/6,2/6]
Oligopotent	[2/6,3/6]
Multipotent	[3/6,4/6]
Pluripotent	[4/6,5/6]
Totipotent	[5/6,1]

(3) Finally, CytoTRACE2 smooths SPS^B by averaging over each cell's value over its k-nearest neighbors, defined in the 30-dimensional PC space of log normal gene expression, using Euclidean distance. The final CytoTRACE2 *potency* score for cell j is computed as a distance-weighted average within the cell's neighborhood $N(j)$:

$$\mathrm{CytoTRACE2}_j=\frac{\sum_{c\in N(j)}\mathrm{SPS}_c^B\left(1-d_c\right)^2}{\sum_{c\in N(j)}\left(1-d_c\right)^2},$$

where d_c is the Euclidean distance between cell j and its neighbor cell c in the PC space.

The intuition behind CytoTRACE2 is that a cell has higher potency when gene sets associated with high-potency categories (for example, totipotent) are more highly expressed in that cell, while gene sets associated with differentiated states are less highly expressed.

The four hyperparameters used in CytoTRACE's smoothing step are retained in the first smoothing step of CytoTRACE2. Additionally, CytoTRACE2 sets the number of gene sets (M) for each potency category, a fixed number of bins (24) for selecting background genes, and the mean (100) and variance ($100\times\frac{n_{\mathrm{bin}}-100}{n_{\mathrm{bin}}}$) for the Gaussian distribution used to sample the number of background genes, so that CytoTRACE2 has eight hyperparameters in total.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762



Figure 3. A radar plot of the 16 cell potency estimation methods, illustrating their number of inputs, number of hyperparameters, and the number of differences between publication and implementation we used, reflecting the simplicity (fewer inputs and hyperparameters) and consistency (fewer differences) of the methods

DISCUSSION

In this review, we summarized 16 cell potency estimation methods into 3 categories, including average-based methods, entropy-based methods, and knowledge-based methods. For each of the 16 methods, we provided a high-level intuition and used a unified set of mathematical notations to detail the procedure. We also summarized the 16 methods' usage complexities, including the number of hyperparameters, the number of required inputs, and the existence of differences between the method description in publications and the method implementation in software packages in Figure 3. Last, we summarized the differences among these 16 methods regarding their inputs and outputs in Table 2.

By closely examining the 16 methods, we identified some open questions for future work. First, as many of the 16 methods require users to input a PPI network, a natural question is the PPI network choice. CCAT evaluated its performance across three different PPI networks (as shown in Figure 3 of the CCAT publication[36]) and stated, "the robust

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

association of CCAT with differentiation potency is due to the subtle positive correlation between transcriptome and connectome, which, as shown, is itself robust to the choice of PPI network."[36] In this study, we also conducted a sensitivity analysis to assess the robustness of CCAT to perturbations in the PPI network. Using the *in vitro* NPCs (C1) dataset from the CytoTRACE publication, we randomly permuted 10% to 100% of the gene labels (nodes) in the PPI adjacency matrix provided with the CCAT package and reran CCAT with the same scRNA-seq data and each perturbed network. As shown in Figure S1, CCAT's potency estimates remained largely stable up to 70% node permutation, with noticeable performance degradation only under near-complete randomization.

We applied the same evaluation strategy to the other methods that take a PPI network as input (ORIGINS, NCG, SCENT, and SPIDE; results for MCE are not included due to its substantially longer runtime and a runtime error encountered during these experiments) and observed a similar pattern: as the perturbation level increases, potency estimation accuracy decreases. This holds even though the network plays a less direct role for these methods than for CCAT, where node degree directly determines gene importance. These observations raise the question of how much information in the PPI network actually drives potency estimates, and where along the spectrum of network perturbations the biologically meaningful signal leveraged by these methods resides.

Second, the fundamental assumption of CytoTRACE is that gene count (i.e., the number of expressed genes) reflects a cell's potency. As gene count is highly correlated with cell library size (Figure S5D of the CytoTRACE publication[23]), we asked whether cell library size itself carries information relevant to cell potency. To investigate, we implemented a variant of CytoTRACE that selects the top 200 genes on the basis of their correlation with library size rather than gene count and applied it to the 12 single-lineage datasets from the CytoTRACE publication. This variant achieved performance comparable to the original CytoTRACE (Figure S2); however, given the strong positive monotonic relationship between gene count and library size consistent with Heaps' law,[54] this similarity is expected. As a result, this analysis does not establish that library size provides an independent signal for potency estimation; rather, it suggests that the contributions of gene count and library size cannot be disentangled within the CytoTRACE framework. More broadly, the biological relevance of library size remains an open question. On one hand, a recent spatial transcriptomics study found an association between cell library size and tissue structure and showed that preserving library size improves spatial domain detection accuracy.[55] On the other hand, library size does not always reflect potency: B cell activation produces a nearly 10-fold increase in total RNA,[56] and total mRNA counts vary substantially across organs and cell types without implying differences in potency.[57] Thus, while our preliminary results suggest that library size may be informative for potency estimation, its precise relationship to cell potency warrants further investigation.

Third, and most importantly, we observed that there is currently no consensus on which quantitative characteristics of scRNA-seq data best define cell potency, highlighting an opportunity for further benchmarking and methodological development in the field. Specifically, the average-based methods, CytoTRACE and NCG, define a cell's potency on the basis of the expression levels of a set of "important" genes, with each method using its own criteria to determine importance while underweighting the expression levels of other genes. Similarly, knowledge-based methods assign higher potency to a cell when genes with larger importance scores, or gene sets associated with more potent categories, show higher activation in that cell. In contrast, some entropy-based methods, such as stemID, dpath, scEnergy, and cmEntropy, operate on the intuition that a cell has higher potency when its genes have similar expression levels, without explicit specification of gene importance. These differing perspectives underscore the need for a clearer theoretical and empirical understanding of how gene expression patterns relate to cell potency, which could guide the development of more generalizable potency estimation methods.

Additionally, we would like to raise a few methodological points for further consideration. The first open methodological question is how best to define the distance when *k*-means is applied to a gene-gene dissimilarity matrix; SLICE currently uses Euclidean distance on the dissimilarity matrix, and further work could clarify the implications of this choice. The second point concerns dpath's use of the Poisson distribution to model log-transformed TPM values. The Poisson distribution is designed for count data, where the mean equals the variance by assumption. Log-transformed TPM values, however, are continuous, non-negative real numbers that do not satisfy this mean-variance relationship, and their distribution can differ substantially from Poisson. Consequently, the Poisson-based dropout weights used in the weighted NMF are only approximate, and the resulting factorization may be miscalibrated (e.g., by over- or underweighting genes with certain expression levels), which in turn could affect the reliability of the potency estimates. Thirdly, these methods employ various approaches to normalize gene expression levels, and systematic comparisons of normalization strategies for potency estimation remain an open area for future benchmarking. We conducted a small benchmark of normalization methods using CCAT as an example, since it does not prescribe a default normalization strategy. The choice of normalization method moderately affects CCAT's performance (Spearman correlation with ground truth: Census = 0.647, logCP10K = 0.659, logCPM = 0.704, and PFlogPF = 0.702), and adopting a top-performing normalization approach from previous benchmark,[58] PFlogPF, can enhance the CCAT results. This example underscores the potential to improve the performance of existing cell potency estimation methods through optimizing the normalization step. Rather than promoting a universal normalization solution for all scRNA-seq analyses, we highlight the importance of applying carefully evaluated normalization strategies tailored to the specific task of estimating cell potency from scRNA-seq data.

In light of the diverse assumptions and methodological strategies employed by existing approaches, as well as emerging developments such as IDEAS,[59] which uses the intrinsic dimensionality of transcriptional profiles to estimate potency with a particular focus on reprogramming datasets, we consider the task of estimating cell potency from scRNA-seq

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

data to remain an active area of investigation, with three remaining challenges. First, comprehensive benchmarking with single-cell resolution remains an open challenge. Although the 16 methods included internal evaluations, these were primarily conducted on datasets (Table S3) with annotated time points or cell types of known differentiation order, rather than datasets with ground-truth potency labels at the single-cell level. This creates a resolution gap between group-level reference labels and single-cell potency estimates (Figure S3); benchmarks with single-cell-level reference labels are needed to determine how much within-group variation in estimated potency is biologically meaningful. One promising direction is to leverage lineage-tracing data, in which individual stem or progenitor cells are labeled and their clonal descendants are profiled over time.[60] Studies such as Weinreb et al.'s clonal analysis of hematopoiesis[61] exemplify this approach: by counting the number of distinct downstream states accessible from a given progenitor, lineage tracing provides an experimentally grounded and direct measure of differentiation potential. A complementary direction is to use real-data-based synthetic null datasets to characterize potential biases in cell potency estimation.[62] Second, the internal benchmarking results of the 16 methods demonstrated substantial variability in performance across datasets, as shown in Figure 2C of the CytoTRACE publication[23] and Figures 3C and 3D of the FitDevo publication.[37] While these variations are well documented, the underlying factors driving them remain insufficiently understood. Given the conceptual differences among methods, such as whether they focus only on important genes or include all genes, further work is needed to understand how dataset characteristics influence method performance. This understanding will be key for informed method selection and future development. Third, an important open question is how existing cell potency estimation methods perform in challenging biological contexts and across distinct cell populations. Within a single dataset, it remains unclear whether current methods accurately capture developmental capacity in edge-case settings, such as nullipotent cell lines (e.g., F9 cells) and post-mitotic cells, contexts that could provide direct tests of whether estimated potency scores reflect the minimal or absent differentiation potential of these cells. A related challenge arises in PSC models in which lineage-specific regulators are perturbed: in these settings, potency scores should reflect perturbation-induced changes in differentiation potential rather than confounding factors such as variable perturbation efficiency or sequencing noise.[63,64] Finally, because most methods are applied within individual datasets, current potency scores are most reliably interpreted as relative rankings within a given population. Although methods such as CytoTRACE2 have begun to consider cross-dataset comparability, further work is needed to determine when and how potency scores can be compared across experiments, tissues, and perturbation conditions.

In addition to furthering our understanding of cell potency estimation, a broader open question is regarding the distinct advantages of computational methods that use different perspectives (pseudotime inference, RNA velocity inference, and cell potency estimation) to infer cellular dynamics from static scRNA-seq data. In particular, there is a lack of consensus on how the three types of computational methods should be used jointly, or whether one or more types are more suitable for certain biological contexts. In some cases, the outputs from the three types of methods are treated as equivalent. For example, CellRank2[65] accepts outputs of all three types of methods for constructing a cell-cell transition matrix, which it uses for downstream inference of cell states and cell fate probabilities. Additionally, Qin et al.[13] utilize the sum of the estimated potency from CCAT and the magnitude of RNA velocity from scVelo to represent the pluripotency of each cell. In other cases, the outputs of different types of methods are used sequentially. For example, estimated cell potency can facilitate trajectory inference in two ways: (1) some potency estimation methods (stemID, SLICE, dpath, scEnergy, and SCENT) directly construct cell trajectories from their potency estimates; and (2) other studies have used potency estimates to identify root cells for existing trajectory inference methods, as mentioned in the CCAT, ORIGINS, and stemFinder publications. These different usages may hinder the reproducibility of scRNA-seq data analysis results. Therefore, a comprehensive benchmarking and a deeper understanding of these three types of methods are necessary to improve the trustworthiness of discoveries made from scRNA-seq data.

Looking ahead, we believe integrating more well-established potency-related biological knowledge can enhance computational methods' ability to quantify cell potency and uncover biological insights. First, leveraging well-characterized potency-related gene regulatory networks (GRNs) can possibly provide a more accurate estimation of cell potency, as a cell's pluripotency is known to be influenced by regulatory activities among genes.[66–68] Second, incorporating the cell-cell interaction information may improve the accuracy of cell potency quantification, as stem cells reside in fixed anatomical areas called niches, and activities like cell-cell interactions inside niches influence the proliferation and differentiation of stem cells.[69–71] Last, utilizing single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq) data alongside scRNA-seq data may enhance computational methods' performance in quantifying cell potency, as chromatin accessibility and transcription factors' activities change during the cell differentiation process.[72–74]

In conclusion, we envision that experimental and computational approaches can work cooperatively to enhance the ability to quantify cell potency and identify more potency marker genes across various tissues and conditions. As a performance check, a robust cell potency estimation method should be able to recognize known potency marker genes. Such methods can then be used to discover novel marker genes, which biologists can experimentally validate (e.g., by knocking out these potential marker genes[75,76]). If these novel genes are proven to be potency-related, they can serve as new features to help quantify cell potency. Additionally, with increased knowledge of potency-related marker genes, biologists can further investigate the key signaling pathways and regulatory activities among these genes. This cooperation between computational algorithms and experimental validation can lead to more reliable and insightful findings regarding cellular differentiation.

To facilitate the single-cell community's future benchmark studies, we have curated Table S3, which includes the datasets used in these cell potency estimation methods, although these datasets do not provide single-cell-level cell potency ground truths.

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to, and will be fulfilled by, the lead contact, Jingyi Jessica Li (lijy03@fredhutch.org).

Materials availability

This study did not generate new materials.

Data and code availability

- This paper analyzes existing, publicly available data. The processed data used for the CytoTRACE analyses were kindly shared with us by the CytoTRACE authors. All data needed to reproduce the analyses are archived at Zenodo: https://doi.org/10.5281/zenodo.22785227.
- All code needed to reproduce the analyses is archived at Zenodo: https://doi.org/10.5281/zenodo.22785227.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work is supported by NSF DGE-2034835 to Q.W. and NIH/NIGMS R35GM140888, NIH/NHGRI R01HG014687, NSF DBI-1846216, and DMS-2113754, Johnson & Johnson WiSTEM2D Award, Sloan Research Fellowship, UCLA David Geffen School of Medicine W.M. Keck Foundation Junior Faculty Award, and the Chan-Zuckerberg Initiative Single-Cell Biology Data Insights to J.J.L.

The authors appreciate the comments and feedback from the Junction of Statistics and Biology members (https://jsb-lab.org). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-2034835. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

AUTHOR CONTRIBUTIONS

Q.W., Z.Z., D.S., and J.J.L. conceived the project. Q.W., Z.Z., and J.J.L. reviewed the cell potency estimation methods. Q.W., Z.Z., Q.L., D.S., and J.J.L. discussed, drew, and revised the figures. Q.W. summarized the datasets mentioned in the original methods' publications. Q.W. and J.J.L. wrote the manuscript. Z.Z., Q.L., D.S., and J.J.L. revised the manuscript. J.J.L. supervised the project. All authors participated in discussions and approved the final manuscript.

DECLARATION OF INTERESTS

J.J.L. serves on the *Cell Systems* advisory board.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - PPI network perturbation analysis
  - Comparison of gene-count- and library-size-based gene selection in CytoTRACE
  - Impact of normalization method on CCAT performance
  - Explanation of the difference between the geometric and arithmetic means in CytoTRACE

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2026.101762.

Received: December 2, 2025
Revised: June 28, 2026
Accepted: October 5, 2026

REFERENCES

1. National Cancer Institute. Cell differentiation. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/cell-differentiation.
2. Zhu, F., and Nie, G. (2025). Cell reprogramming: methods, mechanisms and applications. Cell Regen. *14*, 12. https://doi.org/10.1186/s13619-025-00229-x.
3. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386. https://doi.org/10.1038/nbt.2859.
4. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). Rna velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.
5. Aponte, P.M., and Caicedo, A. (2017). Stemness in cancer: stem cells, cancer stem cells, and their microenvironment. Stem Cells Int. *2017*, 5619472. https://doi.org/10.1155/2017/5619472.
6. Herman-Edelstein, M., Thomas, M.C., Thallas-Bonke, V., Saleem, M., Cooper, M.E., and Kantharidis, P. (2011). Dedifferentiation of immortalized human podocytes in response to transforming growth factor-β: a model for diabetic podocytopathy. Diabetes *60*, 1779–1788. https://doi.org/10.2337/db10-1110.
7. Jimenez, S.A., and Piera-Velazquez, S. (2024). Cellular transdifferentiation: a crucial mechanism of fibrosis in systemic sclerosis. Curr. Rheumatol. Rev. *20*, 388–404. https://doi.org/10.2174/0115733971261932231025045400.
8. Teschendorff, A.E., and Feinberg, A.P. (2021). Statistical mechanics meets single-cell biology. Nat. Rev. Genet. *22*, 459–476. https://doi.org/10.1038/s41576-021-00341-z.
9. Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics *19*, 477. https://doi.org/10.1186/s12864-018-4772-0.
10. Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-sne. J. Mach. Learn. Res. *9*, 2579–2605.
11. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.
12. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.
13. Qin, X., Cardoso Rodriguez, F.C., Sufi, J., Vlckova, P., Claus, J., and Tape, C.J. (2023). An oncogenic phenoscape of colonic stem cell polarization. Cell *186*, 5554–5568.e18. https://doi.org/10.1016/j.cell.2023.11.004.
14. Pang, L., Xiang, F., Yang, H., Shen, X., Fang, M., Li, R., Long, Y., Li, J., Yu, Y., and Pang, B. (2024). Single-cell integrative analysis reveals consensus cancer cell states and clinical relevance in breast cancer. Sci. Data *11*, 289. https://doi.org/10.1038/s41597-024-03127-0.
15. Giraud, J., Chalopin, D., Ramel, E., Boyer, T., Zouine, A., Derieppe, M.A., Larmonier, N., Adotevi, O., Le Bail, B., Blanc, J.F., et al. (2024). Thbs1+

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

myeloid cells expand in SLD hepatocellular carcinoma and contribute to immunosuppression and unfavorable prognosis through trem1. Cell Rep. *43*, 113773. https://doi.org/10.1016/j.celrep.2024.113773.

16. Cui, W., Chen, S., Hu, T., Zhou, T., Qiu, C., Jiang, L., Cheng, X., Ji, J., Yao, K., and Han, H. (2024). Nanoceria-mediated cyclosporin a delivery for dry eye disease management through modulating immune–epithelial crosstalk. ACS Nano *18*, 11084–11102. https://doi.org/10.1021/acsnano.3c11514.
17. Wu, R., Zhang, X., Zhang, X., Sun, L., Xia, T., and Zhang, L.J. (2023). Deciphering the age-dependent changes of pulmonary fibroblasts in mice by single-cell transcriptomics. Front. Cell Dev. Biol. *11*, 1287133. https://doi.org/10.3389/fcell.2023.1287133.
18. Singh, V.K., Saini, A., Kalsan, M., Kumar, N., and Chandra, R. (2016). Describing the stem cell potency: the various methods of functional assessment and in silico diagnostics. Front. Cell Dev. Biol. *4*, 134. https://doi.org/10.3389/fcell.2016.00134.
19. Barker, N., Van Es, J.H., Kuipers, J., Kujala, P., Van Den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., et al. (2007). Identification of stem cells in small intestine and colon by marker gene lgr5. Nature *449*, 1003–1007. https://doi.org/10.1038/nature06196.
20. Singh, S., Lian, Q., Budiman, T., Taketo, M.M., Simons, B.D., and Gupta, V. (2023). Heterogeneous murine peribiliary glands orchestrate compartmentalized epithelial renewal. Dev. Cell *58*, 2732–2745.e5. https://doi.org/10.1016/j.devcel.2023.10.004.
21. International Stem Cell Initiative (2018). Assessment of established techniques to determine developmental and malignant potential of human pluripotent stem cells. Nat. Commun. *9*, 1925. https://doi.org/10.1038/s41467-018-04011-3.
22. International Society for Stem Cell Research. (2026). Section 2: Pluripotency and the undifferentiated state. https://www.isscr.org/basic-research-standards/pluripotency.
23. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405–411. https://doi.org/10.1126/science.aax0249.
24. Senra, D., Guisoni, N., and Diambra, L. (2022). Origins: a protein network-based approach to quantify cell pluripotency from scrna-seq data. MethodsX *9*, 101778. https://doi.org/10.1016/j.mex.2022.101778.
25. Ni, X., Geng, B., Zheng, H., Shi, J., Hu, G., and Gao, J. (2022). Accurate estimation of single-cell differentiation potency based on network topology and gene ontology information. IEEE/ACM Trans. Comput. Biol. Bioinform. *19*, 3255–3262. https://doi.org/10.1109/TCBB.2021.3112951.
26. Grün, D., Muraro, M.J., Boisset, J.C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., Van Es, J., Jansen, E., Clevers, H., et al. (2016). De novo prediction of stem cell identity using single-cell transcriptome data. Cell Stem Cell *19*, 266–277. https://doi.org/10.1016/j.stem.2016.05.010.
27. Guo, M., Bao, E.L., Wagner, M., Whitsett, J.A., and Xu, Y. (2017). Slice: determining cell differentiation and lineage based on single cell entropy. Nucleic Acids Res. *45*, e54. https://doi.org/10.1093/nar/gkw1278.
28. Gong, W., Rasmussen, T.L., Singh, B.N., Koyano-Nakagawa, N., Pan, W., and Garry, D.J. (2017). Dpath software reveals hierarchical haemato-endothelial lineages of Etv2 progenitors based on single-cell transcriptome analysis. Nat. Commun. *8*, 14362. https://doi.org/10.1038/ncomms14362.
29. Jin, S., MacLean, A.L., Peng, T., and Nie, Q. (2018). scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. Bioinformatics *34*, 2077–2086. https://doi.org/10.1093/bioinformatics/bty058.
30. Kannan, S., Farid, M., Lin, B.L., Miyamoto, M., and Kwon, C. (2021). Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level. PLoS Comput. Biol. *17*, e1009305. https://doi.org/10.1371/journal.pcbi.1009305.
31. Teschendorff, A.E., and Enver, T. (2017). Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. Nat. Commun. *8*, 15599. https://doi.org/10.1038/ncomms15599.
32. Shi, J., Teschendorff, A.E., Chen, W., Chen, L., and Li, T. (2020). Quantifying Waddington's epigenetic landscape: a comparison of single-cell potency measures. Brief. Bioinform. *21*, 248–261. https://doi.org/10.1093/bib/bby093.
33. Xu, Z., Zheng, R., Chen, Y., Wang, E., and Li, M. (2022). A single cell potency inference method based on the local cell-specific network entropy. In 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 247–252. https://doi.org/10.1109/BIBM55620.2022.9995196.
34. Noller, K., and Cahan, P. (2024). Cell cycle expression heterogeneity predicts degree of differentiation. Brief. Bioinform. *25*, bbae536. https://doi.org/10.1093/bib/bbae536.
35. Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell *173*, 338–354.e15. https://doi.org/10.1016/j.cell.2018.03.034.
36. Teschendorff, A.E., Maity, A.K., Hu, X., Weiyan, C., and Lechner, M. (2021). Ultra-fast scalable estimation of single-cell differentiation potency from scRNA-Seq data. Bioinformatics *37*, 1528–1534. https://doi.org/10.1093/bioinformatics/btaa987.
37. Zhang, F., Yang, C., Wang, Y., Jiao, H., Wang, Z., Shen, J., and Li, L. (2022). FitDevo: accurate inference of single-cell developmental potential using sample-specific gene weight. Brief. Bioinform. *23*, bbac293. https://doi.org/10.1093/bib/bbac293.
38. Kang, M., Gulati, G.S., Brown, E.L., Qi, Z., Avagyan, S., Almagro Armenteros, J.J., Gleyzer, R., Zhang, W., Steen, C.B., D'Silva, J.P., et al. (2025). Improved reconstruction of single-cell developmental potential with CytoTRACE 2. Nat. Methods *22*, 2258–2263. https://doi.org/10.1038/s41592-025-02857-2.
39. Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with palantir. Nat. Biotechnol. *37*, 451–460. https://doi.org/10.1038/s41587-019-0068-4.
40. Cheng, S., Li, Z., Gao, R., Xing, B., Gao, Y., Yang, Y., Qin, S., Zhang, L., Ouyang, H., Du, P., et al. (2021). A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell *184*, 792–809.e23. https://doi.org/10.1016/j.cell.2021.01.010.
41. Li, J., Simmons, A.J., Hawkins, C.V., Chiron, S., Ramirez-Solano, M.A., Tasneem, N., Kaur, H., Xu, Y., Revetta, F., Vega, P.N., et al. (2024). Identification and multimodal characterization of a specialized epithelial cell type associated with cCohn's disease. Nat. Commun. *15*, 7204. https://doi.org/10.1038/s41467-024-51580-7.
42. Zhang, Z., Wang, Z.X., Chen, Y.X., Wu, H.X., Yin, L., Zhao, Q., Luo, H.Y., Zeng, Z.L., Qiu, M.Z., and Xu, R.H. (2022). Integrated analysis of single-cell and bulk rna sequencing data reveals a pan-cancer stemness signature predicting immunotherapy response. Genome Med. *14*, 45. https://doi.org/10.1186/s13073-022-01050-w.
43. DAVID Bioinformatics. (2005). Linear search algorithm in gene name batch viewer. https://davidbioinformatics.nih.gov/helps/linear_search.html.
44. Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). The david gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. *8*, R183. https://doi.org/10.1186/gb-2007-8-9-r183.
45. Guo, M., Bao, E.L., Wagner, M., Whitsett, J.A., Xu, Y. SLICE demo. GitHub. https://github.com/xu-lab/SLICE/tree/master/demo.
46. Banerji, C.R.S., Miranda-Saavedra, D., Severini, S., Widschwendter, M., Enver, T., Zhou, J.X., and Teschendorff, A.E. (2013). Cellular network entropy as the energy potential in waddington's differentiation landscape. Sci. Rep. *3*, 3039. https://doi.org/10.1038/srep03039.
47. Teschendorff, A.E., Sollich, P., and Kuehn, R. (2014). Signalling entropy: A novel network-theoretical framework for systems analysis and interpretation of functional omic data. Methods *67*, 282–293. https://doi.org/10.1016/j.ymeth.2014.03.013.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

48. Demetrius, L., and Manke, T. (2005). Robustness and network evolution—an entropic principle. Physica A *346*, 682–696. https://doi.org/10.1016/j.physa.2004.07.011.

49. Chen, W., Morabito, S.J., Kessenbrock, K., Enver, T., Meyer, K.B., and Teschendorff, A.E. (2019). Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome. Commun. Biol. *2*, 306. https://doi.org/10.1038/s42003-019-0554-8.

50. Chen, R., Liu, Y., and Xie, J. (2024). Construction of a pathomics model for predicting mrnasi in lung adenocarcinoma and exploration of biological mechanism. Heliyon *10*, e37100. https://doi.org/10.1016/j.heliyon.2024.e37100.

51. Xavier, P.L.P., Marção, M., Simões, R.L.S., Job, M.E.G., de Francisco Strefezzi, R., Fukumasu, H., and Malta, T.M. (2024). Machine learning determines stemness associated with simple and basal-like canine mammary carcinomas. Heliyon *10*, e26714. https://doi.org/10.1016/j.heliyon.2024.e26714.

52. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

53. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.

54. Lazzardi, S., Valle, F., Mazzolini, A., Scialdone, A., Caselle, M., and Osella, M. (2023). Emergent statistical laws in single-cell transcriptomic data. Phys. Rev. E *107*, 044403. https://doi.org/10.1103/PhysRevE.107.044403.

55. Bhuva, D.D., Tan, C.W., Salim, A., Marceaux, C., Pickering, M.A., Chen, J., Kharbanda, M., Jin, X., Liu, N., Feher, K., et al. (2024). Library size confounds biology in spatial transcriptomics data. Genome Biol. *25*, 99. https://doi.org/10.1186/s13059-024-03241-7.

56. Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.-R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., et al. (2013). Global regulation of promoter melting in naive lymphocytes. Cell *153*, 988–999. https://doi.org/10.1016/j.cell.2013.04.033.

57. Kim, Y.K., Cho, B., Cook, D.P., Trcka, D., Wrana, J.L., and Ramalho-Santos, M. (2023). Absolute scaling of single-cell transcriptomes identifies pervasive hypertranscription in adult stem and progenitor cells. Cell Rep. *42*, 111978. https://doi.org/10.1016/j.celrep.2022.111978.

58. Booeshaghi, A.S., Hallgrímsdóttir, I.B., Gálvez-Merchán, Á., and Pachter, L. (2022). Depth normalization for single-cell genomics count data. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.06.490859.

59. Staiano, M., Cirone, N., Biondo, M., Osella, M., and Scialdone, A. (2025). Intrinsic dimensionality of single-cell transcriptomic data reveals potency landscapes during cell reprogramming. Preprint at bioRxiv. https://doi.org/10.1101/2025.07.21.665922.

60. Zhang, Y., Zeng, F., Han, X., Weng, J., and Gao, Y. (2020). Lineage tracing: technology tool for exploring the development, regeneration, and disease of the digestive system. Stem Cell Res. Ther. *11*, 438. https://doi.org/10.1186/s13287-020-01941-y.

61. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381. https://doi.org/10.1126/science.aaw3381.

62. Song, D., Chen, S., Lee, C., Wang, C., Liu, P., Yang, Y., Li, K., Cen, Y., Yan, G., Wang, Q., et al. (2026). Synthetic Control Enables Reliable Cluster Validation and Marker Discovery in Omics Data: From Single-Cell and Spatial to Population-Scale. Preprint at bioRxiv. https://doi.org/10.1101/2023.07.21.550107.

63. Yang, L., Zhu, Y., Yu, H., Cheng, X., Chen, S., Chu, Y., Huang, H., Zhang, J., and Li, W. (2020). scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. Genome Biol. *21*, 19. https://doi.org/10.1186/s13059-020-1928-4.

64. Tian, R., Gachechiladze, M.A., Ludwig, C.H., Laurie, M.T., Hong, J.Y., Nathaniel, D., Prabhu, A.V., Fernandopulle, M.S., Patel, R., Abshari, M., et al. (2019). CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. Neuron *104*, 239–255.e12. https://doi.org/10.1016/j.neuron.2019.07.014.

65. Weiler, P., Lange, M., Klein, M., Pe'er, D., and Theis, F. (2024). CellRank 2: unified fate mapping in multiview single-cell data. Nat. Methods *21*, 1196–1205. https://doi.org/10.1038/s41592-024-02303-9.

66. Li, M., and Belmonte, J.C. (2017). Ground rules of the pluripotency gene regulatory network. Nat. Rev. Genet. *18*, 180–191. https://doi.org/10.1038/nrg.2016.156.

67. Ng, S.W.K., Xie, S.Z.J., Wagenblast, E., Takayama, N., Jin, L., Mbong, N., Murison, A., Erwin, S.E., Mitchell, A., Van Galen, P., et al. (2018). Identification of gene regulatory networks governing stemness properties of human HSC and LSC. Blood *132*, 3832. https://doi.org/10.1182/blood-2018-99-114314.

68. Mason, M.J., Fan, G., Plath, K., Zhou, Q., and Horvath, S. (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. BMC Genomics *10*, 327. https://doi.org/10.1186/1471-2164-10-327.

69. Jones, D.L., and Wagers, A.J. (2008). No place like home: anatomy and function of the stem cell niche. Nat. Rev. Mol. Cell Biol. *9*, 11–21. https://doi.org/10.1038/nrm2319.

70. Suda, T., Arai, F., and Shimmura, S. (2005). Regulation of stem cells in the niche. Cornea *24*, S12–S17. https://doi.org/10.1097/01.ico.0000178742.98716.65.

71. Ottone, C., Krusche, B., Whitby, A., Clements, M., Quadrato, G., Pitulescu, M.E., Adams, R.H., and Parrinello, S. (2014). Direct cell–cell contact with the vascular niche maintains quiescent neural stem cells. Nat. Cell Biol. *16*, 1045–1056. https://doi.org/10.1038/ncb3045.

72. Jadhav, U., Saxena, M., O'Neill, N.K., Saadatpour, A., Yuan, G.C., Herbert, Z., Murata, K., and Shivdasani, R.A. (2017). Dynamic reorganization of chromatin accessibility signatures during dedifferentiation of secretory precursors into Lgr5+ intestinal stem cells. Cell Stem Cell *21*, 65–77.e5. https://doi.org/10.1016/j.stem.2017.05.001.

73. Aughey, G.N., Estacio Gomez, A.E., Thomson, J., Yin, H., and Southall, T.D. (2018). CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. eLife *7*, e32341. https://doi.org/10.7554/eLife.32341.

74. Sijacic, P., Bajic, M., McKinney, E.C., Meagher, R.B., and Deal, R.B. (2018). Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. Plant J. *94*, 215–231. https://doi.org/10.1111/tpj.13882.

75. Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). m^6A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707–719. https://doi.org/10.1016/j.stem.2014.09.019.

76. Wang, Q., Zou, Y., Nowotschin, S., Kim, S.Y., Li, Q.V., Soh, C.L., Su, J., Zhang, C., Shu, W., Xi, Q., et al. (2017). The p53 family coordinates Wnt and nodal inputs in mesendodermal differentiation of embryonic stem cells. Cell Stem Cell. Preprint at bioRxiv *20*, 70–86. https://doi.org/10.1016/j.stem.2016.10.002.

Please cite this article in press as: Wang et al., Categorization and analysis of 16 computational methods for estimating cell potency from single-cell RNA-seq data, Cell Systems (2026), https://doi.org/10.1016/j.cels.2026.101762

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Cortical interneurons (C1)	CytoTRACE	Cortical_interneurons_C1 in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Dentate gyrus timepoints (10x; full)	CytoTRACE	Dentate_gyrus_timepoints_10x_All in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Dentate gyrus timepoints (10x; subset)	CytoTRACE	Dentate_gyrus_timepoints_10x in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Early zebrafish (Drop-seq; full)	CytoTRACE	Early_zebrafish_drop_seq_All in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Early zebrafish (Drop-seq; subset)	CytoTRACE	Early_zebrafish_drop_seq in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Endometrium (CEL-seq)	CytoTRACE	Endometrium_CEL_seq in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
in vitro NPCs (C1)	CytoTRACE	Invitro_NPCs_C1 in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Medial ganglionic eminence (C1)	CytoTRACE	Medial_ganglionic_eminence_C1 in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Neural stem cells (Drop-seq)	CytoTRACE	Neural_stem_cells_Dropseq in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Oligodendrocyte timepoints (C1)	CytoTRACE	Oligodendrocytes_timepoints_C1 in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Thymus (Drop-seq)	CytoTRACE	Thymus_Dropseq in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Whole planaria (Drop-seq)	CytoTRACE	Whole_planaria_Dropseq in data.zip at Zenodo: https://doi.org/10.5281/zenodo.22785227
Software and algorithms		
CytoTRACE (version 0.3.3)	CRAN	https://cytotrace.stanford.edu/
ORIGINS (version 0.1.0)	GitHub	https://github.com/danielasenraoka/ORIGINS
SCENT (version 1.0.3)	GitHub	https://github.com/aet21/SCENT
CCAT (Included in SCENT's package)	GitHub	https://github.com/aet21/SCENT
NCG	GitHub	https://github.com/Xinzhe-Ni/NCG
SPIDE	GitHub	https://github.com/CSUBioGroup/SPIDE
R (version 4.3.1)	R Project	https://www.r-project.org/
Other		
Analysis code	This paper	Zenodo: https://doi.org/10.5281/zenodo.22785227

METHOD DETAILS

PPI network perturbation analysis

To examine how much the information encoded in a specific PPI network contributes to cell-potency estimates, we evaluated five PPI-dependent methods using progressively perturbed versions of a reference PPI network while holding the scRNA-seq data constant. We used the *in vitro* neural progenitor cell (NPC) dataset from the CytoTRACE study, which comprises human pluripotent stem cells collected on days 0, 1, 3, 5, 7, 10, 14, and 30. Sampling day was used to define the ground-truth developmental ordering, with day 0 representing the highest potency and day 30 the lowest. The net17Jan16 network from the CCAT package was used as the reference PPI network. To generate perturbed networks, we randomly selected 10%, 20%, …, 100% of the network nodes and permuted their gene labels. This procedure disrupted the correspondence between gene identity and network position while preserving the global network topology. The unperturbed reference network was included as the 0% perturbation condition.

The reference and perturbed networks were supplied, together with the same NPC expression data, to ORIGINS, NCG, SCENT, CCAT, and SPIDE to estimate cell potency. The gene-gene similarity matrix required by NCG was obtained from the NCG GitHub

repository (https://github.com/Xinzhe-Ni/NCG). For each method and perturbation level, performance was quantified using the weighted Spearman rank correlation between estimated potency and the reverse-coded ground-truth developmental ordering (with earlier time points assigned higher values), with weights determined by the number of cells at each sampling time point. Higher correlation coefficients indicated better agreement with the known developmental progression.

Comparison of gene-count- and library-size-based gene selection in CytoTRACE

To assess whether cell library size carries information relevant to cell potency, we implemented a CytoTRACE variant in which the top 200 genes were selected based on the correlation of their expression with cell library size rather than with gene count, while all other steps remained unchanged. We evaluated this variant alongside the original CytoTRACE implementation using 12 single-lineage datasets from the CytoTRACE study.

For each dataset, the performance of each CytoTRACE implementation was quantified using the weighted Spearman rank correlation between estimated cell potency and the reverse-coded ground-truth developmental ordering (with earlier time points or more potent cell types assigned higher values). Weights were determined by the number of cells associated with each time point or cell type. Higher correlation coefficients indicated better agreement with the known developmental ordering.

Impact of normalization method on CCAT performance

Because CCAT does not have a default normalization method, we used it as an illustrative case to assess how normalization affects cell-potency estimation. The raw count matrix of the NPC dataset used in the PPI network perturbation analysis was independently normalized using four methods: Census, LogCP10K, LogCPM, and PFlogPF. Each normalized expression matrix was supplied to CCAT together with the reference PPI network provided in the CCAT package (net17Jan16). For each normalization method, performance was quantified using the weighted Spearman rank correlation between estimated cell potency and the reverse-coded ground-truth developmental ordering (with earlier time points assigned higher values). Weights were determined by the number of cells at each sampling time point. Higher correlation coefficients indicated better agreement with the known developmental progression.

Explanation of the difference between the geometric and arithmetic means in CytoTRACE

We understand the fact that the arithmetic mean of logarithmic expression levels is equal to the logarithm of the geometric mean of (unlogarithmic) expression levels, provided the logarithm is taken without a pseudocount. However, this equality does not hold in CytoTRACE's algorithm, where a pseudocount is used in the log transformation. Based on the following two quotes from CytoTRACE's publication,[23] it is reasonable to interpret that the geometric mean was calculated using logarithmic expression levels, with no mention of taking the logarithm after calculating the geometric mean.

Quote 1: "The resulting gene expression matrix X^* was log2-normalized with a pseudo-count of 1. We found that by using X^* to derive GCS, as opposed to the original matrix X, we could obtain a larger gain in performance over gene counts alone ($P = 0.006$ vs. 0.04, respectively; paired two-sided t-test; Figure S8C). This finding prompted us to apply this transformation as a pre-processing step for GCS and CytoTRACE throughout this work."

Quote 2: "Indeed, a gene counts signature (GCS), defined as the geometric mean of genes that are most correlated with gene counts, outperformed all evaluated features in the training cohort."

Based on Quote 2, it is reasonable to interpret that the GCS for each cell is the geometric mean of the selected genes, whose expression levels can be interpreted as logarithmic based on Quote 1. However, this interpretation is inconsistent with the code, which defines GCS as the arithmetic mean of the logarithmic expression levels with a pseudocount. Moreover, the correspondence between the geometric mean and arithmetic mean no longer holds when a pseudocount is introduced. Therefore, our reported difference is valid.