

# PCMDB: a curated and comprehensive resource of plant cell markers

Jingjing Jin<sup>1,†</sup>, Peng Lu<sup>1,†</sup>, Yalong Xu<sup>1</sup>, Jiemeng Tao<sup>1</sup>, Zefeng Li<sup>1</sup>, Shuaibin Wang<sup>2</sup>, Shizhou Yu<sup>3</sup>, Chen Wang<sup>1</sup>, Xiaodong Xie<sup>1</sup>, Junping Gao<sup>2</sup>, Qiansi Chen<sup>1</sup>, Lin Wang<sup>1</sup>, Wenxuan Pu<sup>2</sup> and Peijian Cao<sup>1,\*</sup>

<sup>1</sup>China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China, <sup>2</sup>China Tobacco Hunan Industrial Co., Ltd., Changsha 410007, China and <sup>3</sup>Molecular Genetics Key Laboratory of China Tobacco, Guizhou Academy of Tobacco Science, Guiyang 550081, China

Received July 30, 2021; Revised September 29, 2021; Editorial Decision September 30, 2021; Accepted October 02, 2021

## ABSTRACT

The advent of single-cell sequencing opened a new era in transcriptomic and genomic research. To understand cell composition using single-cell studies, a variety of cell markers have been widely used to label individual cell types. However, the specific database of cell markers for use by the plant research community remains very limited. To overcome this problem, we developed the Plant Cell Marker DataBase (PCMDB, <http://www.tobacodb.org/pcmdb/>), which is based on a uniform annotation pipeline. By manually curating over 130 000 research publications, we collected a total of 81 117 cell marker genes of 263 cell types in 22 tissues across six plant species. Tissue- and cell-specific expression patterns can be visualized using multiple tools: eFP Browser, Bar, and UMAP/TSNE graph. The PCMDB also supports several analysis tools, including SCSA and SingleR, which allows for user annotation of cell types. To provide information about plant species currently unsupported in PCMDB, potential marker genes for other plant species can be searched based on homology with the supported species. PCMDB is a user-friendly hierarchical platform that contains five built-in search engines. We believe PCMDB will constitute a useful resource for researchers working on cell type annotation and the prediction of the biological function of individual cells.

## INTRODUCTION

Single-cell sequencing has revolutionized biological research, enabling the characterization of cell types across multiple species, tissues, and cells (1). A large number of

single-cell RNA sequencing (scRNA-seq) studies have been performed to uncover cell lineage relationships across plant tissues, including roots (2–11), leaves (7,12), shoot apical meristems (13), ears (14), seedlings (7) and anther germinal cells (15). This technique has also been used to explore cell-specific transcriptional responses to environmental stimuli, such as low-nitrogen, high-salinity, and iron-deficient environments (7), or the presence of heat or sugar (2,11). Some studies have even attempted to reveal the evolution of cell trajectories between different plants (3).

Cell markers are genes that are highly expressed in one cell type, but lowly expressed or not expressed in other types. To understand cell composition using single-cell studies, one approach is to use many known cell markers (supervised method) to ‘vote’ on the identity of a cell in question (16). There are a variety of known cell markers that are extremely valuable for cell type annotation, advancing our understanding of cell composition. Experimental evidence on cell markers has accumulated over the years, and has been discovered using various techniques, including qRT-PCR, green fluorescent protein (GFP) reporter systems, beta-glucuronidase (GUS) gene reporter systems and western blotting. In recent decades, cell markers can be also identified by significant differential expression between cell groups in transcriptome studies, including bulk RNA-seq/microarray or scRNA-seq studies (4,17). However, these cell markers are scattered over thousands of literatures, which would make it difficult to collect for specific cells for researchers.

The CellMarker database provides comprehensive cell marker resources for humans and mice (18). However, analogous database for use by the plant research community remains very limited. Some current plant scRNA-seq-related databases focus mainly on scRNA-seq data from a specific plant species. For example, the PscB database only contains information for root tissue in *Arabidopsis* (19). The recently available PlantscRNAdb database collects public

\*To whom correspondence should be addressed. Tel: +86 371 67672071; Fax: +86 371 67672071; Email: peijiancao@163.com

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

data for marker genes in four plant species (20). However, this database still has some limitations. It lacks standardized annotation for tissues and cells, and information about their hierarchical structure. Data from this database may therefore not be compatible across different plant species. The experimental marker candidates were curated from the literature by searching for the keywords ‘marker(s)’ and ‘specific expression’, which limits the available resources (Supplementary Table S1). PlantscRNAdb lacks some important features, including sequence information and expression evidence from bulk RNA-seq analysis, which makes further studies more difficult. Finally, PlantscRNAdb focuses mainly on the search for and visualization of marker genes and contains very few additional tools. In summary, the establishment of a plant cell marker-related database with more uniform and strict annotation standards, greater data integrity, more comprehensive annotations, and additional analytical tools is highly desirable.

We developed PCMDB (the Plant Cell Marker DataBase) in order to create a one-stop plant cell marker database that uses standard collection criteria to provide a comprehensive overview of cell markers from experimental research, bulk tissue and cell RNA-seq studies, and single-cell sequencing studies. The current version of PCMDB includes 81 117 manually curated cell marker genes, 3 119 of which are experimental marker genes, across six plant species. These markers involve 263 cell types from 22 tissues. PCMDB includes the tools eFP Browser, Bar, and UMAP/TSNE graph, which can be used to visualize specific expression patterns of cell marker genes. PCMDB also supports two analytical tools based on marker gene information, specifically SCSA, a marker-based cell type prediction tool, and SingleR, for reference data-based cell type prediction (21,22). These tools allow cell type prediction of user-specific single-cell datasets. Our database also supports homology-based search, allowing users to predict potential marker genes for species currently unsupported in PCMDB. In summary, PCMDB constitutes a novel, valuable resource for cell type labeling and function prediction in single-cell sequencing studies.

## MATERIALS AND METHODS

### Data sources

PCMDB contains three different types of cell markers for six model plants—*Arabidopsis*, rice, maize, soybean, tomato and tobacco—including experimentally validated marker genes, differentially expressed marker genes based on bulk RNA-seq data, and differentially expressed genes across specific cells identified by single-cell RNA sequencing. Although some bulk RNA-seq studies were not restricted to a specific cell type, these data can provide additional information about the marker genes identified by the other two approaches. To ensure the standardization of identification of tissue and cell types, we unified the names of the tissues and cell types into a standard reference list based on the Plant Ontology database (23).

To obtain cell markers that have been experimentally verified, we searched PubMed for each species using cell names from Plant Ontology. We obtained a total of 125 490 publications, including 31 967 for *Arabidopsis*, 17 945 for rice,

16 552 for maize, 11 593 for soybean, 10 150 for tomato and 37 283 for tobacco (Supplementary Table S2, round 1). We manually checked the abstracts of these papers and filtered the studies including non-gene related work, foreign gene transgenic work in target plant and so on. Then, we generally checked the entire text and selected the studies containing cell markers or genes inferred through biological experiments, including GFP reporter systems, GUS gene reporter systems and western blotting. Expression patterns for these potential candidates were further checked using eFP Browser of each plant. This strategy decreased the number of publications to 2 883 for *Arabidopsis*, 996 for rice, 649 for maize, 336 for soybean, 277 for tomato and 272 for tobacco (Supplementary Table S2, round 2). We then further carefully checked the entire text and confirmed the annotation/function in the official plant websites, and finally extracted relevant information, including the cell marker name, cell marker ID, and cell/tissue type for each plant (Supplementary Table S2, round 3).

Considering the important contribution of cell marker identification using single-cell RNA sequencing, we searched the literature of PubMed using a list of key words, including ‘single cell’, ‘single cell sequencing’, ‘single cell RNA sequencing’, ‘single cell RNA-seq’, ‘single cell RNA seq’, ‘single cell transcriptome’, ‘single cell transcriptomics’ and ‘scRNA seq’ for each plant. This search allowed us to obtain a total of 45 single-cell sequencing-related publications. Of these, after manual curation for reviews and method-related studies, we retained 14 for *Arabidopsis*, 3 for rice, 3 for maize and 1 for tomato (Supplementary Table S2). We then carefully checked the entire manuscript and corresponding supplementary materials and extracted the relevant information for cell markers identified by the corresponding study for each plant.

We used an additional search strategy to identify differentially expressed cell markers for specific tissues that might be missing from the previous results. For each plant species, we used different keywords and combinations, including ‘genome wide/genome-wide’, ‘transcriptome’, ‘landscape’ and ‘global’, to search for bulk RNA-seq related manuscripts for different plants. In total, we obtained the following publications for different species: 11 011 for *Arabidopsis*, 3 169 for rice, 1 593 for maize, 1 821 for soybean, 609 for tomato and 1 727 for tobacco. We then performed a manual check on the abstracts and text, as described above, and kept only those publications containing useful information (Supplementary Table S2). Then, we carefully screened the entire manuscript and supplementary materials and extracted the differentially expressed markers for each tissue or cell.

The records obtained using the three different strategies were merged to create the final cell marker list.

### Data analysis

**Data processing and annotation.** During the process of collecting data from the publications, we carefully curated important information for each cell marker, including cell marker name, species, tissue type, cell type, source and corresponding publication information (PMID/DOI, journal, title and abstract). Cell marker ID/Name and

annotations were consistent with those available in official plant databases (*Arabidopsis*: TAIR (24); rice: MSU RGAP (25); maize: MaizeGDB (26); soybean: SoyBase (27) and tomato and tobacco: Sol Genomics Network (28)). With the exception of the official ID, other identifiers, such as NCBI Gene/RNA/Protein IDs, are also provided in order to facilitate future usability.

Based on the unification of tissue and cell types from the Plant Ontology database, we adopted a hierarchical structure for these different cell types. To allow for proper visualization of the hierarchical structure, only three levels were retained in the final list.

Finally, all of the information related to cell markers was stored in a MySQL database, as shown in Figure 1.

**Single-cell sequencing data analysis.** Because of the lack of available clustering results from single-cell studies, we re-analyzed these studies using a unified pipeline. For studies providing data accession numbers, we downloaded the raw data from the NCBI database (29) and processed the raw data to generate expression matrices. Only data generated using the 10X Genomics approach, which were almost 90% of the total data, were processed. Fastq-dump (v2.8.0) was used to convert the SRA data into fastq files (29). Then, clean reads were mapped to a reference genome using the 10X Genomics CellRanger (v6.0.1) with default parameters (30). All downstream analyses were performed with Seurat (v4.0.3) (31). In brief, the gene-cell matrices were loaded into the Seurat package implemented in R (v4.0.2). To remove low-quality cells, we filtered out the cells with unique gene counts fewer than 200. Genes expressed in at least three single cells were kept. The SCTransform function was used to scale and normalize the raw data. The top 5 000 highly variable genes were selected for downstream analysis. Using principal component (PC) analysis, the scaled data were reduced to 20 PCs by setting  $npcs = 20$ . Clusters were identified using the FindClusters function with a resolution of 0.5. In the case of multiple samples, datasets were combined into a single dataset using canonical correlation analysis with the IntegrateData function. The marker genes used in the corresponding study were used to label each cluster.

**Database development.** Similar with PLncDB V2.0 (32), PCMDB was constructed using Python (<https://www.python.org/>), Vue.js (<https://vuejs.org/>), ElementUI (<https://element.eleme.io/#/>), and Django (<https://www.djangoproject.com/>). Network proxy services were provided through nginx (<https://www.nginx.com/>). PCMDB can be accessed at <http://www.tobacodb.org/pcmdb/> and requires no registration.

## RESULTS

### Data content of PCMDB

A total of 81 117 cell markers are represented in the current version of PCMDB, including 19 260 for *Arabidopsis*, 19 359 for rice, 16 828 for maize, 12 357 for soybean, 12 198 for tomato and 1 115 for tobacco (Figure 1, Supplementary Table S1). Among them, 3 119 marker genes are supported by experimental evidence, including GFP, GUS and western

blotting (Figure 1, Supplementary Table S1). We derived these markers from 1 622 experimental, 58 bulk RNA-seq and 21 single-cell studies (Supplementary Tables S1 and S2), involved in a total of 263 cell types across 22 tissues (Supplementary Table S1). On average, around 17 experimental cell markers were available for each cell type, with the root cap containing the highest number of cell markers (Figure 1). The top five most frequently used cell markers in the root cap of *Arabidopsis* were *WOX5*, *EIR1*, *GL2*, *PLT1* and *LBD16* (Figure 2B). For bulk RNA-seq related cell markers, the seed contained the highest number of cell markers, although the root endodermis had the highest number of cell markers identified using scRNA-seq data (Supplementary Figures S1 and S2).

### Features of PCMDB

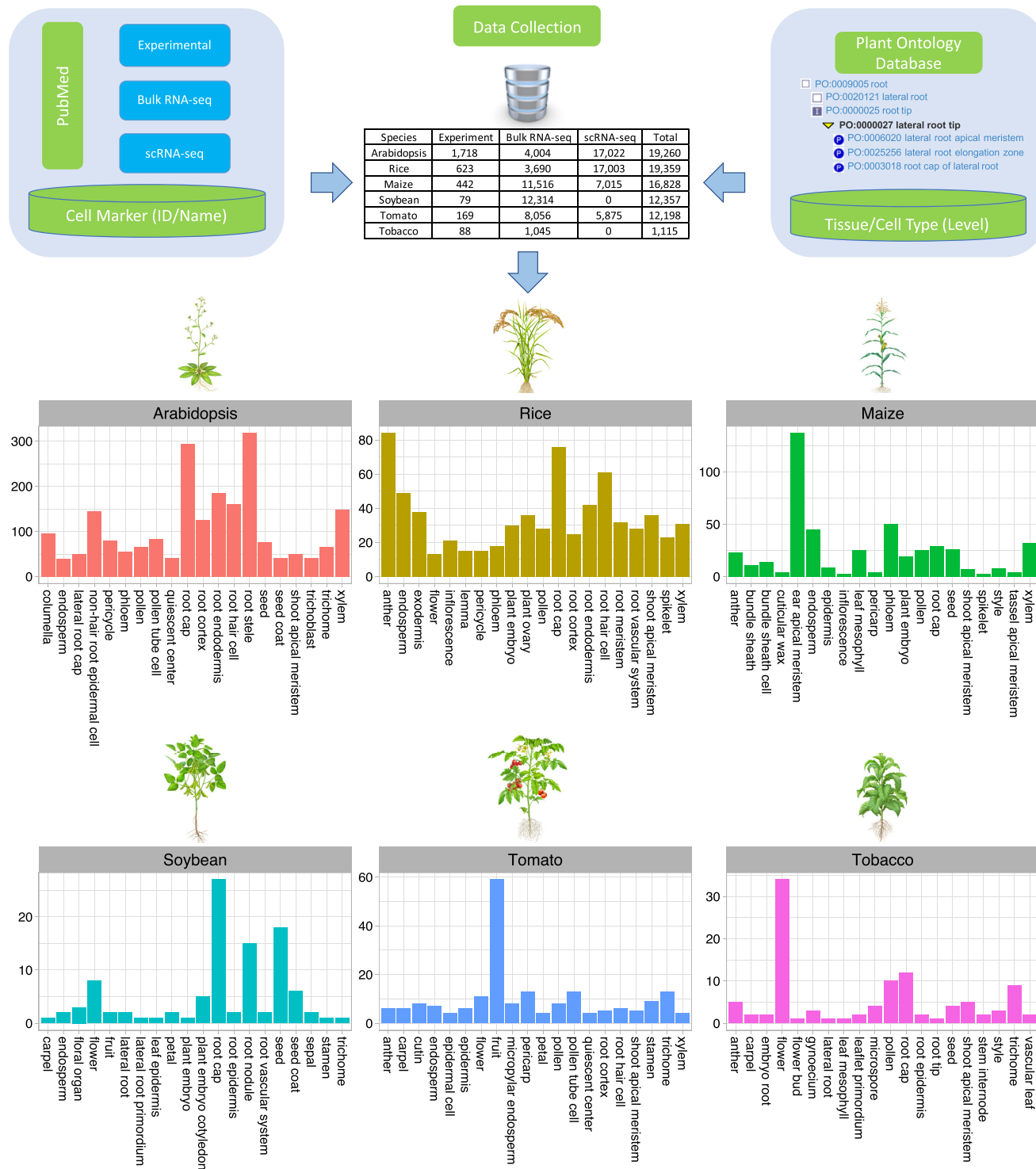
PCMDB includes all basic information for any given marker gene, including nucleotide and peptide sequences, genomic coordinates and putative function. Moreover, PCMDB contains several unique marker gene annotation features. First, users are able to find tissue- and cell-specific expression patterns for the different marker genes, using the eFP Browser (Figure 2D). Second, the detailed expression information of candidate markers predicted from tissue/cell-specific RNA-seq data can be visualized using bar plots (Figure 2E). Third, PCMDB also provides expression matrix and cluster results for each of the different scRNA-seq studies used, and users are able to easily obtain an overview of the gene expression information among each cell of a specific tissue (Figure 2F). Moreover, in order to meet the high demand for personalized cell type annotation of scRNA-seq studies, PCMDB also supports several additional tools that facilitate different research purposes, including marker-based cell type prediction (SCSA) (Figure 3A) and reference data-based cell type prediction (SingleR) (Figure 3B). To compensate for the lack of other plant species in PCMDB, we support cell marker search based on homology, which allows the prediction of potential marker genes in different plant species (Figure 3C).

### PCMDB functions

PCMDB provides convenient access to functional search engines and powerful analytical tools (Figure 2C). Users can browse the data using shortcuts and multiple webpage layers. All data can be downloaded in a customized fashion.

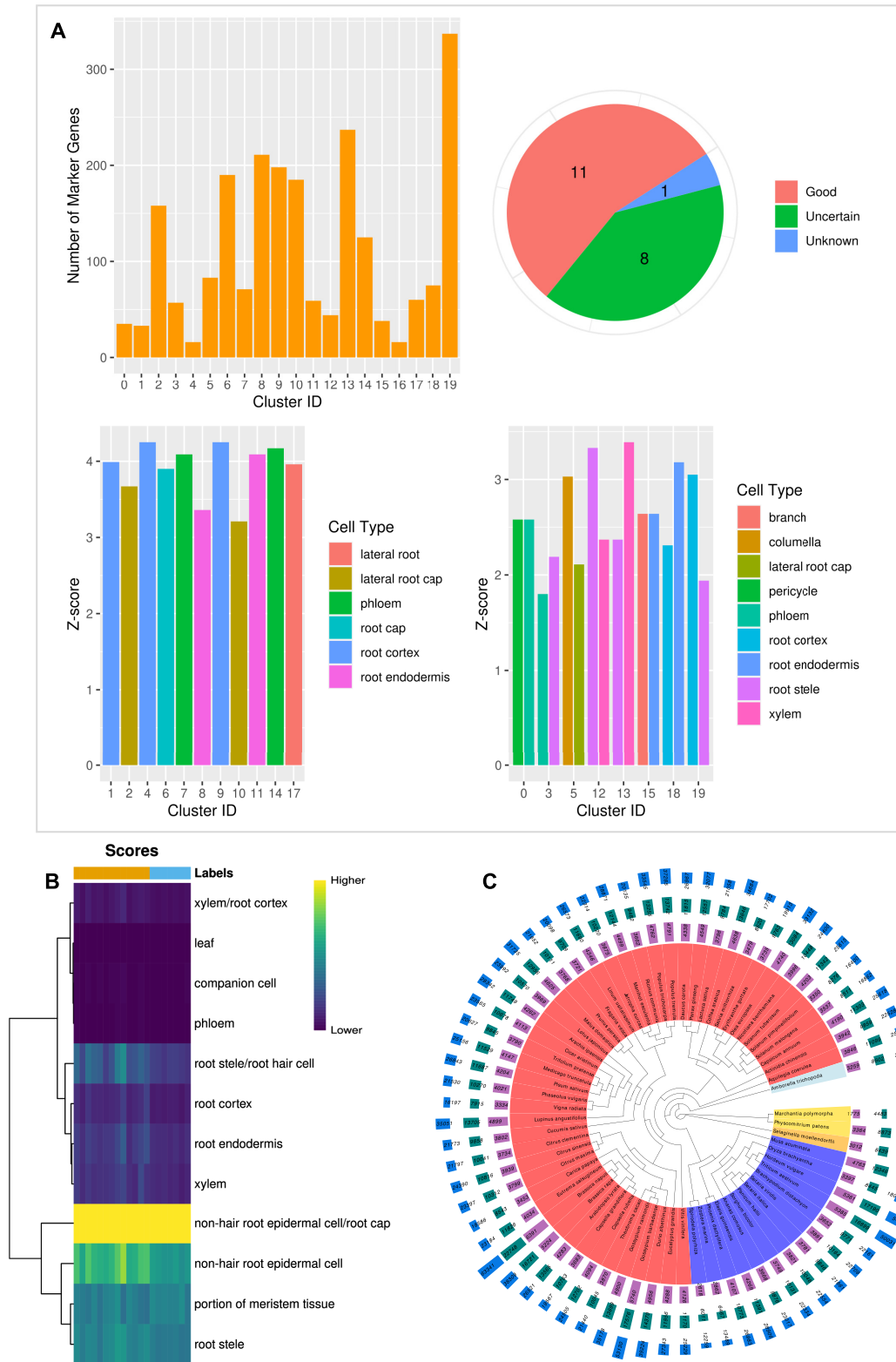
**Browse.** Users can quickly explore cell marker genes of interest, and cell types, using species shortcuts on the home page or through the Browse tag available in the toolbar (Figure 2A). After clicking on different cell types, a summary and complete list of marker genes are displayed. The Browse tag can guide users to access cell markers of diverse cell types in different tissues. The hierarchical cell structures for different plant species are shown on the left tree-based panel.

**Search.** Users can search the whole database using five search engines. Using either keywords or the gene identifier search engine located on the webpage, users can search



**Figure 1.** PCMDB data collection pipeline. Note: the table contains the number of marker genes for each source. The bar plot represents the number of experimental marker genes for the top 20 cell types identified in each plant species.





**Figure 3.** PCMDB tools and number of potential marker genes for currently unsupported species. **(A)** The result of SCSA prediction using the default example data. Left top: The number of marker genes for each cluster of the input data. Right top: The number of clusters classified into different types (Good, Uncertain, and Unknown) by SCSA. Left bottom: Z-score of cluster label for clusters classified into the Uncertain type. Right bottom: Z-score of top 2 cluster labels for clusters classified into the Good type. **(B)** Heatmap based on the score of each cluster label output by SingleR using the default example data. **(C)** The number of potential marker genes for currently unsupported species uncovered using homology search by *Arabidopsis*. From inner to outer, phylogenetic tree of 67 species (different colors mean different clades based on NCBI Taxonomy Browser), the numbers of marker gene candidates identified by experimental markers, bulk RNA-seq markers, and scRNA-seq-related markers.

selecting the species and source types with user-defined downloaded data. We have also uploaded a backup copy of the data available in PCMDB to Zenodo (<https://zenodo.org/>), ID 5101271.

**Submit.** In an effort to make PCMDB a one-stop community resource, we have started accepting submissions of plant marker genes accompanied by the necessary information. All submitted marker genes will be processed using our standard procedures, as described in the Material and Methods. The submitted marker genes will be grouped into the appropriate categories.

### Case study

We tested the usefulness of cell markers collected in PCMDB, taking the single-cell dataset from Zhang *et al.* as example (4). Twenty clusters were obtained by our analysis pipeline mentioned for this dataset. Using the 103 known marker genes mentioned in their paper, 14 clusters had labeled annotation with the SCSA tool, whereas labels for six clusters were still unknown due to unavailable marker genes (Supplementary Figure S3A). However, using all experimental related marker genes of the *Arabidopsis* root (number: 923) in PCMDB, only 1 cluster remained unlabeled (Supplementary Figure S3B). Moreover, using our approach resulted in more clusters (12 clusters) having ‘Good’ type label results, compared to 8 using the 103 genes by Zhang *et al.*, and the quality of cluster labelling was comparable between these two analyses (Supplementary Figure S3B). Similarly, we can also predict the label for clusters using bulk RNA-seq related marker genes (number: 162) (Supplementary Figure S3C), scRNA-seq related marker genes (number: 11 654) (Supplementary Figure S3D), and all marker genes (number: 11 788) (Supplementary Figure S3E) for *Arabidopsis* root. Comparing label annotation results by these three different types of marker genes of PCMDB (Supplementary Figure S3B, C and D), we found that the score for label annotation was highest with experimental marker genes, followed by scRNA-seq candidates.

### DISCUSSION

Cell type annotation is a crucial step in the analysis of single-cell RNA sequencing data, and cell marker genes are useful for the identification and classification of cell types. The rapid increase of single-cell transcriptomic studies demands the establishment of a comprehensive and reliable resource of cell markers that allows users to easily predict the potential functions of specific cells of interest. Currently, the CellMarker database provides comprehensive cell maker resources for both humans and mice, PscB only supports single-cell data visualization from *Arabidopsis* root, and PlantscRNAdb collects publicly available marker genes for four plant species. However, the unstandardized annotation of tissues/cells and the lack of bulk RNA-seq expression and sequence features for marker genes impacts the widespread usage of PlantscRNAdb. Hence, we implemented a standardized pipeline to collect cell marker genes for six plant species and developed several tools to apply the makers to cell type annotation (Figures

1 and 2). The current marker genes in PCMDB may not fit the strict definition of cell marker genes, especially for the bulk RNA-seq-related candidates, which may be involved at the tissue level. As we intend to develop a comprehensive database for plant cell marker genes, bulk RNA-seq candidates were included in PCMDB, which can provide supporting information, as was done in PlantscRNAdb. For example, 25.09% (431/1 718) of the experimentally identified marker genes overlapped with bulk RNA-seq candidates, and 86.9% (1 493/1 718) overlapped with scRNA-seq candidates in *Arabidopsis*. This evidence can be mutually verified, providing more information for users. As in PlantscRNAdb and CellMarker, another limitation of PCMDB is that marker genes may be involved in different cell types, based on different pieces of evidence, and we included all of these records. From the result of case study, we can find that the cell markers in PCMDB can be effectively applied in cell type annotation in a real scRNA-seq dataset, and evidence of different sources can be complementary for each other. In other cases, when few or no experimental marker candidates are available (for example, for the root tissue of tomato), marker genes from scRNA-seq can also help users with cell label annotation. Similarly, with absent or few experimental and scRNA-seq marker candidates (e.g. for leaf tissue of soybean), bulk RNA-seq candidates may also help with cell label annotation. The key point is that users should give different weights or confidence levels for these three different sources, and experimental candidates should have the highest.

PCMDB includes more plant species and a higher number of studies and cell marker genes than PlantscRNAdb (Supplementary Table S1). More than 90% of the marker genes in PlantscRNAdb are contained in PCMDB (Supplementary Table S1). As shown in Supplementary Table S1, the number of cell types is also significantly higher for the three types of cell marker genes. By collecting expression information from bulk RNA-seq and scRNA-seq studies, it is possible to identify specific expression patterns for marker genes (Figure 2). Hence, users can easily predict the potential functions of markers of interest. An innovative feature of PCMDB is the enabling of users to predict cell types on their own single-cell dataset using two analytical tools: SCSA and SingleR (Figure 3). We also provide information about potential marker genes in plant species that are not present in PCMDB, by homology search (Figure 3). Although homology results can provide some clues to users, it's not guaranteed for other plant species, especially when some tissues/cells are absent in some plants.

PCMDB provides a user-friendly interface with which to browse and access all data via multi-layer webpages, powerful search engines, and download ports (Figure 2). We will continue tracking marker gene related studies in literatures and updating the database with novel cell marker genes annually. Plant species that are currently not included in PCMDB will be integrated into the next updates of the database. In the future, we will also integrate different types of cell markers, such as long noncoding RNA and other functional regulatory elements, into PCMDB. Finally, we believe PCMDB constitutes a comprehensive and valuable resource for cell marker genes in plants, and can be used to identify cell types in single-cell studies.

## DATA AVAILABILITY

PCMDB is freely available at <http://www.tobacodb.org/pcmdb/>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Mr. Zeqing Guo, Keqiang Hu and Yongsheng Yan for their IT support. We also want to thank Enago ([www.enago.cn](http://www.enago.cn)) for their language editing of our manuscript.

## FUNDING

Zhengzhou Tobacco Research Institute [CNTC: 110202001020(JY-03), 110201901024(SJ-03)]; Joint Laboratory of HNTI and ZTRI for Tobacco Gene Research and Utilization; Guizhou Academy of Tobacco Science [CNTC: 110202001027(JY-10)]; China Association for Science and Technology [Young Elite Scientists Sponsorship Program 2016QNR001]. Funding for open access charge: China Association for Science and Technology.

*Conflict of interest statement.* None declared.

## REFERENCES

- Cao, Y., Zhu, J., Jia, P. and Zhao, Z. (2017) scRNASeqDB: a database for RNA-Seq based gene expression profiles in human single cells. *Genes*, **8**, 368.
- Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubb, K.L., Trapnell, C., Fields, S., Queitsch, C. and Cuperus, J.T. (2019) Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell*, **31**, 993–1011.
- Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., Li, R., Hu, G., Zhang, P., Ma, Y. *et al.* (2021) Transcriptional landscape of rice roots at the single-cell resolution. *Mol. Plant*, **14**, 384–394.
- Zhang, T.Q., Xu, Z.G., Shang, G.D. and Wang, J.W. (2019) A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant*, **12**, 648–660.
- Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K. and Timmermans, M.C.P. (2019) Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell*, **48**, 840–852.
- Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H. and Wang, J.W. (2021) Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.*, **12**, 2053.
- Wang, Y., Huan, Q., Li, K. and Qian, W. (2021) Single-cell transcriptome atlas of the leaf and root of rice seedlings. *J. Genet. Genomics*, <https://doi.org/10.1016/j.jgg.2021.06.001>.
- Ryu, K.H., Huang, L., Kang, H.M. and Schiefelbein, J. (2019) Single-Cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.*, **179**, 1444–1456.
- Farmer, A., Thibivilliers, S., Ryu, K.H., Schiefelbein, J. and Libault, M. (2021) Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level. *Mol. Plant*, **14**, 372–383.
- Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y. *et al.* (2020) Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Molecular Plant*, **13**, 1178–1193.
- Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O'Malley, R.C., Brady, S.M. *et al.* (2019) High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.*, **27**, 2241–2247.
- Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezruczyk, M., Miras, M., Zollner, N., Hartwig, T., Wudick, M.M. *et al.* (2021) Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell*, **33**, 511–530.
- Satterlee, J.W., Strable, J. and Scanlon, M.J. (2020) Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 33689.
- Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N. *et al.* (2021) Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell*, **56**, 557–568.
- Nelms, B. and Walbot, V. (2019) Defining the developmental program leading to meiosis in maize. *Science*, **364**, 52–56.
- Efroni, I. and Birnbaum, K.D. (2016) The potential of single-cell profiling in plants. *Genome Biol.*, **17**, 65.
- Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinnyen, J.R., Mace, D., Ohler, U. and Benfey, P.N. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*, **318**, 801–806.
- Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M. *et al.* (2019) CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.*, **47**, D721–D728.
- Ma, X., Denyer, T. and Timmermans, M.C.P. (2020) PscB: a browser to explore plant single cell RNA-sequencing data sets. *Plant Physiol.*, **183**, 464–467.
- Chen, H., Yin, X., Guo, L., Yao, J., Ding, Y., Xu, X., Liu, L., Zhu, Q.H., Chu, Q. and Fan, L. (2021) PlantscRNAdb: a database for plant single-cell RNA analysis. *Mol. Plant*, **14**, 855–857.
- Cao, Y., Wang, X. and Peng, G. (2020) SCSA: a cell type annotation tool for single-cell RNA-seq data. *Front. Genet.*, **11**, 490.
- Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R. *et al.* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.*, **20**, 163–172.
- Avraham, S., Tung, C.W., Ilic, K., Jaiswal, P., Kellogg, E.A., McCouch, S., Pujar, A., Reiser, L., Rhee, S.Y., Sachs, M.M. *et al.* (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res.*, **36**, D449–D454.
- Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The *Arabidopsis* information resource: making and mining the “gold standard” annotated reference plant genome. *Genesis*, **53**, 474–485.
- Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McComb, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S. *et al.* (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, **6**, 4.
- Woodhouse, M.R., Cannon, E.K., Portwood, J.L. 2nd, Harper, L.C., Gardiner, J.M., Schaeffer, M.L. and Andorf, C.M. (2021) A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biol.*, **21**, 385.
- Grant, D., Nelson, R.T., Cannon, S.B. and Shoemaker, R.C. (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.*, **38**, D843–D846.
- Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H. *et al.* (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res.*, **43**, D1036–D1041.
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetverin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. *et al.* (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **34**, D173–D180.
- Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, **8**, 14049.
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M. 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
- Jin, J., Lu, P., Xu, Y., Li, Z., Yu, S., Liu, J., Wang, H., Chua, N.H. and Cao, P. (2021) PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Res.*, **49**, D1489–D1495.