STEM CELL REPORTS
Perspective
ISSCR
OPEN ACCESS

A novel paradigm for single-cell annotation in stem cell research

Trevor Atkeson,[1,*] Sean B. Wilson,[2] Elias R. Ruiz-Morales,[3] Melissa H. Little,[2,4] Roser Vento-Tormo,[3,5] Drew Neavin,[1,7] and Joseph E. Powell[1,6,7]

[1]Translational Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW 2021, Australia
[2]Novo Nordisk Foundation Center for Stem Cell Medicine, University of Copenhagen, Copenhagen 2100, Denmark
[3]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK
[4]Novo Nordisk Foundation Center for Stem Cell Medicine, Murdoch Children's Research Institute, Parkville, VIC 3052, Australia
[5]Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
[6]UNSW Cellular Genomics Futures Institute, University of New South Wales, Kensington, Sydney, NSW 2052, Australia
[7]These authors contributed equally
*Correspondence: t.atkeson@garvan.org.au
https://doi.org/10.1016/j.stemcr.2025.102707

SUMMARY

Stem cell-derived models are a powerful tool for studying biology and are increasingly paving the way for cell-based therapies. However, understanding precisely which cell types are present and how closely they recapitulate *in vivo* cells remains challenging. Single-cell genomics, coupled with annotation methods, provides a framework for evaluating the congruence of stem cells with *in vivo* biology. Here, we explore approaches to cell annotation and discuss the challenges of implementing these methods in stem cell-derived models. We provide recommendations for the application of these methods, as well as our vision for the future of stem cell annotation using cell manifolds.

INTRODUCTION

Human pluripotent stem cells (hPSCs) have revolutionized the study of biology through their ability to model human cells in an *in vitro* experimental setting. For the purposes of this perspective, hPSCs encompass embryonic stem cells (Thomson et al., 1998) and induced pluripotent stem cells (iPSCs) (Zhu and Huangfu, 2013). hPSCs may be differentiated to recapitulate organ systems across the three germ layers, yielding ectodermal (e.g., brain and skin), mesodermal (e.g., heart, kidney, and blood), and endodermal (e.g., lung, intestine, and pancreas) tissue models (Biancotti et al., 2010; Haidet-Phillips et al., 2011; Nguyen et al., 2011). iPSCs may be reprogrammed from a donor's adult cells (Takahashi and Yamanaka, 2006), enabling patient-specific disease modeling. Because of this capacity, large-scale development and disease studies employing hPSCs are emerging, providing power for population genetics research and clinical trials in a dish (Carcamo-Orive et al., 2017; DeBoever et al., 2017; Kilpinen et al., 2017; Pashos et al., 2017; Warren et al., 2017). These factors make hPSCs an effective tool for *in vitro* biological research.

Organs consist of a diverse collection of cells, each with a unique molecular profile shaped by factors such as spatial location, developmental stage, environmental exposure, and disease state. A cell's molecular profile governs its identity and function, and no two cells are identical. hPSC models aim to replicate molecular profiles of cell types, tissues, and organ systems *in vitro* (Park et al., 2023; Revah et al., 2022; Shinozawa et al., 2021) by mimicking embryonic cell developmental processes. The process of cellular differentiation is dynamic, and as stem cells differentiate toward terminal cell type(s), they progress through a continuum of intermediate states with corresponding changes in their molecular profile, phenotype, and function. As knowledge of the factors that determine the progression of a cell down a given differentiation trajectory deepens, hPSC-derived models have improved physiological accuracy (Yu et al., 2007), better replicating increasingly specific cell types.

A cell's developmental pathway is determined by how it responds to microenvironment signals, and thus, imperfect differentiation protocols limit the developmental fidelity and subsequent biological relevance of hPSC models. By design, hPSC differentiations condense developmental processes into days or weeks to balance convenience, simplicity, and speed with physiological accuracy. Differentiations often deviate from natural biological pathways due to the challenges of recreating spatiotemporally complex microenvironments found across development *in vitro* (Kim et al., 2020a). An inappropriate cellular response to developmental signals (Yoshimura et al., 2023) can result in off-target or *in vitro*-specific by-products (Figure 1). Despite efforts by the stem cell community, differentiation of hPSC-derived cells often resembles prenatal embryonic tissues, hampering their use in modeling diseases found in mature contexts (Figure 1).

Challenges with replicating development *in vitro* are further complicated by the inter- and intra-line variability inherent to stem cells. Pluripotency and differentiation characteristics of a cell may be influenced by tissue origin, passage number, cell culture conditions, and unknown (stochastic) factors (Laco et al., 2018; Nguyen et al., 2018; Yamatani and Nakai, 2022) (Figure 1). Inter-line variability of pluripotency may limit the blanket application of

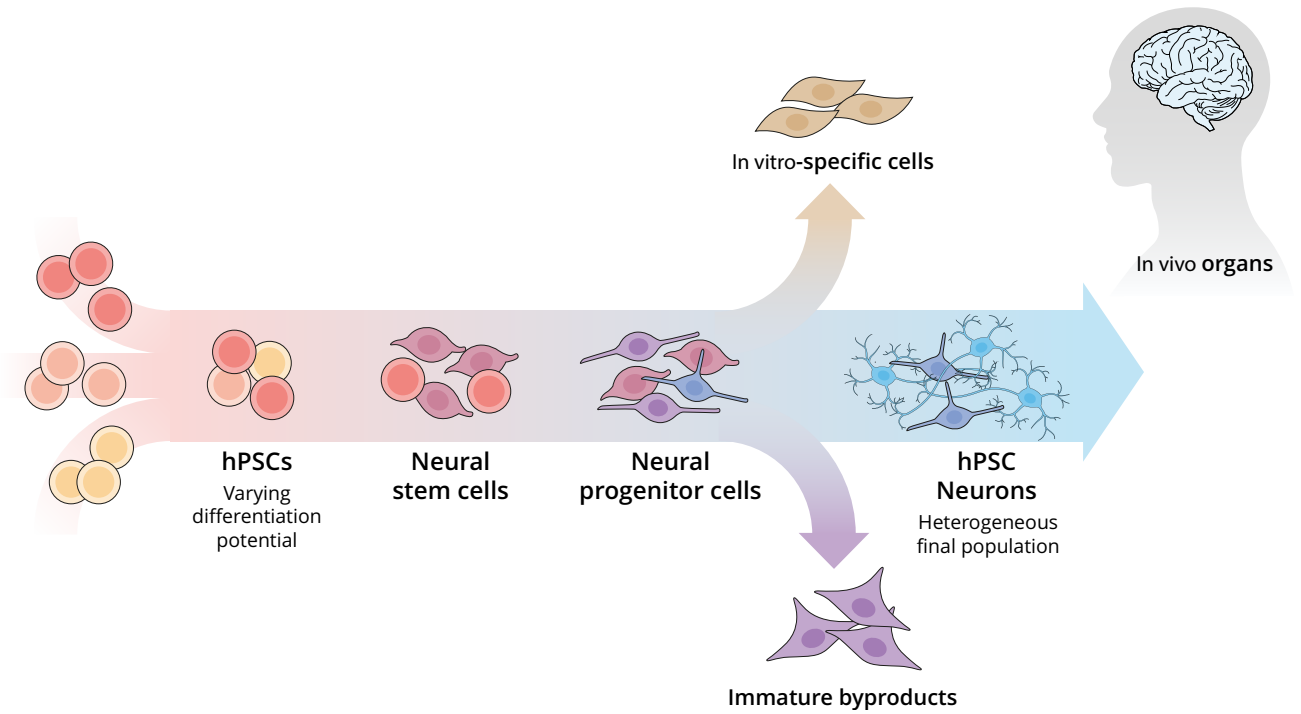


Figure 1. Heterogeneous cell populations across hPSC differentiation

Inter- and intra-line variability in hPSC pluripotency and differentiation properties lead to heterogeneous cell states as cells progress through differentiation at variable rates. Failure of some cells to reach the differentiation target population may yield immature by-products or *in vitro*-specific cell types. In almost all cases, final cell types fail to replicate the diversity, complexity, and maturity found in *in vivo* tissues.

protocols across large numbers of stem cell lines without optimization. Intra-line differences across cells cause differentiating hPSCs to develop at variable rates, resulting in diverse populations at mixed states of maturity (Figure 1).

Failure to fully replicate *in vivo* cell states in a dish limits the relevance and translational potential of stem cell systems. A deeper understanding of cell states throughout *in vivo* development and optimizing differentiations to better replicate these states is essential to improving the utility of stem cell models.

CELL-LEVEL ANNOTATION ENHANCES APPLICATIONS OF STEM CELL RESEARCH

Cell annotation—the process of assigning biological labels to cells—can be deployed to characterize the molecular identity, maturity, and physiological fidelity of hPSC-derived models. One common set of methods uses molecular features such as gene expression profiles, epigenetic modifications, or cell surface markers to distinguish cell identity. Alternatively, variations in features across a cell population can be used to annotate biological trajectories or cell relationships, capturing transitions between states over time and space. By analyzing changes in features, cells can be annotated across dynamic processes such as differentiation, disease progression, or response to perturbations. Here, we use annotation to refer to both of these approaches.

Annotation strategies vary in the number of molecular features and modality (gene expression, epigenetic markings, etc.) considered in determining cell type or state. Traditional cell-type annotation approaches, which measure transcriptomic or proteomic expression of a handful of cell type-specific markers (Table 1), have insufficient granularity to resolve continuous cell states (Figure 2) and are biased to our current biological knowledge. The development of single-cell technologies that capture molecular features at the individual cell level provide higher resolution and enable the detection of rare cell types, presenting an opportunity for more accurate annotation of hPSC models. These technologies include single-cell RNA sequencing to capture gene expression profiles (Tang et al., 2009), single-cell assay for transposase-accessible chromatin (ATAC) sequencing to interrogate chromatin accessibility (Buenrostro et al., 2015), single-cell DNA methylation and single-cell CUT&Tag for epigenetic marks (Bartosovic et al., 2021; Lorthongpanich et al., 2013), and

Table 1. Molecular features used in annotation of hPSC-derived cardiomyocytes

Objective	Annotation method	Annotation features	Examples
Cell identity annotation	Flow cytometry RT-qPCR Western blot	TNNT2, TNNI1, TNNI3, TBX5, NKX2.5 MYH6, MYL7, MHC, ACTN2	Balafkan et al., 2020; Lian et al., 2015; Osafune et al., 2008; Prondzynski et al., 2024; Wang et al., 2023
	Marker-based methods	Cluster markers TNNI1, TTN, ATP2A2, PLN, MYH7, MYL7, NKX2.5, TBX5	Cheng et al., 2023; Churko et al., 2018; Friedman et al., 2018; Grancharova et al., 2021; Wang et al., 2022
	Reference-based methods	Top 3,000–5,000 most variable genes	Litviňuková et al., 2020; Roost et al., 2015; Tyser et al., 2021; Weatherbee et al., 2023
		Reference datasets: Litviňuková et al. (2020) Roost et al. (2015) Tyser et al. (2021) Weatherbee et al. (2023) HCA: cells of the adult human heart HCA: sex-specific control of human heart maturation by the progesterone receptor	
Biological trajectory annotation	Pseudotime	Top 3,000–5,000 most variable genes	Friedman et al., 2018; Lam et al., 2020; Wang et al., 2022
	RNA velocity	Unspliced mRNA (introns) vs. spliced mRNA (exons)	Kannan et al., 2023; Litviňuková et al., 2020; Wang et al., 2022

HCA, Human Cell Atlas.
Selected molecular features for annotation of hPSC-derived cardiomyocytes deployed in various annotation methods.

cellular indexing of transcriptomes and epitopes sequencing, proximity extension assay, and ATAC with select antigen profiling by sequencing to evaluate protein expression (Lundberg et al., 2011; Mimitou et al., 2021; Stoeckius et al., 2017). More recently, "single-cell multiomic" assays have been developed, in which multiple genomic modalities are captured for a single cell (Dhainaut et al., 2022; Liang et al., 2023; Unterman et al., 2022). Multiomic technologies provide a detailed picture of cellular profiles along several cell state axes, uncover molecular interactions across different modalities, and improve identification of rare or previously undescribed cell populations.

As implementation of single-cell technologies increases, techniques to annotate cells are evolving as well, with a shift toward approaches that consider the continuum of cell types and states. For example, in annotation of stem cell-derived cardiomyocytes, more features are being used to account for continuous cell states (Table 1). Accurately placing cells within dynamic landscapes and the broader context of *in vivo* and *in vitro* molecular signatures is critical for utilizing stem cells for biological interrogation.

The broad adoption of stem cell models has resulted in the establishment of many differentiation protocols, resulting in inconsistent outcomes across laboratories and donor lines and making cross-protocol comparisons difficult. There is a need for the systematic assessment of a protocol's cell type/state proportions, maturity in relation to *in vivo* counterparts, and proportion of off-target by-products. Efforts by individual laboratories (Blümke et al., 2023; Cheng et al., 2024; Kim et al., 2020b; Lyra-Leite et al., 2022; Wilson et al., 2022) and larger initiatives (Ramos et al., 2021; Riley and Schekman, 2021) have recognized this need and begun encouraging cross-dataset comparisons and advocating for increased standardization.

Selecting an annotation method can be challenging due to the breadth of methods available and the fast-paced advancement of this field. Here, we review a range of annotation schemes and explore their utility when applied to stem cell models. We recommend strategies for improving annotation, including implementing a stem cell-relevant manifold to foster comprehensive, consensus-based annotation of hPSC-derived models and enhance their predictive ability.

FROM FLOW TO SINGLE-CELL GENOMICS: HOW SINGLE-CELL ANNOTATION METHODS HAVE BEEN APPLIED IN STEM CELL RESEARCH

There are now a large number of tools that may be used to annotate cells (Figure 2). Flow cytometry and marker- and reference-based methods classify a cell's type or state, while dynamic methods such as pseudotime and RNA

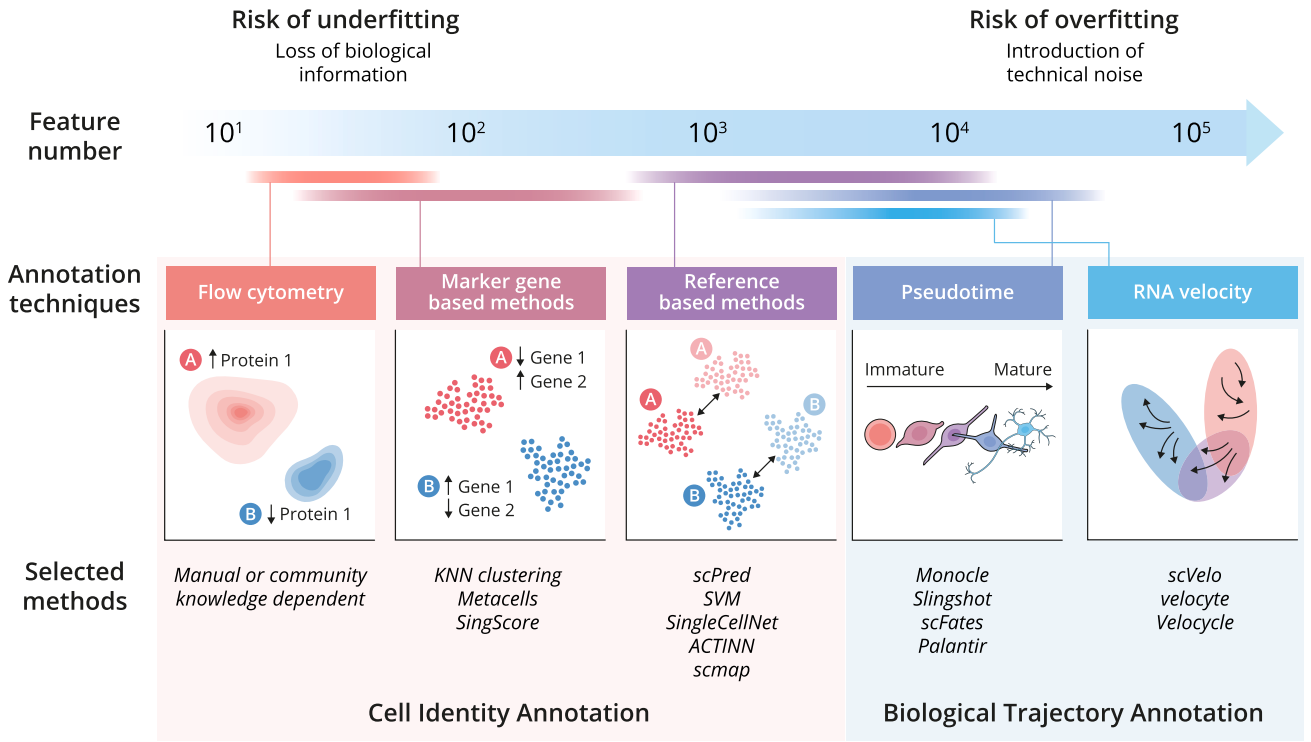


Figure 2. Tools for cell identity and biological trajectory annotation

Techniques are arranged according to the number of features used to capture biological variation. These methods rely on measuring similarities or differences between cells to assign annotations. Using too few features may obscure important biological signals, whereas using too many can introduce non-biological technical noise and increase the risk of overfitting.

velocity annotate a cell's location along a biological trajectory.

Flow cytometry-based annotation

Annotation using flow cytometry techniques typically label cell populations using single-cell protein measurements from a handful of molecular markers (Qiu et al., 2011). While flow cytometry is a cost-effective tool for rapid batch validation or isolation of cellular subsets, several barriers to accurate flow cytometry-based annotation remain. Technical limitations in fluorescence detection constrain the number of antibodies that can be used in parallel, and the limited number of proteins accessible for antibody binding restricts the capacity for complex cytometric analyses. The result of these factors is the annotation of cell types based on a handful of cellular features, masking underlying heterogeneity of cell states (Figure 2). Further, proteins marking specific cell states must be known *a priori*, limiting discovery work. While flow cytometry remains an important research tool that is useful in annotating distinct, well-characterized cell states, single-cell genomics methods now provide several advantages for interrogating cell identity in stem cell models.

Marker-based annotation

Marker-based annotation techniques typically use dimensional reduction (Box 1) to identify highly variable features, which are then used to group similar cells into clusters for annotation. All cells in a cluster are then annotated with the same cell type label based on the level of expression of known molecular features (Pullin and McCarthy, 2024) (Figure 2). These annotation methods are straightforward, computationally efficient, and synergistic with conventional annotation techniques. New cell type markers identified by techniques such as differential expression may then be applied to compare cell type identities across different datasets (Rothová et al., 2022) or annotate new cell types (Villani et al., 2017). Marker-based annotation may use predefined gene signatures (Foroutan et al., 2018) commonly employed in gene set enrichment analyses, which are particularly effective for datasets containing distinct cell types with well-established, high-confidence markers. However, when marker-based methods are applied to stem cell models, the number of independent cell types present across a differentiation is often unknown and biological conclusions may shift when a population is clustered at different resolutions yielding more or less cell type clusters (Alquicira-Hernandez et al., 2019;

Box 1. Dimensional reduction and visualization of single-cell data

In single-cell genomics, the high dimensionality of data, where thousands of molecular features are measured per cell, makes accurate and interpretable visualization difficult. Dimensional reduction techniques are often employed to identify feature combinations that capture variance structure in the data, resulting in a lower-dimensional representation known as a manifold (Kharchenko, 2021). A manifold is composed of multiple dimensions, each of which contain combinations of features that explain sources of variation in the dataset. A single dimension may capture features related to a combination of characteristics such as cell maturity, cell cycle state, differentiation stage, or technical factors such as batch effects (Cuomo et al., 2020). Cells that share similar molecular characteristics occupy comparable coordinates across the dimensions and are consequently close in the manifold. These coordinates can be used to visualize variation in the data typically using the top two dimensions. However, determining the specific collections of processes (biological or technical) represented by a dimension is often challenging and is now an active area of method development.

A common approach to exploring the structure of a manifold and understanding the relationships between cells is to graph cells in relation to their nearest genomic neighbors. Uniform manifold approximation and projection (UMAP) is a popular technique that applies dimensional reduction to visualize data in two- or three-dimensional space (Becht et al., 2018; van der Maaten and Hinton, 2008; McInnes et al., 2018). The generated visualizations can then be overlaid with information such as cell type labels derived from annotation methods. Visualization techniques such as UMAP are commonly used to ensure consistent data integration across batches, validate cell type clusters, and visualize changes in cell populations across different conditions (Chari and Pachter, 2023). However, the dimensional reduction employed by these techniques may introduce significant data distortion (Chari and Pachter, 2023). Additionally, visualizations may vary based on preprocessing (Chari and Pachter, 2023) and lose relevance in sparse or distantly related data (Becht et al., 2018; van der Maaten and Hinton, 2008). While two- or three-dimensional visualization techniques can be used to form hypotheses about the structure of the manifold and explore local relationships between cells, biological conclusions should be validated in the full high-dimensional dataset (Lause et al., 2024).

Garcia-Alonso et al., 2021; Haniffa et al., 2021; Kiselev et al., 2018; Ma and Pellegrini, 2020; Pedregosa et al., 2011; Tan and Cahan, 2019; Zhang et al., 2019). Tools such as clustering trees are increasingly used to help select an appropriate resolution and evaluate cell type relationships as the cluster size changes (La Manno et al., 2016). Additionally, marker-based methods do not account for the influence of donor variability, technical noise, and different microenvironments within a cluster (Burkhardt et al., 2021; Kang et al., 2018). Transitions between cellular states are often gradual in stem cell models, and overlapping patterns of molecular features can render cluster boundaries arbitrary (Adler et al., 2019; Burkhardt et al., 2021; Saunders et al., 2018). While marker-based tools are built on decades of knowledge regarding cell types (Patel et al., 2014; Ramsköld et al., 2012), these techniques may not yield consistent results across research groups (Abdelaal et al., 2019) and scale poorly to larger datasets (Ianevski et al., 2022; Zhang et al., 2019).

Reference-based annotation

Reference-based methods incorporate thousands of features when determining cell identity, better capturing the continuous molecular profiles of stem cell systems (Figure 2). These techniques project unannotated query data into an annotated reference manifold, where identity of the query cells is then predicted based on feature similarity to the reference (Abdelaal et al., 2019; Alquicira-Hernandez et al., 2019; Garcia-Alonso et al., 2021; Kiselev et al., 2018; La Manno et al., 2016; Ma and Pellegrini, 2020; Pedregosa et al.; Tan and Cahan, 2019) (example shown in Box 2). The cell type label from the reference data with the highest probability is assigned to the query cell. A key strength of reference-based methods is a quantitative per-cell confidence score (Salehin et al., 2025) that estimates the similarity of a query cell's features to cell type/state profiles from the reference. This is particularly helpful when annotating cells that share overlapping features across multiple cell states. Reference-based annotation is efficient and allows consistent annotation to be transferred across multiple query datasets (Box 2). However, these techniques rely upon an accurate, comprehensive, well-annotated reference relevant to each specific query (Haniffa et al., 2021, 2023). Misannotated references can transfer batch effects or bias between reference and query datasets, and query cell types absent in the reference may be misannotated or left unannotated.

While reference-based annotation traditionally relies on small, well-curated reference data, the growth of AI has precipitated in the development of powerful models trained on massive and diverse single-cell datasets. After pre-training, these models can be fine-tuned for different research tasks, including cell annotation (Aminzadeh et al., 2024; Cui et al., 2024, 2024; De Donno et al., 2023;

Box 2. Annotation of hPSC-derived kidney organoids using a reference-based method

Here, we illustrate how a reference based annotation tool (Wilson et al., 2022) can be employed to annotate hPSC-derived kidney organoids and compare cell type proportions across differentiation protocols. Kidney development is complex, and cell patterning is driven by reciprocal signaling and self-organization (Wilson et al., 2022). Intermediate cell types in the developing tissue often exhibit overlapping transcriptional signatures (Combes et al., 2019a, 2019b; Lindström et al., 2018; Ransick et al., 2019), confounding annotation of these cells. Kidney organoids modeling this development vary significantly in structure and maturity across differentiation batches and protocols, with corresponding changes in cell type composition and gene expression (Little and Combes, 2019; Phipson et al., 2019; Wilson et al., 2022), necessitating an unbiased tool to annotate and compare organoid cell products. Using *in vivo*-derived cell-type gene signatures, DevKidCC annotates query kidney organoid datasets using a hierarchical framework in which cells are annotated at increasingly fine resolutions (Figure 3A). Even with a few cells in a population, accurate annotation of similar but distinct cell types is achieved using this method, which is not possible with marker-based annotation. For example, the tool was able to distinguish a small population (0.5% of total cells) of ureteric epithelium cells (Figure 3B). While sharing a similar transcriptomic profile with distal nephron epithelium cells, ureteric epithelium cells arise from distinct precursors (Combes et al., 2019a, 2019b; Howden et al., 2021; Lindström et al., 2018; Ransick et al., 2019). DevKidCC enables consistent comparison of organoids across differentiation batches and protocols, leading to an understanding of the key differences in nephron patterning between kidney organoid models across the field (Figure 3C). While DevKidCC is trained for annotation of kidney organoid datasets, its hierarchical annotation framework can be applied to other tissue systems (Wilson et al., 2022) with customized tiers of subclassification where sufficient transcriptomic data are available.

Hao et al., 2024; Heimberg et al., 2023; Hou and Ji, 2024; Lotfollahi et al., 2019; Pearce et al., 2025; Theodoris et al., 2023; Yang et al., 2022). AI models can generalize across a range of biological questions such as predicting responses to drug and disease perturbations (Zhang et al., 2025), discovering gene regulatory networks (Fu et al., 2025; Pearce et al., 2025; Rosen et al., 2023), identifying candidate therapeutic targets (Combes et al., 2019a; Deshpande et al., 2023; Lindström et al., 2018; Ransick et al., 2019; Wilson et al., 2022), and improving differentiation protocols (Kong et al., 2022). They can integrate datasets collected from diverse biological and technical conditions (Boiarsky et al., 2023) to form composite datasets encompassing multiple tissues, patients, and disease populations (Alquicira-Hernandez et al., 2019; Deshpande et al., 2023), enabling more accurate annotation across diverse research scenarios. While AI models already incorporate genomic data from tens of millions of cells, they have largely ignored using data from hPSC systems, limiting their predictive power in these areas. As more single-cell hPSC data become available, AI models will achieve broader coverage of possible cell states and greater power to uncover the molecular patterns that define cell identity. This, in turn, will enable the imputation of unmeasured cell profiles (Baruzzo et al., 2020; Luo et al., 2024; Marouf et al., 2020) or modalities (Hu et al., 2023; Liu et al., 2025; Song et al., 2024).

Biological trajectory annotation

Cellular genomic measurements are a snapshot of a cell's molecular profile. For cells that change states over time (e.g., in a differentiation), genomic measurements provide incomplete coverage of continuous, transitioning cell states across a biological trajectory. Pseudotime annotation approaches (Cannoodt et al., 2016; Street et al., 2018; Wolf et al., 2019) organize cells along an inferred pseudo-temporal trajectory based on molecular feature similarities, aiming to capture their progression along a biological pathway (Charrout et al., 2020). The pseudo-temporal order does not correspond to real time but is determined from the continuum of feature changes across cell states. It is important to remember that pseudotime approaches do not typically annotate cells with classical, categorical cell-type labels, but instead provide a continuous cell state metric. In some instances, alternative annotation labels can be overlaid with the pseudo-temporal position or the coordinates of a cell in state space can be used as indirect annotations.

Pseudotime annotation approaches help characterize dynamic cell-state processes such as differentiation (Street et al., 2018; Sumanaweera et al., 2023), activation, aging, or the cell cycle (Saelens et al., 2019). These annotation tools are important in hPSC experiments, as individual cells in a differentiation transit across continuous state changes over varying time scales, resulting in heterogeneous cell populations (Figure 1). Pseudotime approaches have been deployed in stem cell models to identify developmental checkpoints (Fischer et al., 2019) and uncover previously unidentified gene-regulatory networks driving differentiation (Cuomo et al., 2020; Maulding et al., 2024; Trapnell et al., 2014).

The choice of pseudotime tool depends on prior knowledge of the sampled cells and the biological system, such

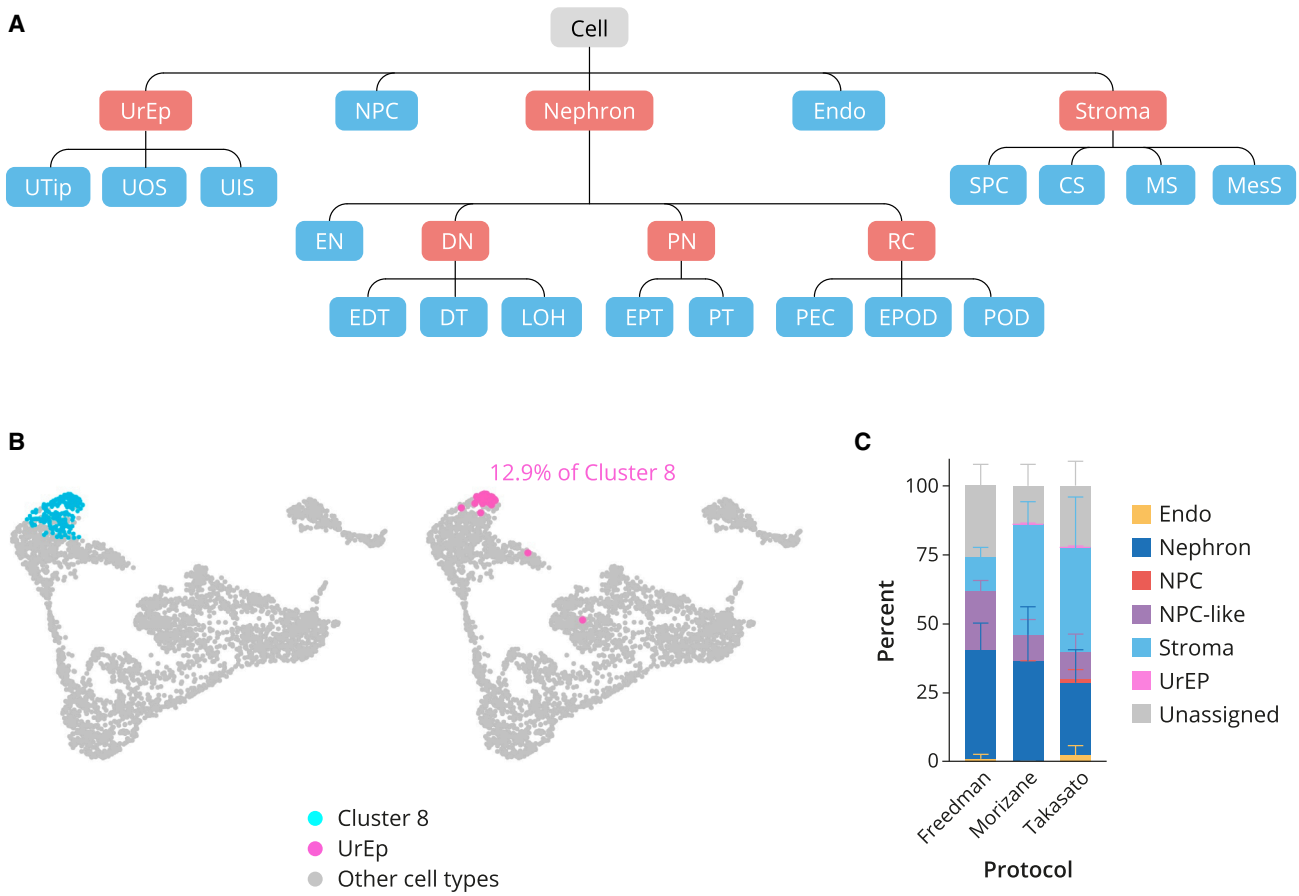


Figure 3. Hierarchical reference-based cell type annotation using DevKidCC

(A) Lineage tree of cell types used to annotate hPSC-derived kidney organoids. Red boxes represent intermediate annotations that may be annotated at finer cell type resolution, while blue boxes represent annotations that are not further subdivided.

(B) DevKidCC successfully identified a small population of ureteric epithelium cells (pink) that would be misannotated as distal nephron epithelium cells (teal) using marker-based annotation techniques even after sub-clustering only the epithelial populations. Data from Lawlor et al. (2021).

(C) Annotation with DevKidCC demonstrates cell type proportion differences across three commonly used hPSC-derived kidney organoid protocols (Freedman et al., 2015; Morizane et al., 2015; Takasato et al., 2015). Error bars represent ±1 standard deviation.

as known starting points, pre-annotated cell types, or trajectory shape and complexity. Pseudotime trajectories range from simple (linear, cyclical, and bifurcating) to complex (connected and disconnected graphs) (Saelens et al., 2019). Effective pseudotime analysis requires single-cell datasets with sufficient transient states (Fang et al., 2024) and often involves trialing multiple methods to minimize algorithm bias (Charrout et al., 2020) and validating trajectory (Saelens et al., 2019) shape using approaches that rely on fewer assumptions. Historically, pseudotime methods have not accounted for cross-biological sample variation, limiting comparisons across conditions and increasing sample-specific false discovery rates. However, emerging tools now address inter-individual variability (Hou et al., 2023) and provide uncertainty metrics (Campbell and Yau, 2019), improving interpretability and confidence.

While pseudotime approaches place cells along an inferred biological trajectory, RNA velocity techniques predict a cell's near-future transcriptional state on the sub-hour scale (Gorin et al., 2022) (Figure 2). By quantifying accumulation or depletion of nascent unspliced RNA, these tools determine the direction of rapid transcriptional changes in processes such as embryogenesis or differentiation (Bergen et al., 2020; La Manno et al., 2018; Lederer et al., 2024). Pseudotime and RNA velocity tools are complementary, where pseudotime reconstructs global cell state trajectories, while RNA velocity resolves directionality and short-term dynamics of cells. RNA velocity can be used to infer root and final cell states in a differentiation and forecast unsampled

intermediate populations (Lange et al., 2022) or future cell fates (La Manno et al., 2018). RNA velocity is particularly useful where single-cell sampling of shorter time frames is impossible, especially in cases where rapid cell state changes are difficult to predict and measure (Li et al., 2024). When applied to stem cell models, RNA velocity has been deployed to detect directionality in complex asynchronous differentiations (Setty et al., 2019) and decipher gene expression dynamics across cell types in organoids (Antón-Bolaños et al., 2024). However, RNA velocity models may lack in reproducibility (Gorin et al., 2022) or assume constant transcription rates for a given state (Bergen et al., 2021; Setty et al., 2019), neglecting genes undergoing multiple rate kinetics across a stem cell differentiation (Barile et al., 2021; Battich et al., 2015; La Manno et al., 2018) and clouding analysis of trajectory dynamics.

A wide range of tools and approaches are available to hPSC researchers to annotate cells, as types, states, or across biological trajectories. Selecting an annotation approach should be dependent on the particular biological question.

EVOLVING CELL MANIFOLDS AND ANNOTATION TECHNIQUES TO IMPROVE BIOLOGICAL INTERPRETATION OF STEM CELL MODELS

Challenges in annotation specific to stem cells and practical strategies for improvement

Current annotation approaches have limitations when applied to hPSC models and in many instances can lead to misannotation of cells or loss of key information. There are a few important factors to consider when thinking about improving annotation frameworks for hPSC models. First, some cells in a differentiation will stochastically vary from a protocol's desired cell type. Therefore, annotation methods need to account for a wide variety of potential cell types/states to avoid misannotating these by-products. Second, identifying how closely hPSCs resemble *in vivo* counterparts is important, but annotation techniques must be applicable across and between a wide range of potentially heterogeneous differentiation protocols. Finally, dynamic changes in cell types/states define most hPSC models. Therefore, methods that accurately quantify this variation at the level of cells will improve downstream applications of annotated cells.

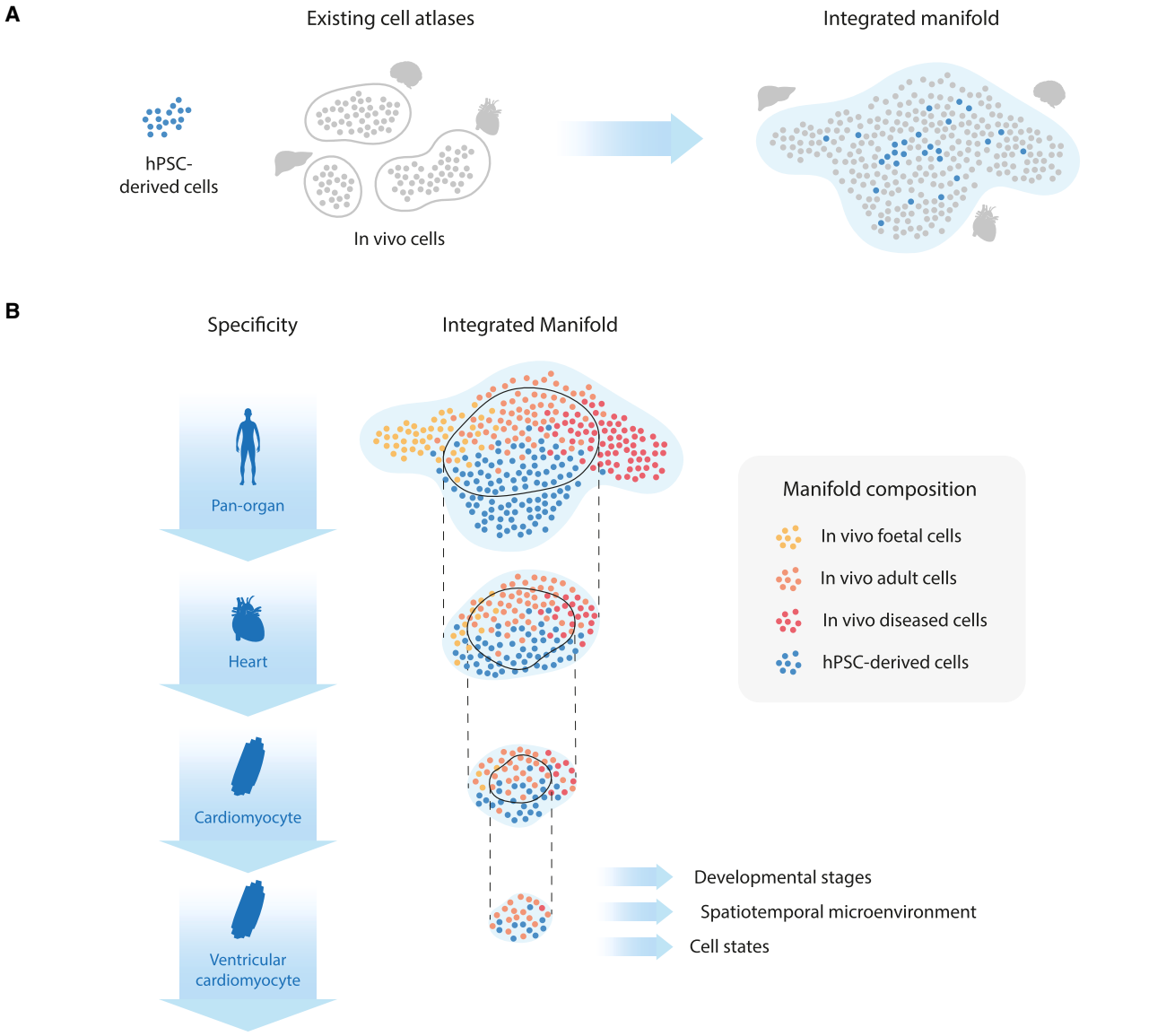
Improving annotation frameworks for hPSC models faces several key challenges. Characterization of the molecular features comprising cell identity remains incomplete, and the field lacks consensus cell state markers to use for annotation. hPSC-derived cells are immature and may contain combinations of molecular features not found *in vivo*. Inter- and intra-line variability of pluripotency leads stem cells to progress through development at variable rates depending on a range of internal and external conditions, making comparisons between laboratories and protocols challenging. A scarcity of well-annotated reference datasets across development limits the application of reference-based methods.

Given the specific challenges in stem cell annotation outlined, we have developed a set of recommendations to enhance the accuracy of annotating these models. First, we recommend expanding *in vivo* sample collection across development, disease states, and genetic backgrounds. Second, we suggest optimizing annotation methods to better capture biological diversity, minimize technical noise, and establish standardized, quantitative benchmarking for annotation performance in hPSC systems. Third, we propose integrating stem cells and their derivatives into existing cell atlases, expanding single-cell multiomic dataset collection, and enhancing cross-modality imputation techniques. Last, we advocate for the development of infrastructure and tools to promote data sharing across research groups and incentivize the use of community-driven, biologically relevant annotation techniques. Collectively, we believe that achieving these goals would improve the accuracy and reliability of stem cell annotation, refining the biological fidelity of hPSC models.

A comprehensive stem cell-relevant manifold for annotation provides the foundation for effective annotation of hPSCs

While sizable single-cell atlases exist (Cao et al., 2017; 2019; Chen et al., 2024; Han et al., 2018; Hawrylycz et al., 2012; Plass et al., 2018; Regev et al., 2017; Schaum et al., 2018), these platforms currently do not contain sufficient numbers of *in vivo* samples across developmental time points or stem cells and their derivatives (Heimberg et al., 2023) from across the stem cell field (He et al., 2023; Xu et al., 2023) (Figure 4A). Expanding existing atlases to include wider and deeper coverage of these two data types will help bridge the biological gap between *in vivo* and hPSC-derived cells.

Our recommendations support the creation of a stem cell-relevant manifold that encompasses cell states observed both *in vivo* and *in vitro*. This manifold would include samples spanning diverse disease groups, genetic ancestries, and developmental stages, as well as hPSCs and their derivatives from multiple laboratories, lines, and protocols (Figure 4B). We envision a pan-organ structure that encompasses all tissues, enabling detection of cells differentiating toward adjacent or alternative fates and providing insights to refine differentiation protocols. *In vitro* artifacts will also be represented but clearly labeled as off-target populations to facilitate their identification in query samples. To maximize utility, the manifold should be organized hierarchically, from broad organ-level scale to



(legend on next page)

specific cell types and states, enabling annotation at multiple resolutions (Figure 4B). Construction would start with individual organs and expand continually and incrementally as new datasets are generated by the field. Because contributing samples will vary in cell numbers and annotation resolution, standardization of cell types/states labels will be critical to ensure consistency. Integrating this extensive corpus will require comprehensive biological (e.g., donor, disease state, and age) and technical metadata (e.g., methods of cell collection, generation, differentiation, and sequencing) (Haghverdi et al., 2018; Stuart et al., 2019). While annotation techniques could leverage the entire manifold, researchers may also select organs or datasets of interest for annotation of query samples—following approaches similar to those used in the CZI CELLxGENE Census (Program et al., 2023).

Application of a stem cell-relevant manifold will enhance stem cell research

By integrating diverse populations of input cells across cell types, tissues, developmental stages, disease states, intersex populations, and genetic ancestries, the manifold would establish a broader consensus of cell identity than any single dataset alone (Figure 4A). This increases the likelihood of accurately annotating cell types observed only under specific conditions. Europeans remain overrepresented in genomics studies (Ghosh et al., 2022; Nguyen et al., 2023), and undersampled populations present vast wells of unexplored genetic ancestral diversity (The H3Africa Consortium et al., 2014) that stem cell researchers could leverage (Azeez et al., 2022; Choudhury et al., 2020; Lek et al., 2016). Generating a manifold from multiple individuals would allow the consequences of allelic variation to be captured and modeled by the manifold. Although its initial aim is not to study genetic regulation, expanding the manifold's scope and depth would facilitate such analyses. However, two challenges remain: obtaining genotype data from single-cell samples and accounting for compounded variation introduced by different protocols, researchers, and laboratories, which could hinder genetic analyses. To enhance functionality, the manifold should contain metadata for self-reported ethnicity, sex, tissue, disease state, or source publication, as in approaches like CELLxGENE (Program et al., 2023).

The manifold would serve as a flexible framework for a range of annotation techniques and applications (Figure 4C). For example, assessing a drug's impact on a biological pathway would involve selecting and annotating features specific to that pathway, while investigating genetic influences on cell fate commitment would require annotation using features associated with differentiation and development. Its broad scope could provide precise cell state signatures for marker-based annotation, serve as a comprehensive resource for reference-based methods, and act as training data for AI approaches. Critically, researchers could also leverage combinations of manifold datasets to benchmark annotation tools.

Beyond annotation, the manifold would inform strategies to enhance differentiation maturity and fidelity and encourage the selection of optimal protocols given experimental requirements. For instance, comparing cell composition and maturity between brain organoids and *in vivo* brain tissue could be used to track adherence to physiological development and reveal discrepancies in gene expression or cellular organization (Figure 4C). Alignment of pseudotemporal datasets using dynamic time-warping (Cacchiarelli et al., 2018; Sugihara et al., 2022) or other trajectory tools (Alpert et al., 2018; Kong et al., 2022; Laidlaw et al., 2024; Sumanaweera et al., 2023) would contexualize the identity and maturity of a query sample, informing iterative refinement of differentiation protocols (Kong et al., 2022) to better recapitulate *in vivo* environments. Trajectory confidence scores (Campbell and Yau, 2019) could further determine how faithfully differentiations replicate expected biological trajectories. However, as illustrated by DevKidCC, reliable annotation with hPSC-specific references depends on ensuring that the manifold adequately represents the full range of protocols used across the field.

By comparing hPSC differentiations with one another, the manifold could benchmark protocols or pinpoint sources of technical variation driving differences between laboratories, donors, and scientists using the same protocol (Neavin et al., 2023). Direct comparisons of all protocols would reveal which produce rare cell types or off-target by-products. These insights would guide the selection of protocols that yield the most biologically relevant cell types/states for the research question. Comparing population-level target cell yields across hPSC lines could uncover pluripotency gene programs driving successful

Figure 4. Generation of a stem cell-relevant manifold

(A) Existing atlases contain small numbers of hPSC-derived cells and are substandard in annotating these cell types. Our proposed manifold would contain hPSCs combined with *in vivo* datasets for more effective annotation and downstream analyses of stem cell models.

(B) By integrating single-cell datasets collected from *in vivo* and hPSC-derived cells, a hierarchical stem cell-relevant manifold could be created to annotate increasingly specific organs, cell types, or cell states.

(C) Benefits of this stem cell-relevant manifold include consensus-based query annotation, annotation tool flexibility, cell type identity context, and multiomic imputation.

differentiation and identify lines suitable to be used across laboratories (Ellis et al., 2025; Pantazis et al., 2022) for direct comparisons of the same line.

The manifold should integrate multiple genomic modalities to provide complementary viewpoints of genomic regulation, enabling identification of patterns across molecular feature types (Figure 4C). By drawing connections across analogous cells, the manifold should enable visualization of equivalent multiomic data and the imputation of unmeasured modalities from transcriptomic profiles. Notably, while bulk RNA sequencing approaches offer lower resolution, they have been widely applied. The manifold can aid in deconvoluting bulk sequencing data (Cho et al., 2024; Chu et al., 2022; Tsoucas et al., 2019; Wang et al., 2019) by estimating cell type proportions and their corresponding gene expression profiles at single-cell resolution. Given the manifold's size and comprehensiveness, this can be done more precisely and in a context-specific manner.

Finally, while the manifold represents the forefront of our understanding of stem cells and their differentiation mechanisms, caution should be taken when applying it to clinical decision making or human health. Instead, the manifold could be used to select and hone differentiation protocols to use as cell therapies. Pursuing these treatments requires two complementary approaches. Quantitatively, the manifold can be used to refine differentiation protocols to maximize similarity between *in vitro* and *in vivo* cell states (Kong et al., 2022) to ensure stem cell-derived therapies perform as intended. Qualitatively, decisions to pursue stem cell-based treatments depend on treatment context, including a patient's disease and its severity, injection site, and the potential risks and goals of treatment. Together, these approaches can inform when and how to deploy stem cell therapies while highlighting the improvements that are required before clinical translation. Early efforts, such as dopaminergic neuronal grafts (Pavan et al., 2025) for Parkinson disease, illustrate the potential of these treatments but underscore disease context-dependent questions such as immune cloaking, long-term graft survival, and controlled proliferation. As emphasized above, the manifold should be an iterative, community-built resource, with predictive power and clinical relevance increasing as coverage of cell states grows. Supporting efforts by atlases such as the Human Cell Atlas to increase the collection of diseased and developmental samples would further inform downstream stem cell applications.

Improving single-cell annotation methods will support key developments in the stem cell field. Better annotations of hPSC-derived cells can guide the refinement of differentiation protocols, aligning *in vitro* models more closely with *in vivo* states. Continued development and broader adoption of annotation techniques incorporating thousands of features will foster consensus on cell identity and reduce cross-laboratory reproducibility challenges (Volpato et al., 2018). Finally, improved annotation will inform a more comprehensive understanding of cell relationships and the mechanisms driving development or dysregulation (MacLean et al., 2018).

Challenges in development of a stem cell-relevant manifold

The single-cell community has laid the foundation for developing a comprehensive stem cell-relevant manifold by ensuring that data are accessible, coordinated, and well documented. Nevertheless, some potential challenges remain. The scale of samples for which data must be generated is substantial, and the collection, processing, and analysis of these data are a time-intensive and expensive undertaking. Historically, single-cell sampling has been biased toward cells that are easy to collect and process, limiting the downstream annotation accuracy of large (Hegenbarth et al., 2022), rare, or fragile cells (Miranda et al., 2023). However, advancements in sample collection protocols for challenging tissues such as the brain and heart along with the continued improvement of single-cell technologies would ameliorate some of these challenges moving forward. While synthetic data (Crowell et al., 2023; Treppner et al., 2021) can help address gaps in the manifold, its use should be approached with caution.

Single-cell datasets are increasing in size and complexity, necessitating the development of integration methods (Argelaguet et al., 2021) that address complex, nonlinear, nested batch effects (Kopp et al., 2022; Luecken et al., 2022) while preserving biological variation. These methods must extend beyond cell types to integrate pseudotime metrics (Sugihara et al., 2022) and cell cycle variation, and must be both scalable and practical (Luecken et al., 2022) to implement. Although models incorporating prior cell type information best preserve biological variation (Andreatta et al., 2024; Luecken et al., 2022; Sikkema et al., 2023), hPSC cell identity is often unknown or includes off-target cell types. Quantitative trajectory or cluster conservation scores will, therefore, be useful to ensure proper integration. As AI-based integration methods (Aminzadeh et al., 2024; De Donno et al., 2023; Lopez et al., 2018; Lotfollahi et al., 2019; Xu et al., 2021) become more prevalent, they are expected to surpass traditional methods in both scale and accuracy, enabling the integration of hundreds of datasets at low computational costs and even across species (Pearce et al., 2025; Tarashansky et al., 2021). However, AI-based methods require ongoing monitoring and refinement. In addition to integrating new datasets, integration with existing atlases will be

critical to identify and mitigate gaps in represented cell types/states.

As the manifold expands, efficient computational processing would be critical and annotation must balance run time with accuracy. Established single-cell repositories (Program et al., 2023; Rood et al., 2025) and annotation models (Heimberg et al., 2023; Program et al., 2023) provide scalable examples for optimizing annotation efficiency while managing computational demands. Comprehensive metadata should accompany the manifold, covering information such as disease status, sample collection and processing, spatial location, and donor ethnicity, with personal identifiers anonymized. The immaturity of single-cell proteomic, metabolomic, and other emerging genomic technologies hampers the expansion of multiomic modalities. Additionally, the sparsity of datasets from these costly or emerging techniques may limit the ability to draw connections across modalities.

Finally, while we have presented challenges in manifold development and integration, researchers conducting single-cell experiments must take into account and minimize factors that can confound downstream biological annotation. Using multiple lines, or commonly shared references such as H9 (WiCell Research Institute) or KOLF2.1J (Jackson Laboratories), helps mitigate line-specific effects such as passage number, donor sex, age at donation, and tissue of origin. Thorough metadata collection and sufficient coverage of potential confounders are essential to reliably detect and correct their effects on annotation. In parallel, experiment-specific sources of variation like cell type-specific loss during processing, reagent lot differences, or the researcher handling samples should be minimized across cell culture, differentiation, and single-cell workflows. Careful control of these factors ensures that the full spectrum of cell-type variation is captured and that downstream biological discoveries are not distorted by confounding effects.

Conclusions

While the manifold concept described above is ambitious in size and scope, if realized, it would serve as a valuable shared annotation resource for the stem cell field. A single, comprehensive manifold generated from *in vivo* and *in vitro* cells would enhance annotation accuracy and improve the fidelity of hPSC differentiations. It would also deepen the understanding of intermediate states across development (MacLean et al., 2018) and disease progression, and improve the predictive power of AI tools. Moreover, understanding of dysregulation yielded by more accurate annotation would help identify genetic targets and guide drug development.

Comprehensive annotation of hPSC cell states will improve our ability to replicate developmental and disease processes *in vitro*, advancing insights into cell function. As we highlight in this perspective, single-cell annotation of stem cell systems remains a challenge. However, implementation of the described stem cell-relevant manifold would improve the annotation of hPSC-derived cells, refine differentiation protocols, broaden understanding of cell states, and uncover mechanisms underlying cellular development and dysregulation. These advancements will accelerate transformative improvements in stem cell models, uncovering the molecular mechanisms that drive development and disease.

ACKNOWLEDGMENTS

J.E.P is supported by the National Health and Medical Research Council Fellowship (Australia, APP1107599) and with the support of the Fok Family in memory of Dr. and Mrs. Wing Kan Fok. S.B.W. and M.H.L. are supported by the Novo Nordisk Foundation Center for Stem Cell Medicine (Denmark, NNF21CC0073729). R.V-T. is supported by Wellcome Trust Grant (United Kingdom, 220540/Z/20/A). E.R.R-M. is supported by the IBSA Foundation Fellowship (Switzerland). Figures were generated with the assistance of BioRender.

AUTHOR CONTRIBUTIONS

T.A. led writing, discussion of conceptual frameworks. S.B.W., E.R.R-M., M.H.L., and R.V-T. contributed to discussion and edited the manuscript. D.R.N. and J.E.P. led conceptual thinking, contributed to discussion, and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M.J.T., and Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol. *20*, 194. https://doi.org/10.1186/s13059-019-1795-z.

Adler, M., Korem Kohanim, Y., Tendler, A., Mayo, A., and Alon, U. (2019). Continuum of Gene-Expression Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. Cell Syst. *8*, 43–52.e5. https://doi.org/10.1016/j.cels.2018.12.008.

Alpert, A., Moore, L.S., Dubovik, T., and Shen-Orr, S.S. (2018). Alignment of single-cell trajectories to compare cellular expression dynamics. Nat. Methods *15*, 267–270. https://doi.org/10.1038/nmeth.4628.

Alquicira-Hernandez, J., Sathe, A., Ji, H.P., Nguyen, Q., and Powell, J.E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biol. *20*, 264. https://doi.org/10.1186/s13059-019-1862-5.

Aminzadeh, F., Wu, J., He, J., Saberi, M., and Vafaee, F. (2024). Single-Cell Data Integration and Cell Type Annotation through

Contrastive Adversarial Open-set Domain Adaptation. Preprint at bioRxiv. https://doi.org/10.1101/2024.10.04.616599.

Andreatta, M., Hérault, L., Gueguen, P., Gfeller, D., Berenstein, A.J., and Carmona, S.J. (2024). Semi-supervised integration of single--cell transcriptomics data. Nat. Commun. *15*, 872. https://doi.org/10.1038/s41467-024-45240-z.

Antón-Bolaños, N., Faravelli, I., Faits, T., Andreadis, S., Kastli, R., Trattaro, S., Adiconis, X., Wei, A., Sampath Kumar, A., Di Bella, D.J., et al. (2024). Brain Chimeroids reveal individual susceptibility to neurotoxic triggers. Nature, 1–8. https://doi.org/10.1038/s41586-024-07578-8.

Argelaguet, R., Cuomo, A.S.E., Stegle, O., and Marioni, J.C. (2021). Computational principles and challenges in single-cell data integration. Nat. Biotechnol. *39*, 1202–1215. https://doi.org/10.1038/s41587-021-00895-7.

Azeez, I.A., Awogbindin, I.O., Olayinka, J.N., Folarin, R.O., Adamu, A.S., Ior, L.D., Shehu, A.M., Mukhtar, A.I., Ajeigbe, O.F., Emokpae, A.O., et al. (2022). Neural stem cell research in Africa: current realities and future prospects. Biol. Open *11*, bio059574. https://doi.org/10.1242/bio.059574.

Balafkan, N., Mostafavi, S., Schubert, M., Siller, R., Liang, K.X., Sullivan, G., and Bindoff, L.A. (2020). A method for differentiating human induced pluripotent stem cells toward functional cardiomyocytes in 96-well microplates. Sci. Rep. *10*, 18498. https://doi.org/10.1038/s41598-020-73656-2.

Barile, M., Imaz-Rosshandler, I., Inzani, I., Ghazanfar, S., Nichols, J., Marioni, J.C., Guibentif, C., and Göttgens, B. (2021). Coordinated changes in gene expression kinetics underlie both mouse and human erythroid maturation. Genome Biol. *22*, 197. https://doi.org/10.1186/s13059-021-02414-y.

Bartosovic, M., Kabbe, M., and Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. Nat. Biotechnol. *39*, 825–835. https://doi.org/10.1038/s41587-021-00869-9.

Baruzzo, G., Patuzzi, I., and Di Camillo, B. (2020). SPARSim single cell: a count data simulator for scRNA-seq data. Bioinformatics *36*, 1468–1475. https://doi.org/10.1093/bioinformatics/btz752.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. Cell *163*, 1596–1610. https://doi.org/10.1016/j.cell.2015.11.018.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. https://doi.org/10.1038/nbt.4314.

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.

Bergen, V., Soldatov, R.A., Kharchenko, P.V., and Theis, F.J. (2021). RNA velocity—current challenges and future perspectives. Mol. Syst. Biol. *17*, e10282. https://doi.org/10.15252/msb.202110282.

Biancotti, J.-C., Narwani, K., Buehler, N., Mandefro, B., Golan-Lev, T., Yanuka, O., Clark, A., Hill, D., Benvenisty, N., and Lavon, N. (2010). Human Embryonic Stem Cells as Models for Aneuploid Chromosomal Syndromes. Stem Cell. *28*, 1530–1540. https://doi.org/10.1002/stem.483.

Blümke, A., Ijeoma, E., Simon, J., Wellington, R., Purwaningrum, M., Doulatov, S., Leber, E., Scatena, M., and Giachelli, C.M. (2023). Comparison of osteoclast differentiation protocols from human induced pluripotent stem cells of different tissue origins. Stem Cell Res. Ther. *14*, 319. https://doi.org/10.1186/s13287-023-03547-6.

Boiarsky, R., Singh, N., Buendia, A., Getz, G., and Sontag, D. (2023). A Deep Dive into Single-Cell RNA Sequencing Foundation Models. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.19.563100.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490. https://doi.org/10.1038/nature14590.

Burkhardt, D.B., Stanley, J.S., Tong, A., Perdigoto, A.L., Gigante, S. A., Herold, K.C., Wolf, G., Giraldez, A.J., van Dijk, D., and Krishnaswamy, S. (2021). Quantifying the effect of experimental perturbations at single-cell resolution. Nat. Biotechnol. *39*, 619–629. https://doi.org/10.1038/s41587-020-00803-5.

Cacchiarelli, D., Qiu, X., Srivatsan, S., Manfredi, A., Ziller, M., Overbey, E., Grimaldi, A., Grimsby, J., Pokharel, P., Livak, K.J., et al. (2018). Aligning Single-Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of Myogenic Reprogramming Outcome. Cell Syst. *7*, 258–268.e3. https://doi.org/10.1016/j.cels.2018.07.006.

Campbell, K.R., and Yau, C. (2019). A descriptive marker gene approach to single-cell pseudotime inference. Bioinformatics *35*, 28–35. https://doi.org/10.1093/bioinformatics/bty498.

Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guilliams, M., Lambrecht, B., Preter, K.D., and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. Preprint at bioRxiv, 079509. https://doi.org/10.1101/079509.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science *357*, 661–667. https://doi.org/10.1126/science.aam8940.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

Carcamo-Orive, I., Hoffman, G.E., Cundiff, P., Beckmann, N.D., D'Souza, S.L., Knowles, J.W., Patel, A., Papatsenko, D., Abbasi, F., Reaven, G.M., et al. (2017). Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity. Cell Stem Cell *20*, 518–532.e9. https://doi.org/10.1016/j.stem.2016.11.005.

Chari, T., and Pachter, L. (2023). The specious art of single-cell genomics. PLoS Comput. Biol. *19*, e1011288. https://doi.org/10.1371/journal.pcbi.1011288.

Charrout, M., Reinders, M.J.T., and Mahfouz, A. (2020). Untangling biological factors influencing trajectory inference from single

cell data. NAR Genom. Bioinform. *2*, lqaa053. https://doi.org/10.1093/nargab/lqaa053.

Chen, X., Huang, Y., Huang, L., Huang, Z., Hao, Z.-Z., Xu, L., Xu, N., Li, Z., Mou, Y., Ye, M., et al. (2024). A brain cell atlas integrating single-cell transcriptomes across human brain regions. Nat. Med. *30*, 2679–2691. https://doi.org/10.1038/s41591-024-03150-z.

Cheng, B., Fang, W., Pastor, S., March, A.R., Porras, T., Wu, H.-W., Velez, M., Parekh, C., Maris, J.M., Asgharzadeh, S., and Huang, M. (2024). Comparison of human pluripotent stem cell differentiation protocols to generate neuroblastoma tumors. Sci. Rep. *14*, 23050. https://doi.org/10.1038/s41598-024-73947-y.

Cheng, S., Brenière-Letuffe, D., Ahola, V., Wong, A.O.T., Keung, H. Y., Gurung, B., Zheng, Z., Costa, K.D., Lieu, D.K., Keung, W., and Li, R.A. (2023). Single-cell RNA sequencing reveals maturation trajectory in human pluripotent stem cell-derived cardiomyocytes in engineered tissues. iScience *26*, 106302. https://doi.org/10.1016/j.isci.2023.106302.

Cho, H.J., Xie, E., Zhang, A., and Bekiranov, S. (2024). Generating Synthetic Single Cell Data from Bulk RNA-seq Using a Pretrained Variational Autoencoder. 2024.05.18.594837. https://doi.org/10.1101/2024.05.18.594837.

Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y.J., et al. (2020). High-depth African genomes inform human migration and health. Nature *586*, 741–748. https://doi.org/10.1038/s41586-020-2859-7.

Chu, T., Wang, Z., Pe'er, D., and Danko, C.G. (2022). Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat. Cancer *3*, 505–517. https://doi.org/10.1038/s43018-022-00356-3.

Churko, J.M., Garg, P., Treutlein, B., Venkatasubramanian, M., Wu, H., Lee, J., Wessells, Q.N., Chen, S.-Y., Chen, W.-Y., Chetal, K., et al. (2018). Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. Nat. Commun. *9*, 4906. https://doi.org/10.1038/s41467-018-07333-4.

Combes, A.N., Phipson, B., Lawlor, K.T., Dorison, A., Patrick, R., Zappia, L., Harvey, R.P., Oshlack, A., and Little, M.H. (2019a). Single cell analysis of the developing mouse kidney provides deeper insight into marker gene expression and ligand-receptor crosstalk. Development *146*, dev178673. https://doi.org/10.1242/dev.178673.

Combes, A.N., Zappia, L., Er, P.X., Oshlack, A., and Little, M.H. (2019b). Single-cell analysis reveals congruence between kidney organoids and human fetal kidney. Genome Med. *11*, 3. https://doi.org/10.1186/s13073-019-0615-0.

Crowell, H.L., Morillo Leonardo, S.X., Soneson, C., and Robinson, M.D. (2023). The shaky foundations of simulating single-cell RNA sequencing data. Genome Biol. *24*, 62. https://doi.org/10.1186/s13059-023-02904-1.

Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat. Methods *21*, 1470–1480. https://doi.org/10.1038/s41592-024-02201-0.

Cuomo, A.S.E., Seaton, D.D., McCarthy, D.J., Martinez, I., Bonder, M.J., Garcia-Bernardo, J., Amatya, S., Madrigal, P., Isaacson, A., Buettner, F., et al. (2020). Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. Nat. Commun. *11*, 810. https://doi.org/10.1038/s41467-020-14457-z.

De Donno, C., Hediyeh-Zadeh, S., Moinfar, A.A., Wagenstetter, M., Zappia, L., Lotfollahi, M., and Theis, F.J. (2023). Population-level integration of single-cell datasets enables multi-scale analysis across samples. Nat. Methods *20*, 1683–1692. https://doi.org/10.1038/s41592-023-02035-2.

DeBoever, C., Li, H., Jakubosky, D., Benaglio, P., Reyna, J., Olson, K.M., Huang, H., Biggs, W., Sandoval, E., D'Antonio, M., et al. (2017). Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. Cell Stem Cell *20*, 533–546.e7. https://doi.org/10.1016/j.stem.2017.03.009.

Deshpande, A., Loth, M., Sidiropoulos, D.N., Zhang, S., Yuan, L., Bell, A.T.F., Zhu, Q., Ho, W.J., Santa-Maria, C., Gilkes, D.M., et al. (2023). Uncovering the spatial landscape of molecular interactions within the tumor microenvironment through latent spaces. Cell Syst. *14*, 285–301.e4. https://doi.org/10.1016/j.cels.2023.03.004.

Dhainaut, M., Rose, S.A., Akturk, G., Wroblewska, A., Nielsen, S.R., Park, E.S., Buckup, M., Roudko, V., Pia, L., Sweeney, R., et al. (2022). Spatial CRISPR genomics identifies regulators of the tumor microenvironment. Cell *185*, 1223–1239.e20. https://doi.org/10.1016/j.cell.2022.02.015.

Ellis, J., Woltjen, K., Mital, S., Saito, M.K., Hotta, A., and Loring, J.F. (2025). Diversifying the reference iPSC line concept. Cell Stem Cell *32*, 873–877. https://doi.org/10.1016/j.stem.2025.05.004.

Fang, M., Gorin, G., and Pachter, L. (2024). Trajectory inference from single-cell genomics data with a process time model. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.26.577510.

Fischer, D.S., Fiedler, A.K., Kernfeld, E.M., Genga, R.M.J., Bastidas-Ponce, A., Bakhti, M., Lickert, H., Hasenauer, J., Maehr, R., and Theis, F.J. (2019). Inferring population dynamics from single-cell RNA-sequencing time series data. Nat. Biotechnol. *37*, 461–468. https://doi.org/10.1038/s41587-019-0088-0.

Foroutan, M., Bhuva, D.D., Lyu, R., Horan, K., Cursons, J., and Davis, M.J. (2018). Single sample scoring of molecular phenotypes. BMC Bioinf. *19*, 404. https://doi.org/10.1186/s12859-018-2435-4.

Freedman, B.S., Brooks, C.R., Lam, A.Q., Fu, H., Morizane, R., Agrawal, V., Saad, A.F., Li, M.K., Hughes, M.R., Werff, R.V., et al. (2015). Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. Nat. Commun. *6*, 8715. https://doi.org/10.1038/ncomms9715.

Friedman, C.E., Nguyen, Q., Lukowski, S.W., Helfer, A., Chiu, H.S., Miklas, J., Levy, S., Suo, S., Han, J.-D.J., Osteil, P., et al. (2018). Single-Cell Transcriptomic Analysis of Cardiac Differentiation from Human PSCs Reveals HOPX-Dependent Cardiomyocyte Maturation. Cell Stem Cell *23*, 586–598.e8. https://doi.org/10.1016/j.stem.2018.09.009.

Fu, X., Mo, S., Buendia, A., Laurent, A.P., Shao, A., Alvarez-Torres, M.D.M., Yu, T., Tan, J., Su, J., Sagatelian, R., et al. (2025). A

foundation model of transcription across human cell types. Nature *637*, 965–973. https://doi.org/10.1038/s41586-024-08391-z.

Garcia-Alonso, L., Handfield, L.-F., Roberts, K., Nikolakopoulou, K., Fernando, R.C., Gardner, L., Woodhams, B., Arutyunyan, A., Polanski, K., Hoo, R., et al. (2021). Mapping the temporal and spatial dynamics of the human endometrium *in vivo* and *in vitro*. Nat. Genet. *53*, 1698–1711. https://doi.org/10.1038/s41588-021-00972-2.

Ghosh, S., Nehme, R., and Barrett, L.E. (2022). Greater genetic diversity is needed in human pluripotent stem cell models. Nat. Commun. *13*, 7301. https://doi.org/10.1038/s41467-022-34940-z.

Gorin, G., Fang, M., Chari, T., and Pachter, L. (2022). RNA velocity unraveled. PLoS Comput. Biol. *18*, e1010492. https://doi.org/10.1371/journal.pcbi.1010492.

Grancharova, T., Gerbin, K.A., Rosenberg, A.B., Roco, C.M., Arakaki, J.E., DeLizo, C.M., Dinh, S.Q., Donovan-Maiye, R.M., Hirano, M., Nelson, A.M., et al. (2021). A comprehensive analysis of gene expression changes in a high replicate and open-source dataset of differentiating hiPSC-derived cardiomyocytes. Sci. Rep. *11*, 15845. https://doi.org/10.1038/s41598-021-94732-1.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427. https://doi.org/10.1038/nbt.4091.

Haidet-Phillips, A.M., Hester, M.E., Miranda, C.J., Meyer, K., Braun, L., Frakes, A., Song, S., Likhite, S., Murtha, M.J., Foust, K.D., et al. (2011). Astrocytes from familial and sporadic ALS patients are toxic to motor neurons. Nat. Biotechnol. *29*, 824–828. https://doi.org/10.1038/nbt.1957.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell *173*, 1307. https://doi.org/10.1016/j.cell.2018.05.012.

Haniffa, M., Taylor, D., Linnarsson, S., Aronow, B.J., Bader, G.D., Barker, R.A., Camara, P.G., Camp, J.G., Chédotal, A., Copp, A., et al. (2021). A roadmap for the Human Developmental Cell Atlas. Nature *597*, 196–205. https://doi.org/10.1038/s41586-021-03620-1.

Haniffa, M., Maartens, A., and Teichmann, S.A. (2023). How developmental cell atlases inform stem cell embryo models. Nat. Methods *20*, 1849–1851. https://doi.org/10.1038/s41592-023-02072-x.

Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., Wang, T., Ma, J., Zhang, X., and Song, L. (2024). Large-scale foundation model on single-cell transcriptomics. Nat. Methods *21*, 1481–1491. https://doi.org/10.1038/s41592-024-02305-7.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391–399. https://doi.org/10.1038/nature11405.

He, Z., Dony, L., Fleck, J.S., Szałata, A., Li, K.X., Slišković, I., Lin, H.-C., Santel, M., Atamian, A., Quadrato, G., et al. (2023). An integrated transcriptomic cell atlas of human neural organoids. 2023.10.05.561097. https://doi.org/10.1101/2023.10.05.561097.

Hegenbarth, J.-C., Lezzoche, G., De Windt, L.J., and Stoll, M. (2022). Perspectives on Bulk-Tissue RNA Sequencing and Single-Cell RNA Sequencing for Cardiac Transcriptomics. Front. Mol. Med. *2*, 839338. https://doi.org/10.3389/fmmed.2022.839338.

Heimberg, G., Kuo, T., DePianto, D., Heigl, T., Diamant, N., Salem, O., Scalia, G., Biancalani, T., Rock, J., Turley, S., et al. (2023). Scalable querying of human cell atlases via a foundational model reveals commonalities across fibrosis-associated macrophages. 2023.07.18.549537. https://doi.org/10.1101/2023.07.18.549537.

Hou, W., and Ji, Z. (2024). Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. Nat. Methods *21*, 1462–1465. https://doi.org/10.1038/s41592-024-02235-4.

Hou, W., Ji, Z., Chen, Z., Wherry, E.J., Hicks, S.C., and Ji, H. (2023). A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. Nat. Commun. *14*, 7286. https://doi.org/10.1038/s41467-023-42841-y.

Howden, S.E., Wilson, S.B., Groenewegen, E., Starks, L., Forbes, T.A., Tan, K.S., Vanslambrouck, J.M., Holloway, E.M., Chen, Y.-H., Jain, S., et al. (2021). Plasticity of distal nephron epithelia from human kidney organoids enables the induction of ureteric tip and stalk. Cell Stem Cell *28*, 671–684.e6. https://doi.org/10.1016/j.stem.2020.12.001.

Hu, T., Allam, M., Cai, S., Henderson, W., Yueh, B., Garipcan, A., Ievlev, A.V., Afkarian, M., Beyaz, S., and Coskun, A.F. (2023). Single-cell spatial metabolomics with cell-type specific protein profiling for tissue systems biology. Nat. Commun. *14*, 8260. https://doi.org/10.1038/s41467-023-43917-5.

Ianevski, A., Giri, A.K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat. Commun. *13*, 1246. https://doi.org/10.1038/s41467-022-28803-w.

Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94. https://doi.org/10.1038/nbt.4042.

Kannan, S., Miyamoto, M., Zhu, R., Lynott, M., Guo, J., Chen, E.Z., Colas, A.R., Lin, B.L., and Kwon, C. (2023). Trajectory reconstruction identifies dysregulation of perinatal maturation programs in pluripotent stem cell-derived cardiomyocytes. Cell Rep. *42*, 112330. https://doi.org/10.1016/j.celrep.2023.112330.

Kharchenko, P.V. (2021). The triumphs and limitations of computational methods for scRNA-seq. Nat. Methods *18*, 723–732. https://doi.org/10.1038/s41592-021-01171-x.

Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J., et al. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. Nature *546*, 370–375. https://doi.org/10.1038/nature22403.

Kim, J., Koo, B.-K., and Knoblich, J.A. (2020a). Human organoids: model systems for human biology and medicine. Nat. Rev. Mol. Cell Biol. *21*, 571–584. https://doi.org/10.1038/s41580-020-0259-3.

Kim, T.W., Koo, S.Y., and Studer, L. (2020b). Pluripotent Stem Cell Therapies for Parkinson Disease: Present Challenges and Future

Opportunities. Front. Cell Dev. Biol. *8*, 729. https://doi.org/10.3389/fcell.2020.00729.

Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. Nat. Methods *15*, 359–362. https://doi.org/10.1038/nmeth.4644.

Kong, W., Fu, Y.C., Holloway, E.M., Garipler, G., Yang, X., Mazzoni, E.O., and Morris, S.A. (2022). Capybara: A computational tool to measure cell identity and fate transitions. Cell Stem Cell *29*, 635–649.e11. https://doi.org/10.1016/j.stem.2022.03.001.

Kopp, W., Akalin, A., and Ohler, U. (2022). Simultaneous dimensionality reduction and integration for single-cell ATAC-seq data using deep learning. Nat. Mach. Intell. *4*, 162–168. https://doi.org/10.1038/s42256-022-00443-1.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566–580.e19. https://doi.org/10.1016/j.cell.2016.09.027.

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.

Laco, F., Woo, T.L., Zhong, Q., Szmyd, R., Ting, S., Khan, F.J., Chai, C.L.L., Reuveny, S., Chen, A., and Oh, S. (2018). Unraveling the Inconsistencies of Cardiac Differentiation Efficiency Induced by the GSK3β Inhibitor CHIR99021 in Human Pluripotent Stem Cells. Stem Cell Rep. *10*, 1851–1866. https://doi.org/10.1016/j.stemcr.2018.03.023.

Laidlaw, R.F., Briggs, E.M., Matthews, K.R., McCulloch, R., and Otto, T.D. (2024). TrAGEDy: Trajectory Alignment of Gene Expression Dynamics. Preprint at bioRxiv. https://doi.org/10.1101/2022.12.21.521424.

Lam, Y.Y., Keung, W., Chan, C.H., Geng, L., Wong, N., Brenière-Letuffe, D., Li, R.A., and Cheung, Y.F. (2020). Single-Cell Transcriptomics of Engineered Cardiac Tissues From Patient-Specific Induced Pluripotent Stem Cell–Derived Cardiomyocytes Reveals Abnormal Developmental Trajectory and Intrinsic Contractile Defects in Hypoplastic Right Heart Syndrome. J. Am. Heart Assoc. *9*, e016528. https://doi.org/10.1161/JAHA.120.016528.

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat. Methods *19*, 159–170. https://doi.org/10.1038/s41592-021-01346-6.

Lause, J., Kobak, D., and Berens, P. (2024). The art of seeing the elephant in the room: 2D embeddings of single-cell data do make sense. Preprint at bioRxiv. https://doi.org/10.1101/2024.03.26.586728.

Lawlor, K.T., Vanslambrouck, J.M., Higgins, J.W., Chambon, A., Bishard, K., Arndt, D., Er, P.X., Wilson, S.B., Howden, S.E., Tan, K.S., et al. (2021). Cellular extrusion bioprinting improves kidney organoid reproducibility and conformation. Nat. Mater. *20*, 260–271. https://doi.org/10.1038/s41563-020-00853-9.

Lederer, A.R., Leonardi, M., Talamanca, L., Herrera, A., Droin, C., Khven, I., Carvalho, H.J.F., Valente, A., Mantes, A.D., Arabí, P.M., et al. (2024). Statistical inference with a manifold-constrained RNA velocity model uncovers cell cycle speed modulations. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.18.576093.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291. https://doi.org/10.1038/nature19057.

Li, S., Zhang, P., Chen, W., Ye, L., Brannan, K.W., Le, N.-T., Abe, J.I., Cooke, J.P., and Wang, G. (2024). A relay velocity model infers cell-dependent RNA velocity. Nat. Biotechnol. *42*, 99–108. https://doi.org/10.1038/s41587-023-01728-5.

Lian, X., Bao, X., Zilberter, M., Westman, M., Fisahn, A., Hsiao, C., Hazeltine, L.B., Dunn, K.K., Kamp, T.J., and Palecek, S.P. (2015). Chemically defined, albumin-free human cardiomyocyte generation. Nat. Methods *12*, 595–596. https://doi.org/10.1038/nmeth.3448.

Liang, Q., Cheng, X., Wang, J., Owen, L., Shakoor, A., Lillvis, J.L., Zhang, C., Farkas, M., Kim, I.K., Li, Y., et al. (2023). A multi-omics atlas of the human retina at single-cell resolution. Cell Genom. *3*, 100298. https://doi.org/10.1016/j.xgen.2023.100298.

Lindström, N.O., Tran, T., Guo, J., Rutledge, E., Parvez, R.K., Thornton, M.E., Grubbs, B., McMahon, J.A., and McMahon, A.P. (2018). Conserved and Divergent Molecular and Anatomic Features of Human and Mouse Nephron Patterning. J. Am. Soc. Nephrol. *29*, 825–840. https://doi.org/10.1681/ASN.2017091036.

Little, M.H., and Combes, A.N. (2019). Kidney organoids: accurate models or fortunate accidents. Genes Dev. *33*, 1319–1345. https://doi.org/10.1101/gad.329573.119.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., et al. (2020). Cells of the adult human heart. Nature *588*, 466–472. https://doi.org/10.1038/s41586-020-2797-4.

Liu, G., Zhao, Y., Zhao, Y., Wang, T., Cai, Q., Wang, X., Wen, Z., Lin, L., Yang, G., and Chen, J. (2025). SCARF: Single Cell ATAC-seq and RNA-seq Foundation model. 2025.04.07.647689. https://doi.org/10.1101/2025.04.07.647689.

Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058. https://doi.org/10.1038/s41592-018-0229-2.

Lorthongpanich, C., Cheow, L.F., Balu, S., Quake, S.R., Knowles, B. B., Burkholder, W.F., Solter, D., and Messerschmidt, D.M. (2013). Single-Cell DNA-Methylation Analysis Reveals Epigenetic Chimerism in Preimplantation Embryos. Science *341*, 1110–1112. https://doi.org/10.1126/science.1240617.

Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. Nat. Methods *16*, 715–721. https://doi.org/10.1038/s41592-019-0494-8.

Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., and Theis, F.J. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nat. Methods *19*, 41–50. https://doi.org/10.1038/s41592-021-01336-8.

Lundberg, M., Eriksson, A., Tran, B., Assarsson, E., and Fredriksson, S. (2011). Homogeneous antibody-based proximity extension

assays provide sensitive and specific detection of low-abundant proteins in human blood. Nucleic Acids Res. *39*, e102. https://doi.org/10.1093/nar/gkr424.

Luo, E., Hao, M., Wei, L., and Zhang, X. (2024). scDiffusion: conditional generation of high-quality single-cell data using diffusion model. Bioinformatics *40*, btae518. https://doi.org/10.1093/bioinformatics/btae518.

Lyra-Leite, D.M., Gutiérrez-Gutiérrez, Ó., Wang, M., Zhou, Y., Cyganek, L., and Burridge, P.W. (2022). A review of protocols for human iPSC culture, cardiac differentiation, subtype-specification, maturation, and direct reprogramming. STAR Protoc. *3*, 101560. https://doi.org/10.1016/j.xpro.2022.101560.

Ma, F., and Pellegrini, M. (2020). ACTINN: automated identification of cell types in single cell RNA sequencing. Bioinformatics *36*, 533–538. https://doi.org/10.1093/bioinformatics/btz592.

van der Maaten, L., and Hinton, G. (2008). Viualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

MacLean, A.L., Hong, T., and Nie, Q. (2018). Exploring intermediate cell states through the lens of single cells. Curr. Opin. Syst. Biol. *9*, 32–41. https://doi.org/10.1016/j.coisb.2018.02.009.

Marouf, M., Machart, P., Bansal, V., Kilian, C., Magruder, D.S., Krebs, C.F., and Bonn, S. (2020). Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. Nat. Commun. *11*, 166. https://doi.org/10.1038/s41467-019-14018-z.

Maulding, N.D., Seninge, L., and Stuart, J.M. (2024). Associating transcription factors to single-cell trajectories with DREAMIT. Genome Biol. *25*, 220. https://doi.org/10.1186/s13059-024-03368-7.

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.

Mimitou, E.P., Lareau, C.A., Chen, K.Y., Zorzetto-Fernandes, A.L., Hao, Y., Takeshima, Y., Luo, W., Huang, T.-S., Yeung, B.Z., Papalexi, E., et al. (2021). Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat. Biotechnol. *39*, 1246–1258. https://doi.org/10.1038/s41587-021-00927-2.

Miranda, A.M.A., Janbandhu, V., Maatz, H., Kanemaru, K., Cranley, J., Teichmann, S.A., Hübner, N., Schneider, M.D., Harvey, R.P., and Noseda, M. (2023). Single-cell transcriptomics for the assessment of cardiac disease. Nat. Rev. Cardiol. *20*, 289–308. https://doi.org/10.1038/s41569-022-00805-7.

Morizane, R., Lam, A.Q., Freedman, B.S., Kishi, S., Valerius, M.T., and Bonventre, J.V. (2015). Nephron organoids derived from human pluripotent stem cells model kidney development and injury. Nat. Biotechnol. *33*, 1193–1200. https://doi.org/10.1038/nbt.3392.

Neavin, D.R., Steinmann, A.M., Farbehi, N., Chiu, H.S., Daniszewski, M.S., Arora, H., Bermudez, Y., Moutinho, C., Chan, C.-L., Bax, M., et al. (2023). A village in a dish model system for population-scale hiPSC studies. Nat. Commun. *14*, 3240. https://doi.org/10.1038/s41467-023-38704-1.

Nguyen, H.N., Byers, B., Cord, B., Shcheglovitov, A., Byrne, J., Gujar, P., Kee, K., Schüle, B., Dolmetsch, R.E., Langston, W., et al. (2011). LRRK2 Mutant iPSC-Derived DA Neurons Demonstrate Increased Susceptibility to Oxidative Stress. Cell Stem Cell *8*, 267–280. https://doi.org/10.1016/j.stem.2011.01.013.

Nguyen, Q.H., Lukowski, S.W., Chiu, H.S., Senabouth, A., Bruxner, T.J.C., Christ, A.N., Palpant, N.J., and Powell, J.E. (2018). Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. Genome Res. *28*, 1053–1066. https://doi.org/10.1101/gr.223925.117.

Nguyen, T.B., Lac, Q., Abdi, L., Banerjee, D., Deng, Y., and Zhang, Y. (2022). Harshening stem cell research and precision medicine: The states of human pluripotent cells stem cell repository diversity, and racial and sex differences in transcriptomes. Front. Cell Dev. Biol. *10*, 1071243. https://doi.org/10.3389/fcell.2022.1071243.

Osafune, K., Caron, L., Borowiak, M., Martinez, R.J., Fitz-Gerald, C.S., Sato, Y., Cowan, C.A., Chien, K.R., and Melton, D.A. (2008). Marked differences in differentiation propensity among human embryonic stem cell lines. Nat. Biotechnol. *26*, 313–315. https://doi.org/10.1038/nbt1383.

Pantazis, C.B., Yang, A., Lara, E., McDonough, J.A., Blauwendraat, C., Peng, L., Oguro, H., Kanaujiya, J., Zou, J., Sebesta, D., et al. (2022). A reference human induced pluripotent stem cell line for large-scale collaborative studies. Cell Stem Cell *29*, 1685–1702.e22. https://doi.org/10.1016/j.stem.2022.11.004.

Park, D.S., Kozaki, T., Tiwari, S.K., Moreira, M., Khalilnezhad, A., Torta, F., Olivié, N., Thiam, C.H., Liani, O., Silvin, A., et al. (2023). iPS-cell-derived microglia promote brain organoid maturation via cholesterol transfer. Nature *623*, 397–405. https://doi.org/10.1038/s41586-023-06713-1.

Pashos, E.E., Park, Y., Wang, X., Raghavan, A., Yang, W., Abbey, D., Peters, D.T., Arbelaez, J., Hernandez, M., Kuperwasser, N., et al. (2017). Large diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. Cell Stem Cell *20*, 558–570.e10. https://doi.org/10.1016/j.stem.2017.03.017.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.

Pavan, C., Davidson, K.C., Payne, N., Frausin, S., Hunt, C.P.J., Moriarty, N., Berrocal Rubio, M.Á., Elahi, Z., Quattrocchi, A.T., Abu-Bonsrah, K.D., et al. (2025). A cloaked human stem-cell-derived neural graft capable of functional integration and immune evasion in rodent models. Cell Stem Cell *32*, 710–726.e8. https://doi.org/10.1016/j.stem.2025.03.008.

Pearce, J.D., Simmonds, S.E., Mahmoudabadi, G., Krishnan, L., Palla, G., Istrate, A.-M., Tarashansky, A., Nelson, B., Valenzuela, O., Li, D., et al. (2025). A Cross-Species Generative Cell Atlas Across 1.5 Billion Years of Evolution: The TranscriptFormer Single-cell Model. Preprint at bioRxiv. https://doi.org/10.1101/2025.04.25.650731.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.,

et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825.

Phipson, B., Er, P.X., Combes, A.N., Forbes, T.A., Howden, S.E., Zappia, L., Yen, H.-J., Lawlor, K.T., Hale, L.J., Sun, J., et al. (2019). Evaluation of variability in human kidney organoids. Nat. Methods *16*, 79–87. https://doi.org/10.1038/s41592-018-0253-2.

Plass, M., Solana, J., Wolf, F.A., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F.J., Kocks, C., and Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science *360*, eaaq1723. https://doi.org/10.1126/science.aaq1723.

Abdulla, S., Aevermann, B., Assis, P., Badajoz, S., Bell, S.M., Bezzi, E., Cakir, B., Chaffer, J., Chambers, S., Michael Cherry, J., et al. (2023). CZ CELL×GENE Discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.30.563174.

Prondzynski, M., Berkson, P., Trembley, M.A., Tharani, Y., Shani, K., Bortolin, R.H., Sweat, M.E., Mayourian, J., Yucel, D., Cordoves, A.M., et al. (2024). Efficient and reproducible generation of human iPSC-derived cardiomyocytes and cardiac organoids in stirred suspension systems. Nat. Commun. *15*, 5929. https://doi.org/10.1038/s41467-024-50224-0.

Pullin, J.M., and McCarthy, D.J. (2024). A comparison of marker gene selection methods for single-cell RNA sequencing data. Genome Biol. *25*, 56. https://doi.org/10.1186/s13059-024-03183-0.

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat. Biotechnol. *29*, 886–891. https://doi.org/10.1038/nbt.1991.

Ramos, D.M., Skarnes, W.C., Singleton, A.B., Cookson, M.R., and Ward, M.E. (2021). Tackling neurodegenerative diseases with genomic engineering: A new stem cell initiative from the NIH. Neuron *109*, 1080–1083. https://doi.org/10.1016/j.neuron.2021.03.022.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782. https://doi.org/10.1038/nbt.2282.

Ransick, A., Lindström, N.O., Liu, J., Zhu, Q., Guo, J.-J., Alvarado, G.F., Kim, A.D., Black, H.G., Kim, J., and McMahon, A.P. (2019). Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. Dev. Cell *51*, 399–413.e7. https://doi.org/10.1016/j.devcel.2019.10.005.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. eLife *6*, e27041. https://doi.org/10.7554/eLife.27041.

Revah, O., Gore, F., Kelley, K.W., Andersen, J., Sakai, N., Chen, X., Li, M.-Y., Birey, F., Yang, X., Saw, N.L., et al. (2022). Maturation and circuit integration of transplanted human cortical organoids. Nature *610*, 319–326. https://doi.org/10.1038/s41586-022-05277-w.

Riley, E.A., and Schekman, R. (2021). Open science takes on Parkinson's disease. eLife *10*, e66546. https://doi.org/10.7554/eLife.66546.

Rood, J.E., Wynne, S., Robson, L., Hupalowska, A., Randell, J., Teichmann, S.A., and Regev, A. (2025). The Human Cell Atlas from a cell census to a unified foundation model. Nature *637*, 1065–1071. https://doi.org/10.1038/s41586-024-08338-4.

Roost, M.S., van Iperen, L., Ariyurek, Y., Buermans, H.P., Arindrarto, W., Devalla, H.D., Passier, R., Mummery, C.L., Carlotti, F., de Koning, E.J.P., et al. (2015). KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. Stem Cell Rep. *4*, 1112–1124. https://doi.org/10.1016/j.stemcr.2015.05.002.

Rosen, Y., Roohani, Y., Agarwal, A., Samotorčan, L., Consortium, T. S., Quake, S.R., and Leskovec, J. (2023). Universal Cell Embeddings: A Foundation Model for Cell Biology. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.28.568918.

Rothová, M.M., Nielsen, A.V., Proks, M., Wong, Y.F., Riveiro, A.R., Linneberg-Agerholm, M., David, E., Amit, I., Trusina, A., and Brickman, J.M. (2022). Identification of the central intermediate in the extra-embryonic to embryonic endoderm transition through single-cell transcriptomics. Nat. Cell Biol. *24*, 833–844. https://doi.org/10.1038/s41556-022-00923-x.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. *37*, 547–554. https://doi.org/10.1038/s41587-019-0071-9.

Salehin, N., Proks, M., and Brickman, J.M. (2025). A Deep Learning Approach to Assessing Cell Identity in Stem Cell-Based Embryo Models. Methods Mol. Biol. https://doi.org/10.1007/7651_2025_654.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F. M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell *174*, 1015–1030.e16. https://doi.org/10.1016/j.cell.2018.07.028.

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.

Shinozawa, T., Kimura, M., Cai, Y., Saiki, N., Yoneyama, Y., Ouchi, R., Koike, H., Maezawa, M., Zhang, R.-R., Dunn, A., et al. (2021). High-Fidelity Drug-Induced Liver Injury Screen Using Human Pluripotent Stem Cell-Derived Organoids. Gastroenterology *160*, 831–846.e10. https://doi.org/10.1053/j.gastro.2020.10.002.

Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madissoon, E., Markov, N.S., Zaragosi, L.-E., Ji, Y., Ansari, M., et al. (2023). An integrated cell atlas of the lung in health and disease. Nat. Med. *29*, 1563–1577. https://doi.org/10.1038/s41591-023-02327-2.

Song, D., Wang, Q., Yan, G., Liu, T., Sun, T., and Li, J.J. (2024). scDesign3 generates realistic in silico data for multimodal single-cell and spatial omics. Nat. Biotechnol. *42*, 247–252. https://doi.org/10.1038/s41587-023-01772-1.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P.

(2017). Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868. https://doi.org/10.1038/nmeth.4380.

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 Genom. *19*, 477. https://doi.org/10.1186/s12864-018-4772-0.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Sugihara, R., Kato, Y., Mori, T., and Kawahara, Y. (2022). Alignment of single-cell trajectory trees with CAPITAL. Nat. Commun. *13*, 5972. https://doi.org/10.1038/s41467-022-33681-3.

Sumanaweera, D., Suo, C., Cujba, A.-M., Muraro, D., Dann, E., Polanski, K., Steemers, A.S., Lee, W., Oliver, A.J., Park, J.-E., et al. (2023). Gene-level alignment of single cell trajectories. 2023.03.08.531713. https://doi.org/10.1101/2023.03.08.531713.

Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367–372. https://doi.org/10.1038/s41586-018-0590-4.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024.

Takasato, M., Er, P.X., Chiu, H.S., Maier, B., Baillie, G.J., Ferguson, C., Parton, R.G., Wolvetang, E.J., Roost, M.S., Chuva de Sousa Lopes, S.M., and Little, M.H. (2015). Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. Nature *526*, 564–568. https://doi.org/10.1038/nature15695.

Tan, Y., and Cahan, P. (2019). SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. Cell Syst. *9*, 207–213.e2. https://doi.org/10.1016/j.cels.2019.06.004.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382. https://doi.org/10.1038/nmeth.1315.

Tarashansky, A.J., Musser, J.M., Khariton, M., Li, P., Arendt, D., Quake, S.R., and Wang, B. (2021). Mapping single-cell atlases throughout Metazoa unravels cell type evolution. eLife *10*, e66747. https://doi.org/10.7554/eLife.66747.

The H3Africa Consortium, Matovu, E., Bucheton, B., Chisi, J., Enyaru, J., Hertz-Fowler, C., Koffi, M., Macleod, A., Mumba, D., Sidibe, I., et al. (2014). Enabling the genomic revolution in Africa. Science *344*, 1346–1348. https://doi.org/10.1126/science.1251546.

Theodoris, C.V., Xiao, L., Chopra, A., Chaffin, M.D., Al Sayed, Z.R., Hill, M.C., Mantineo, H., Brydon, E.M., Zeng, Z., Liu, X.S., and Ellinor, P.T. (2023). Transfer learning enables predictions in network biology. Nature *618*, 616–624. https://doi.org/10.1038/s41586-023-06139-9.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic Stem Cell Lines Derived from Human Blastocysts. Science *282*, 1145–1147. https://doi.org/10.1126/science.282.5391.1145.

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.

Treppner, M., Salas-Bastos, A., Hess, M., Lenz, S., Vogel, T., and Binder, H. (2021). Synthetic single cell RNA sequencing data from small pilot studies using deep generative models. Sci. Rep. *11*, 9403. https://doi.org/10.1038/s41598-021-88875-4.

Tsoucas, D., Dong, R., Chen, H., Zhu, Q., Guo, G., and Yuan, G.-C. (2019). Accurate estimation of cell-type composition from gene expression data. Nat. Commun. *10*, 2975. https://doi.org/10.1038/s41467-019-10802-z.

Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289. https://doi.org/10.1038/s41586-021-04158-y.

Unterman, A., Sumida, T.S., Nouri, N., Yan, X., Zhao, A.Y., Gasque, V., Schupp, J.C., Asashima, H., Liu, Y., Cosme, C., et al. (2022). Single-cell multi-omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19. Nat. Commun. *13*, 440. https://doi.org/10.1038/s41467-021-27716-4.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573. https://doi.org/10.1126/science.aah4573.

Volpato, V., Smith, J., Sandor, C., Ried, J.S., Baud, A., Handel, A., Newey, S.E., Wessely, F., Attar, M., Whiteley, E., et al. (2018). Reproducibility of Molecular Phenotypes after Long-Term Differentiation to Human iPSC-Derived Neurons: A Multi-Site Omics Study. Stem Cell Rep. *11*, 897–911. https://doi.org/10.1016/j.stemcr.2018.08.013.

Wang, J., Morgan, W., Saini, A., Liu, T., Lough, J., and Han, L. (2022). Single-cell transcriptomic profiling reveals specific maturation signatures in human cardiomyocytes derived from LMNB2-inactivated induced pluripotent stem cells. Front. Cell Dev. Biol. *10*, 895162. https://doi.org/10.3389/fcell.2022.895162.

Wang, L., Nguyen, T., Rosa-Garrido, M., Zhou, Y., Cleveland, D.C., and Zhang, J. (2023). Comparative analysis of the cardiomyocyte differentiation potential of induced pluripotent stem cells reprogrammed from human atrial or ventricular fibroblasts. Front. Bioeng. Biotechnol. *11*, 1108340. https://doi.org/10.3389/fbioe.2023.1108340.

Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat. Commun. *10*, 380. https://doi.org/10.1038/s41467-018-08023-x.

Warren, C.R., O'Sullivan, J.F., Friesen, M., Becker, C.E., Zhang, X., Liu, P., Wakabayashi, Y., Morningstar, J.E., Shi, X., Choi, J., et al. (2017). Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of GWAS Variants in Metabolic Disease. Cell Stem Cell *20*, 547–557.e7. https://doi.org/10.1016/j.stem.2017.01.010.

Weatherbee, B.A.T., Gantner, C.W., Iwamoto-Stohl, L.K., Daza, R. M., Hamazaki, N., Shendure, J., and Zernicka-Goetz, M. (2023). Pluripotent stem cell-derived model of the post-implantation human embryo. Nature *622*, 584–593. https://doi.org/10.1038/s41586-023-06368-y.

Wilson, S.B., Howden, S.E., Vanslambrouck, J.M., Dorison, A., Alquicira-Hernandez, J., Powell, J.E., and Little, M.H. (2022). DevKidCC allows for robust classification and direct comparisons of kidney organoid datasets. Genome Med. *14*, 19. https://doi.org/10.1186/s13073-022-01023-z.

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. *20*, 59. https://doi.org/10.1186/s13059-019-1663-x.

Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M.I., and Yosef, N. (2021). Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol. Syst. Biol. *17*, e9620. https://doi.org/10.15252/msb.20209620.

Xu, Q., Halle, L., Hediyeh-zadeh, S., Kuijs, M., Kilik, U., Yu, Q., Frum, T., Adam, L., Parikh, S., Gander, M., et al. (2023). An integrated transcriptomic cell atlas of human endoderm-derived organoids. Preprint at bioRxiv, 567825. https://doi.org/10.1101/2023.11.20.567825.

Yamatani, Y., and Nakai, K. (2022). Comprehensive comparison of gene expression diversity among a variety of human stem cells. NAR Genom. Bioinform. *4*, lqac087. https://doi.org/10.1093/nargab/lqac087.

Yang, F., Wang, W., Wang, F., Fang, Y., Tang, D., Huang, J., Lu, H., and Yao, J. (2022). scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nat. Mach. Intell. *4*, 852–866. https://doi.org/10.1038/s42256-022-00534-z.

Yoshimura, Y., Muto, Y., Ledru, N., Wu, H., Omachi, K., Miner, J.H., and Humphreys, B.D. (2023). A single-cell multiomic analysis of kidney organoid differentiation. Proc. Natl. Acad. Sci. USA *120*, e2219699120. https://doi.org/10.1073/pnas.2219699120.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920. https://doi.org/10.1126/science.1151526.

Zhang, A.W., O'Flanagan, C., Chavez, E.A., Lim, J.L.P., Ceglia, N., McPherson, A., Wiens, M., Walters, P., Chan, T., Hewitson, B., et al. (2019). Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. Nat. Methods *16*, 1007–1015. https://doi.org/10.1038/s41592-019-0529-1.

Zhang, J., Ubas, A.A., de Borja, R., Svensson, V., Thomas, N., Thakar, N., Lai, I., Winters, A., Khan, U., Jones, M.G., et al. (2025). Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. 2025.02.20.639398. https://doi.org/10.1101/2025.02.20.639398.

Zhu, Z., and Huangfu, D. (2013). Human pluripotent stem cells: an emerging model in developmental biology. Dev. Camb. Engl. *140*, 705–717. https://doi.org/10.1242/dev.086165.