REVIEW ARTICLE

Human Cell Atlas and cell-type authentication for regenerative medicine

Yulia Panina[1], Peter Karagiannis[1], Andreas Kurtz[2], Glyn N. Stacey[3,4,5] and Wataru Fujibuchi[1]

Abstract

In modern biology, the correct identification of cell types is required for the developmental study of tissues and organs and the production of functional cells for cell therapies and disease modeling. For decades, cell types have been defined on the basis of morphological and physiological markers and, more recently, immunological markers and molecular properties. Recent advances in single-cell RNA sequencing have opened new doors for the characterization of cells at the individual and spatiotemporal levels on the basis of their RNA profiles, vastly transforming our understanding of cell types. The objective of this review is to survey the current progress in the field of cell-type identification, starting with the Human Cell Atlas project, which aims to sequence every cell in the human body, to molecular marker databases for individual cell types and other sources that address cell-type identification for regenerative medicine based on cell data guidelines.

The Human Cell Atlas

The Human Cell Atlas (HCA) project aims to characterize all cells by single-cell analytical techniques, specifically single-cell RNA sequencing (scRNA-seq) and assay for transposase accessible chromatin sequencing (scATAC-seq) and to link this information to classic knowledge, namely, location, lineage, and type, as well as cell states, state transitions, and cell−cell interactions. The HCA was initiated to unify scRNA-seq data in a manner similar to the data compilation achieved in the Human Genome Project[1]. As of February 2020, the HCA project has participation from 1027 institutes in 71 countries, from which 81 laboratories have already posted scRNA-seq data for 34 organs and tissues, including the liver[2], lung[3], blood and immune systems[4], plasma cells[5], human cortex[6], colon[7], and retina[8], with the total number of sequenced cells reaching 4.5 million. The HCA is an open science project, and its standard operating protocols (SOPs) are available on the Web. A standardized data analysis service called the Data Coordination Platform (DCP) is provided to the community. In the long term, the HCA project aims to define not only normal cells but also cells in specific disease states[1,9].

The HCA project is generating tremendous amounts of data. For example, there are currently 148 members in the immune system group, which is conducting approximately 100 projects and is expected to include more than 1000 projects over time. One project[10] has produced scRNA-seq data for 530,000 cells. The challenge for the HCA project is to create a universal classification system for all of these projects. Adding to this challenge are regional HCAs, such as HCA Asia, that focus on regional diseases. To assign cell types to all these cells, exhaustive analysis methods are needed. Once this goal is accomplished, the HCA can shift to its second goal, which is to understand how normal cell states transition into disease states[11].

Challenge of human cell-type classifications in the HCA

A typical scheme for assigning cell types to scRNA-seq data requires computational analysis and a workflow

Correspondence: Wataru Fujibuchi (fujibuchi-g@cira.kyoto-u.ac.jp)
[1]Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
[2]BIH Center for Regenerative Therapies (BCRT), Charité—Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany
Full list of author information is available at the end of the article

consisting of preprocessing and downstream analysis steps. Preprocessing is necessary to correct biological and technical noise caused by experimental errors and sample variabilities, thus contributing to quality control and normalization, whereas downstream analysis aims to cluster single-cell data on the basis of similarities in profiles and assign cell types using various computational tools depending on the data type[12,13]. Each cell type can be identified by a uniquely characteristic transcript pattern of gene expressions (i.e., RNA signature) or a uniquely characteristic transcript profile[14]. However, because the number of clusters depends on the algorithms and parameters, the results are not generally optimized for separating subtypes or states of the same cell type (e.g., cell cycle phase differences)[15].

In this regard, several clustering approaches have been developed, including partition type (such as *k*-means and self-organizing maps), hierarchical clustering, graph-based (such as spectral clustering or clique detection), mixture model (such as Seurat), density-based, neural network type, ensemble, and affinity propagation type[16–18]. However, most of these approaches require parameter settings, and to complicate matters, some depend on random values. Consequently, the resulting clusters could vary in number and cluster members. Detailed reviews of these approaches can be found in recent papers[15,19]. In addition, a variety of software tools are available for clustering approaches, and the number is constantly growing.

Alternatively, the identification of a known cell type by its RNA signature can be performed using a cell−cell similarity search. In this case, researchers can identify known cell types and find similar or related unknown cell types by using hitherto unknown relationships. Software programs for cell−cell similarity search analyses include CellAtlasSearch[20], CellSim[21], and Cell BLAST[22], all of which are tailored to different needs. For example, Cell BLAST has a "special tuning" mode for handling batch effects between a query and reference. CellSim calculates the similarity of different cells on the basis of cell ontology and molecular networks and has a feature that allows users to identify the cell type by entering a list of genes. Finally, CellAtlasSearch is tailored for handling ultra-large RNA-seq data through parallel screening of tens of thousands of single cells using efficient clustering methods. However, these methods are typically dependent on existing cell data and thus are effective for cell-type identification that demands large amounts of annotated cell-type entries.

Another approach for assigning cell types is the integration of gene expression information and spatial information of the individual cells to provide a molecular description of each cell type in the context of the tissue microenvironment. Recently, developed computational methods allow, in principle, the reconstruction of a spatial map of tissues using scRNA-seq data[23−25]. scRNA-seq-based maps have revealed cell-type-specific functions in the liver[26], blastocyst[27], and growth plate[28]. The expression of cell adhesion genes and specific gene functions defined by Gene Ontology (GO) terms is being used to develop new tools for single-cell 3D transcriptome analysis that enhance spatial prediction[29,30]. Although the identification of cell types in a spatial context is expected to yield more information relevant to the in vivo environment, these cutting-edge approaches are still at the elementary stage and need further improvement before they can be widely used.

To expedite methods for cell-type identification, the HCA project has multiple committees and working groups that coordinate the efforts of independent research groups and unify the results of scRNA-seq. Typically, the research groups perform all the data preparation, acquisition, and analysis, and the HCA committees provide the general framework, guidelines, and data repository space. The DCP team supplies vetted algorithms for data processing on the portal, and the research groups choose the algorithms for data processing, depositing the data into the HCA database under controlled pipelines. Hence, the cell-type classification, which is part of the RNA-seq data analysis, is performed separately by various groups, with the results processed by the cooperating DCP team according to HCA guidelines to ensure consistency in cell-type assignment and authentication.

History of cell-type classification

With the advances in stem cell research that have made it possible to engineer cells for cell therapies and drug discovery, the identification and authentication of cell types have become emerging priorities in the biological community[31]. The greatest problem in the authentication of human cells lies in the lack of integrated standard metrics for cell morphology, gene expression, and molecular markers[32]. Considering that scRNA-seq data are usually confounded by a high degree of noise[33] and that current scientific communities use highly variable methodologies for scRNA-seq data analysis[34], it is very difficult to standardize the data and harmonize it with classical morphology and marker maps. Accordingly, the HCA project has invested both time and resources to study ways of handling technical and biological noise affecting data reproducibility and the degradation of biopsy samples[35].

Historically, the characterization of cell types was based on histological, i.e., anatomical, morphological, and functional, criteria[36]. For example, the earliest attempt to classify cells in the nervous system was based on histological characteristics, such as the locations from which the cells were obtained, cell morphology, and the presence

of certain molecular markers[37]. Since then, the location (i.e., cerebral cortex, cardiac muscle, or stomach), cell morphology (i.e., fibroblast-like or epithelial-like), and molecular markers (such as CD75-positive cells, etc.) have been accepted by the scientific community as the three main pillars for defining cell types[38–41]. To accelerate cell-type identification, several cell marker databases are available (Table 1).

Labome provides a list of 226 markers for epithelial, dendritic, glial, bone marrow, natural killer, and other cell types[42]. CellFinder was the first database website of molecular markers and now features information on 3394 cell types, 50,951 cell lines, and 553,905 protein expressions[43]. However, CellFinder data are diverse for species (mammals, fish, invertebrates, bacteria, viruses, plants, etc.), and the site includes other data, such as microscopic and anatomical images, whole-genome expression profiles from RNA-seq and microarrays, etc. In other words, it is not a cell marker database per se but a collection of various data of different cell types, which makes looking for markers relatively difficult compared to the search in other databases.

The first database compiled exclusively with markers of human and mouse cells was CellMarker[44], which currently features 13,605 cell markers for 467 cell types in 158 human tissue and 9148 cell markers for 389 cell types in 81 mouse tissues. The gene expression data in Cell-Marker originate from scRNA-seq studies, experimental studies, and microarrays. At approximately the same time, PanglaoDB was published[45], providing data on 8230 markers collected from human and mouse scRNA-seq experiments, along with other types of data. The markers are grouped by cell type (178 cell types, 4644 genes, and 29 tissues), and the cell types are subsequently grouped into 26 organs and 3 germ layers. A typical cell type has 28 (median) gene markers, but some, such as fibroblasts, have more than 100 markers. The major drawback of this database is that it lacks the source from which the marker information was obtained and information on how the marker was originally found and used, thus making information verification impossible.

Limitations of the current methods of cell-type classification

Despite the abovementioned databases, the new era of single-cell sequencing, which has revealed that markers can be expressed in different cells or at varying levels when cells are cultured in vitro, demands the reconsideration of marker-based cell-type classification methods. For example, the identification of mesenchymal stromal cells (MSCs) and the results of fate mapping in vivo have been problematic[46]. In addition, many markers often correspond to several cell types rather than a unique one, which may be related to the fact that MSCs encompass a number of different tissue-specific progenitor or stem cells. In other words, some markers are of mixed cell types[47,48]. Moreover, the expression levels of some markers make it difficult to discriminate among cell types[49]. For example, mature monocytes are usually characterized by the expression of CD33, CD11b, CD14, HLA-DR, and CD16, whereas granulocytes are characterized by the expression of CD33, CD11b, CD15, and CD66b. However, CD15 is expressed at low levels on monocytes in some anti-CD15 clones. Conversely, in disease states, CD14 can be variably expressed in neutrophils[50]. Notably, current markers all share the common feature that they are expressed on the cell surface. Intracellular markers such as microRNA (miRNA) could enhance the specific marker profile of a given cell type[51]. The combination of surface and intracellular markers for cell-type classification has yet to be fully developed in any database.
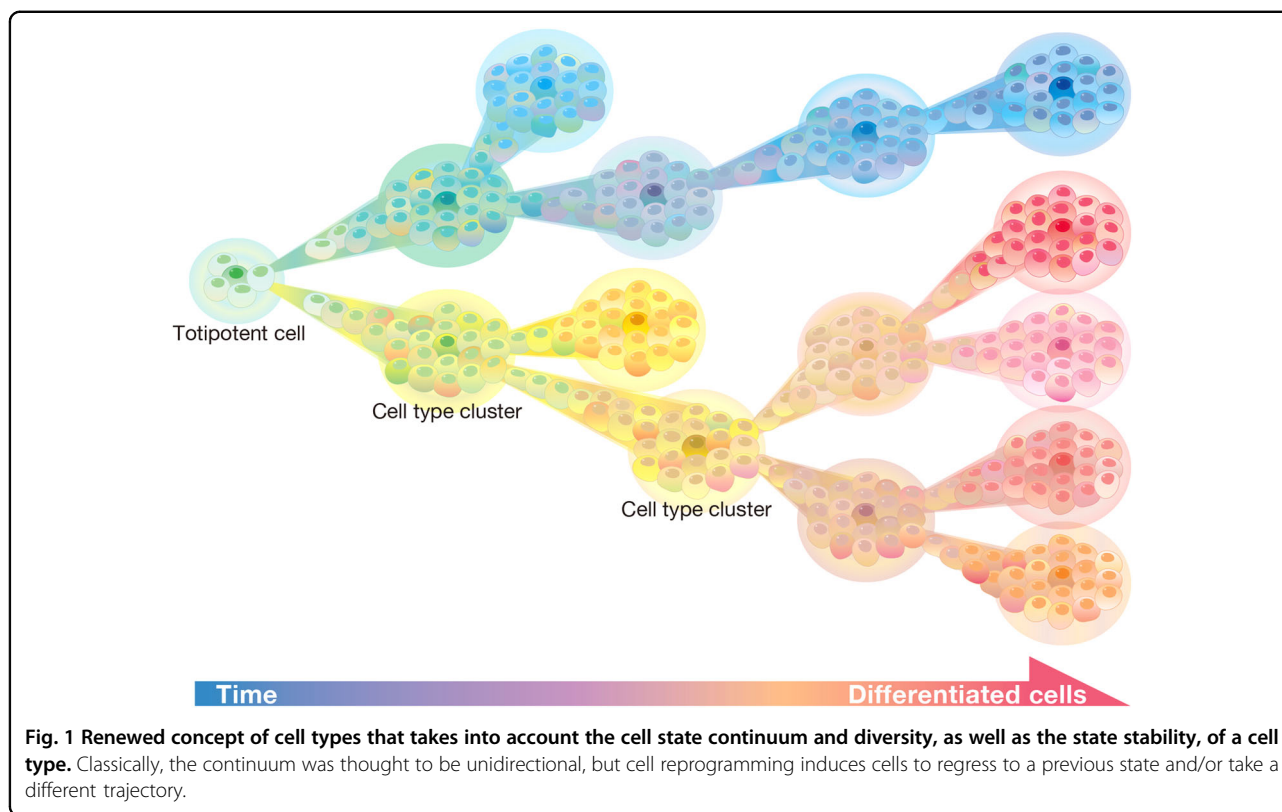
Another problem is the heterogeneity of cell states. In some cases, heterogeneity, such as that during states of immaturity or senescence, blurs discrete states, leaving a continuous spectrum that greatly complicates cell-type classification (Fig. 1). In such cases, cell types may seem to have hundreds of variants[52,53]. This lack of clarity is a major issue in clinical applications using cell therapies, for which cells are cultured in vitro to generate and stabilize a specific cell type. For example, for Parkinson's disease, the differentiation of pluripotent stem cells (PSCs) into specific neural cell types for cell therapies demands extremely precise markers[54]. As with their use as intercellular markers, the application of miRNA would be helpful. Another important problem involves changes in cell behavior and cell composition under different conditions or even within the same culture. These changes compromise stable cell phenotypes across cell populations over the course of an experiment, leading to confusing experimental interpretations[55].

Emerging concepts for cell-type classification

Accordingly, several new concepts for cell-type authentication have been proposed. Evolutionary biologists have developed a classification on the basis of the evolution of gene expression states[56]. In this proposition, the cell type is defined by a set of changes in the "core regulatory complex" (CoRC) of transcription factors that regulate cell-type-specific traits. From this point of view, cell types are defined by evolutionary units that differ according to their evolutionary lineages rather than their phenotypic similarities and are characterized by their ability to evolve gene expression states independently of each other. Thus, a gene regulatory network that defines the cell type would include "master" transcription factors that control downstream effector genes[57–59]. A classic example is the transformation of fibroblasts into skeletal muscle cells by the forced expression of the myoblast

Table 1   Online cell marker databases as of November 1, 2019.

Database	No. of markers	Cell type	Species	Characteristic	PMID	Address
Labome	226	7 major cell types and several others	Human	No distinction between diseased or healthy states. Few cell types are covered compared with other databases in this table. Source references are provided. No search engine. Only experimentally confirmed markers are included. This collection is most useful for searching verified markers.	N/A[42]	https://www.labome.com/method/Cell-Markers.html
CellFinder	553,905	3394 cell types and 50,951 cell lines	Human, mouse	Omics data are prioritized. In many cases, users must analyze raw data. No distinction between verified markers and ambiguous markers. This database is most useful for omics data exploration.	PMC3965082 [43]	http://www.cellfinder.org/
CellMarker	22,753	467 human cell types and 389 mouse cell types	Human, mouse	A major portion of markers are for cancer cells. No distinction between experimentally confirmed markers and ones gathered through omics. No distinction between verified markers and ambiguous markers. References are provided. This database is recommended for a first search of information.	PMC6323899 [44]	http://biocc.hrbmu.edu.cn/CellMarker/
PanglaoDB	8230	178 cell types	Human, mouse	Markers for healthy and diseased adult and embryonic cells. Source references are not provided for the markers. Contains single-cell sequencing data and automatically updates the list of papers on single-cell sequencing. Because of the relatively small number of cell types, this database is good for narrowing down the list of markers given by other databases.	PMC6450036 [45]	https://panglaodb.se/markers.html
SHOGoiN	2594	740 cell types	Human	Markers for healthy human cells including embryonic cells. Source references are provided. No distinction between verified markers and ambiguous markers. This database is most useful for a quick search by cell type.	PMC3204613 [81]	https://stemcellinformatics.org/cell_marker/literatures

Fig. 1 Renewed concept of cell types that takes into account the cell state continuum and diversity, as well as the state stability, of a cell type. Classically, the continuum was thought to be unidirectional, but cell reprogramming induces cells to regress to a previous state and/or take a different trajectory.

determination protein (MYOD) transcription factor[60]. The discovery of induced pluripotent stem cells (iPSCs) cemented the notion that master transcription factors can determine cell type because, through reprogramming to iPSCs and their subsequent differentiation, nearly any cell type can be converted into any other cell type[61,62]. Later works in neuronal lineages[63], the transcription factor competition observed in embryonic stem cells[64], and pancreatic cell transdifferentiation[65] support this idea.

Another evolutionary approach for cell-type classification is the construction of a hierarchy to describe the relationships between cells, which is analogous to how taxonomy hierarchies are created to describe the relationships between species[66]. Based on this approach, we proposed a "periodic table" for cell types[67]. This proposal aims to distinguish cell types from cell states, in which the periods and groups correspond to developmental trajectories and stages of differentiation[68]. scRNA-seq has paved the way for new interpretations of cell states. For example, in the epigenetic landscapes described by Waddington[69], it was originally assumed that cell states follow continuous trajectories that branch at cell-fate decision points, but these decision points have since been refined into transition states[70]. Whereas Waddington assumed that the decision points were deterministic, the transition states in the modernized version are stochastic and the related signaling networks are probabilistic. This

concept is meant to accommodate the gene expression heterogeneity encountered in real-world single-cell data. Indeed, the identification and characterization of cell transition states is one of the biggest challenges in single-cell transcriptomics[71]. Dimension reduction techniques, such as principal components analysis (PCA)[72] and t-distributed stochastic neighbor embedding (tSNE)[73], and graph and community detection algorithms, such as consensus clustering[74], SNN-Cliq[75], and Seurat[18], have been developed and utilized to identify cell transition states. Fully comprehending the influence of cell transition states on cell types at the single-cell level will require both new tools and parameters[76]. Finally, although the results are still in the rudimentary stage, recent research has made automated cluster annotation available for unbiased cell-type identification. This type of annotation is rapid and allows investigators to forgo the manual clustering step by combining annotation and clustering. Numerous tools based on various algorithms have been developed to facilitate automated cell identification methods (for a detailed review, see Abdelaal et al.[77]).

Proposing a data-driven cell-type definition

To help consolidate the many opinions about cell-type classification and provide data-related guidelines for cell-type authentication for clinical application, the International Cell Type Authentication Committee (ICTAC) was

created (https://cell-type.org/). The ICTAC was launched to establish criteria and processes for defining, determining, and authenticating all human cell types. Its mission is to help scientific communities identify cell types and provide systematic information on cell classifications. The ICTAC originated from the International Stem Cell Banking Initiative (ISCBI) (https://www.iscbi.org/)[78], an organization that focuses on practical issues in cell banking and regenerative medicine. More than 300 stem cell and policy professionals from 28 countries are part of the ISCBI community and are working together to advance stem cell research and biobanking along with developing regulations and public policy. The ISCBI is managed by an executive board, with delegates and steering group members of the community closely collaborating with the International Human Pluripotent Stem Cell Registry (hPSCreg)[79,80]. The first task of the ICTAC is to integrate existing cell databases, such as SHOGoiN[81], CellFinder[43], and Cell Ontology[82], to comprehend information on existing cell-type definitions. One of the key functions of the ICTAC is to provide a tool that can process the massive amount of accumulating single-cell data to classify new cell types that do not fit into existing definitions or cannot be identified by cell-matching software, such as CellSim, CellAtlasSearch, or Cell BLAST.

The ICTAC proposed the concept of reference cell types (RCTs), which are defined by an integrative examination of core properties (species, physiological system, source age, and markers) and additional attributes (functions including potency, morphology, developmental origin, omics, and environmental conditions) and are constantly updated by experts of relevant tissues and organs. RCTs provide a framework for identifying and authenticating new and known cell types. Ultimately, RCTs are designed to support cell-type classifications in various communities, including stem cell banking initiatives and massive-scale single-cell sequencing projects.

Cell type authentication for regenerative medicine

In the field of regenerative medicine, PSCs such as the aforementioned iPSCs have tremendous potential because of their capacity to differentiate into most cell types in the human body. Moreover, recent technological progress has made it possible to produce PSCs on a large scale[83] and generate significant resources for a large number of well-characterized and documented PSCs (e.g., https://fujifilmcdi.com/the-cirm-ipsc-bank, https://ebisc.org[84]). However, in addition to scalability, the clinical translation of PSCs depends on fast and reliable ways to assess quality and safety in terms of three important properties: cellular identification, differentiation potency, and malignant potential[85].

For example, the establishment of PSC-derived platelets as a substitute for primary donor cells can compensate for anticipated donor shortages and are useful against platelet transfusion refractoriness[86]. In this system, PSCs need to be differentiated into megakaryocytes, which are then cultured in bioreactors to shed platelets[87]. However, platelet production by iPSC-derived megakaryocytes is heterogeneous, and many biochemical and biophysical approaches have been attempted to enhance and homogenize their production[88]. Ultimately, the characterization of the best megakaryocytes is lacking. In other words, the existing classification of the megakaryocytes is too broad to identify the cells that are optimal for platelet production, thereby resulting in inefficient platelet production. This problem could be solved by identifying markers associated with platelet-producing megakaryocytes and developing differentiation protocols aimed at the selection of the relevant cell types.

Differentiation efficacy depends on PSC quality. There are several methods for ensuring high PSC quality that are based on assessing the potency of PSC differentiation into cells of the three germ layers. The gold standard is the teratoma assay[89,90], in which the differentiation capacity of a PSC line in vivo is assessed by grafting the cells into immunodeficient mice. In addition to differentiation capacity, this assay enables the assessment of viability, histotypic organization, and carcinogenicity at the same time. However, the assay is lengthy and laborious, and it requires experts in pathological assessment and the use of experimental animals[91,92]. Furthermore, in a comparison of the methods used for and the results obtained from teratoma assays performed at 18 centers worldwide, the ISCBI found that both the test methods and test results varied substantially among expert centers[93]. Another quality check method is the embryoid body (EB) assay, which enables the monitoring of differentiation capacity in vitro. EBs are cell aggregates that spontaneously differentiate into three developmental germ layers when cultured in suspension[94]. This approach is more standardized than the teratoma assay and considered more robust by some researchers[95]. Contributing to its favorability, an EB assay can be combined relatively easily with a bioinformatics analysis of gene expression profiles[96–98].

Finally, the detection of pluripotency-specific markers, such as alkaline phosphatase[99], Nanog, and Oct4, as well as other mRNAs and proteins, is another way to perform a quality check of PSC characteristics. The detection of these markers is usually performed using flow cytometry, RT-qPCR, and cell-staining techniques[100]. A number of markers can be used to identify PSC types, such as the naive PSC state or the high-yield expansion PSC state[101–104], and to measure the quality of the PSCs[105]. At the single-cell level, pluripotency can be assessed with scRNA-seq and bioinformatics tools. For example, a

PluriTest® assay is used to distinguish typical PSCs from other PSC-like populations through machine learning that is based on the transcriptomes of ideal cell lines and control PSC lines[106]. However, functional differentiation assays are still required to exclude false-positive results. Other tools, such as SLICE[107], SCENT[108], and Epi-Pluri-Score[109], allow researchers to quantify cell potency and cellular differentiation by entropy analysis. However, a recent multinational study of a range of pluripotency assays concluded that the demonstration of at least some capacity for in vitro or in vivo differentiation was important for the veracity of the results[85].

Guidelines for big data generation, storage, and management

As described above, numerous groups are working internationally to generate human omics data not only for tissues and bulk cells but also for millions of individual cells. The gathered data have the potential to greatly advance experimentation practices and quality checks for cells intended for clinical applications. However, for this goal to be realized, the data must be structured, and new algorithms that can efficiently curate the data are needed. Accordingly, members of the ISCBI have proposed Minimum Information About a Cellular Assay for Regenerative Medicine (MIACARM) guidelines, which are directed first and foremost at stem cell banks, although MIACARM can also be used to structure data formats of cellular assays in general[110]. MIACARM is based on the Minimum Information About a Cellular Assay (MIACA), the first attempt at creating a reporting format for describing functional research on cell lines[111]. Unlike MIACA, MIACARM sets guidelines for human cells used in medical applications or single-cell analysis, including omics. The proposed guidelines have the potential to enhance information flow in stem cell research that aims to produce clinical-grade cells for therapies. As an extension of MIACARM, which currently targets source cell characterization (MIACARM-I) and stem cell characterization (MIACARM-II), the ISCBI community is engaged in plans to provide guidelines for characterizing differentiated cells (MIACARM-III) using ICTAC proposals for cell authentication based on the framework developed for RCTs.

Conclusion

It is obvious that the body constitutes a myriad of different cell types that have specific functions. Advances in microscopy, histology, and, now, omics technologies have made it clear that the list of cell types is much longer than originally imagined and that our understanding of molecular expressions in systems both in vitro and in vivo is still at the nascent stage. Moreover, the concept of cell reprogramming has taught us that cells can be in constant flux, oscillating between cell types. The ability to define these different cell types is crucial for understanding natural development, including the development of diseased states, and for producing cells for clinical therapies. Critical points are a consensus definition of cell types and the data required for their authentication. In addition, it is vital to ensure accurate and traceable links between precious resources of biological materials and the associated data sets to make full use of both biological and electronic resources and promote reproducibility in scientific data. The many existing databases and the massive data already accumulated affirm the need for the scientific community to work together in creating a universal standard.

Author details
[1]Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. [2]BIH Center for Regenerative Therapies (BCRT), Charité—Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. [3]International Stem Cell Banking Initiative, 2 High Street, Barley, Herts SG88HZ, UK. [4]National Stem Cell Resource Centre, Institute of Zoology, Chinese Academy of Sciences, 100190 Beijing, China. [5]Innovation Academy for Stem Cell and Regeneration, Chinese Academy of Sciences, 100101 Beijing, China

Conflict of interest
The authors declare that they have no conflict of interest.

References
1. Pullen, L. C. Human cell atlas poised to transform our understanding of organs. *Am. J. Transplant.* **18**, 1–2 (2018).
2. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
3. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
4. Popescu, D-M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
5. Ledergor, G. et al. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat. Med.* **24**, 1867–1876 (2018).
6. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
7. Smillie, C. S. et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e22 (2019).

8. Lukowski, S. W. et al. A single-cell transcriptome atlas of the adult human retina. *EMBO J.* **38**, e100811 (2019).

9. Regev, A. et al. The human cell atlas. *Elife* **6**, e27041 (2017).

10. Rosa, F., Kurochkin, I., Pires, C. & Pereira, F. HCA census of immune cells. https://data.humancellatlas.org/explore/projects/116965f3-f094-4769-9d28-ae675c1b569c (2019).

11. Ponting, C. P. The Human Cell Atlas: making "cell space" for disease. *Dis. Model. Mech.* **12**, dmm037622 (2019).

12. Prakadan, S. M., Shalek, A. K & Weitz, D. A. Scaling by shrinking: empowering single-cell "omics" with microfluidic devices. *Nat. Rev. Genet.* **18**, 345–361 (2017).

13. Mezger, A. et al. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat. Commun.* **9**, 3647 (2018).

14. Saelens, W., Cannoodt, R. & Saeys, Y. A comprehensive evaluation of module detection methods for gene expression data. *Nat. Commun.* **9**, 1090 (2018).

15. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).

16. Li, P. et al. The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **42**, 1060–1067 (2010).

17. Reeb, P. D., Bramardi, S. J. & Steibel, J. P. Assessing dissimilarity measures for sample-based hierarchical clustering of RNA sequencing data using plasmode datasets. *PLoS ONE* **10**, e0132310 (2015).

18. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

19. Petegrosso, R., Li, Z. & Kuang, R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. Brief. *Bioinformatics* https://doi.org/10.1093/bib/bbz063 (2019).

20. Srivastava, D., Iyer, A., Kumar, V. & Sengupta, D. CellAtlasSearch: a scalable search engine for single cells. *Nucleic Acids Res.* **46**, W141–W147 (2018).

21. Li, L. et al. CellSim: a novel software to calculate cell similarity and identify their co-regulation networks. *BMC Bioinform.* **20**, 111 (2019).

22. Cao, Z.-J., Wei, L., Lu, S., Yang, D.-C. & Gao, G. Cell BLAST: searching large-scale scRNA-seq database via unbiased cell embedding. *BioRxiv* https://doi.org/10.1101/587360 (2019).

23. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

24. Achim, K. et al. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).

25. Durruthy-Durruthy, R., Gottlieb, A. & Heller, S. 3D computational reconstruction of tissues with hollow spherical morphologies using single-cell gene expression data. *Nat. Protoc.* **10**, 459–474 (2015).

26. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).

27. Durruthy-Durruthy, J. et al. Spatiotemporal reconstruction of the human blastocyst by single-cell gene-expression analysis informs induction of naive pluripotency. *Dev. Cell* **38**, 100–115 (2016).

28. Li, J. et al. Systematic reconstruction of molecular cascades regulating GP development using single-cell RNA-seq. *Cell Rep.* **15**, 1467–1480 (2016).

29. Mori, T. et al. Development of 3D tissue reconstruction method from single-cell RNA-seq data. *Genomics Comput. Biol.* **3**, 53 (2017).

30. Mori, T., Takaoka, H., Yamane, J., Alev, C. & Fujibuchi, W. Novel computational model of gastrula morphogenesis to identify spatial discriminator genes by self-organizing map (SOM) clustering. *Sci. Rep.* **9**, 12597 (2019).

31. Masters, J. R. Cell-line authentication: end the scandal of false cell lines. *Nature* **492**, 186 (2012).

32. Clevers, H. et al. What is your conceptual definition of "cell type" in the context of a mature organism? *Cell Syst.* **4**, 255–259 (2017).

33. Brennecke, P. et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).

34. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* **14**, e1006245 (2018).

35. Hon, C.-C., Shin, J. W., Carninci, P. & Stubbington, M. J. T. The Human Cell Atlas: technical approaches and challenges. *Brief. Funct. Genomics* **17**, 283–294 (2018).

36. Junqueira, L. C., Carneiro, J. & Kelly, R. O. *Histologi Dasar (Basic Histology)* (EGC Penebrit Buku Kedokteran, 1980).

37. Ramon y Cajal, S. *Histologie du système nerveux de l'homme & des vertébrés* (Maloine, 1909).

38. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* **8**, 427–437 (2007).

39. Klausberger, T. & Somogyi, P. Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science* **321**, 53–57 (2008).

40. DeFelipe, J. et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat. Rev. Neurosci.* **14**, 202–216 (2013).

41. Sugino, K. et al. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat. Neurosci.* **9**, 99–107 (2006).

42. Yakimchuk, K. Cell markers. *Mater. Methods* **3**, 183 (2013).

43. Stachelscheid, H. et al. CellFinder: a cell data repository. *Nucleic Acids Res.* **42**, D950–D958 (2014).

44. Zhang, X. et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721–D728 (2019).

45. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, baz046 (2019).

46. Zhou, B. O., Yue, R., Murphy, M. M., Peyer, J. G. & Morrison, S. J. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. *Cell Stem Cell* **15**, 154–168 (2014).

47. Debnath, S. et al. Discovery of a periosteal stem cell mediating intramembranous bone formation. *Nature* **562**, 133–139 (2018).

48. Zhang, J. & Link, D. C. Targeting of mesenchymal stromal cells by Cre-Recombinase transgenes commonly used to target osteoblast lineage cells. *J. Bone Miner. Res.* **31**, 2001–2007 (2016).

49. Gustafson, M. P. et al. A method for identification and analysis of non-overlapping myeloid immunophenotypes in humans. *PLoS ONE* **10**, e0121546 (2015).

50. Wagner, C. et al. Expression patterns of the lipopolysaccharide receptor CD14, and the FCgamma receptors CD16 and CD64 on polymorphonuclear neutrophils: data from patients with severe bacterial infections and lipopolysaccharide-exposed cells. *Shock* **19**, 5–12 (2003).

51. Miki, K. et al. Efficient detection and purification of cell populations using synthetic microrna switches. *Cell Stem Cell* **16**, 699–711 (2015).

52. Markram, H. et al. Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793–807 (2004).

53. Parra, P., Gulyás, A. I. & Miles, R. How many subtypes of inhibitory cells in the hippocampus? *Neuron* **20**, 983–993 (1998).

54. Takahashi, J. Stem cells and regenerative medicine for neural repair. *Curr. Opin. Biotechnol.* **52**, 102–108 (2018).

55. Greenblatt, M. B., Ono, N., Ayturk, U. M., Debnath, S. & Lalani, S. The unmixing problem: a guide to applying single-cell RNA sequencing to bone. *J. Bone Miner. Res.* **34**, 1207–1219 (2019).

56. Arendt, D. et al. The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

57. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).

58. Saint-André, V. et al. Models of human core transcriptional regulatory circuitries. *Genome Res.* **26**, 385–396 (2016).

59. Mullen, A. C. et al. Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell* **147**, 565–576 (2011).

60. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).

61. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

62. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**, 101–106 (2008).

63. Hobert, O. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl. Acad. Sci. USA* **105**, 20067–20071 (2008).

64. Sokolik, C. et al. Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst.* **1**, 117–129 (2015).

65. van der Meulen, T. & Huising, M. O. Role of transcription factors in the transdifferentiation of pancreatic islet cells. *J. Mol. Endocrinol.* **54**, R103–R117 (2015).

66. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017).

67. Sakurai, K. & Fujibuchi, W. Close relationships between iPS cells and Periodic Table of chemical elements. *Trans. Res. Inst. Oceanochem* **29**, 17–23 (2016).

68. Xia, B. & Yanai, I. A periodic table of cell types. *Development* **146**, dev169854 (2019).

69. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).

70. Moris, N., Pina, C. & Arias, A. M. Transition states and cell fate decisions in epigenetic landscapes. *Nat. Rev. Genet.* **17**, 693–703 (2016).

71. Moignard, V. et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–276 (2015).

72. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).

73. Weinreb, C., Wolock, S. & Klein, A. M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018).

74. Kiselev, V. Y. et al. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).

75. Xu, C. & Su, Z. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* **31**, 1974–1980 (2015).

76. Zheng, X., Jin, S., Nie, Q. & Zou, X. scRCMF: identification of cell subpopulations and transition states from single cell transcriptomes. *IEEE Trans. Biomed. Eng.* https://doi.org/10.1109/TBME.2019.2937228 (2019).

77. Abdelaal, T. et al. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol.* **20**, 194 (2019).

78. Crook, J. M., Hei, D. & Stacey, G. N. The International Stem Cell Banking Initiative (ISCBI): raising standards to bank on. *Vitr. Cell Dev. Biol. Anim.* **46**, 169–172 (2010).

79. Seltmann, S. et al. hPSCreg-the human pluripotent stem cell registry. *Nucleic Acids Res.* **44**, D757–D763 (2016).

80. Kurtz, A. et al. A standard nomenclature for referencing and authentication of pluripotent stem cells. *Stem Cell Rep.* **10**, 1–6 (2018).

81. Hatano, A. et al. CELLPEDIA: a repository for human cell information for cell studies and differentiation analyses. *Database* **2011**, bar046 (2011).

82. Diehl, A. D. et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semant.* **7**, 44 (2016).

83. Jenkins, M. J. & Farid, S. S. Human pluripotent stem cell-derived products: advances towards robust, scalable and cost-effective manufacturing strategies. *Biotechnol. J.* **10**, 83–95 (2015).

84. De Sousa, P. A. et al. Rapid establishment of the European Bank for induced Pluripotent Stem Cells (EBiSC)—the Hot Start experience. *Stem Cell Res.* **20**, 105–114 (2017).

85. International Stem Cell Initiative. Assessment of established techniques to determine developmental and malignant potential of human pluripotent stem cells. *Nat. Commun.* **9**, 1925 (2018).

86. Sugimoto, N. & Eto, K. Platelet production from induced pluripotent stem cells. *J. Thromb. Haemost.* **15**, 1717–1727 (2017).

87. Ito, Y. et al. Turbulence activates platelet biogenesis to enable clinical scale ex vivo production. *Cell* **174**, 636–648.e18 (2018).

88. Karagiannis, P., Sugimoto, N. & Eto, K. in *Platelets*, 4th edn (ed. Michelson, A. D.) 1173–1189 (Academic Press, 2019).

89. Gropp, M. et al. Standardization of the teratoma assay for analysis of pluripotency of human ES cells and biosafety of their differentiated progeny. *PLoS ONE* **7**, e45532 (2012).

90. International Stem Cell Banking Initiative. Consensus guidance for banking and supply of human embryonic stem cell lines for research purposes. *Stem Cell Rev.* **5**, 301–314 (2009).

91. Buta, C. et al. Reconsidering pluripotency tests: do we still need teratoma assays? *Stem Cell Res.* **11**, 552–562 (2013).

92. Ellis, J. et al. Alternative induced pluripotent stem cell characterization criteria for in vitro applications. *Cell Stem Cell* **4**, 198–199 (2009). author reply 202.

93. Andrews, P. W. et al. Points to consider in the development of seed stocks of pluripotent stem cells for clinical applications: International Stem Cell Banking Initiative (ISCBI). *Regen. Med.* **10**, 1–44 (2015).

94. Kurosawa, H. Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells. *J. Biosci. Bioeng.* **103**, 389–398 (2007).

95. Sheridan, S. D., Surampudi, V. & Rao, R. R. Analysis of embryoid bodies derived from human induced pluripotent stem cells as a means to assess pluripotency. *Stem Cells Int.* **2012**, 738910 (2012).

96. Ng, E. S., Davis, R. P., Azzola, L., Stanley, E. G. & Elefanty, A. G. Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. *Blood* **106**, 1601–1603 (2005).

97. Bock, C. et al. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).

98. Avior, Y., Biancotti, J. C. & Benvenisty, N. TeratoScore: assessing the differentiation potential of human pluripotent stem cells by quantitative expression analysis of teratomas. *Stem Cell Rep.* **4**, 967–974 (2015).

99. O'Connor, M. D. et al. Alkaline phosphatase-positive colony formation is a sensitive, specific, and quantitative indicator of undifferentiated human embryonic stem cells. *Stem Cells* **26**, 1109–1116 (2008).

100. O'Connor, M. D., Kardel, M. D. & Eaves, C. J. Functional assays for human embryonic stem cell pluripotency. *Methods Mol. Biol.* **690**, 67–80 (2011).

101. Collier, A. J. & Rugg-Gunn, P. J. Identifying human naïve pluripotent stem cells—evaluating state-specific reporter lines and cell-surface markers. *Bioessays* **40**, e1700239 (2018).

102. Messmer, T. et al. Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. *Cell Rep.* **26**, 815–824.e4 (2019).

103. Ghimire, S. et al. Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Sci. Rep.* **8**, 5884 (2018).

104. Lipsitz, Y. Y., Woodford, C., Yin, T., Hanna, J. H. & Zandstra, P. W. Modulating cell state to enhance suspension expansion of human pluripotent stem cells. *Proc. Natl. Acad. Sci. USA* **115**, 6369–6374 (2018).

105. Ungrin, M., O'Connor, M., Eaves, C. & Zandstra, P. W. Phenotypic analysis of human embryonic stem cells. *Curr. Protoc. Stem Cell Biol.* **Chapter 1**, Unit 1B.3 (2007).

106. Müller, F.-J. et al. A bioinformatic assay for pluripotency in human cells. *Nat. Methods* **8**, 315–317 (2011).

107. Guo, M., Bao, E. L., Wagner, M., Whitsett, J. A. & Xu, Y. SLICE: determining cell differentiation and lineage based on single cell entropy. *Nucleic Acids Res.* **45**, e54 (2017).

108. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat. Commun.* **8**, 15599 (2017).

109. Lenz, M. et al. Epigenetic biomarker to support classification into pluripotent and non-pluripotent cells. *Sci. Rep.* **5**, 8973 (2015).

110. Sakurai, K., Kurtz, A., Stacey, G. N., Sheldon, M. & Fujibuchi, W. First proposal of minimum information about a cellular assay for regenerative medicine. *Stem Cells Transl. Med.* **5**, 1345–1361 (2016).

111. Wiemann, S., Mehrle, A. & Hahne, F. MIACA—minimum information about a cellular assay, and the cellular assay object model http://miaca.sourceforge.net/ (2015).