scCloudMine: A cloud-based app for visualization, comparison, and exploration of single-cell transcriptomic data

Mathew G. Lewsey¹,², Changyu Yi¹, Oliver Berkowitz¹,², Felipe Ayora³,⁵,*, Maurice Bernado³ and James Whelan¹,²,⁴,*

¹La Trobe Institute for Agriculture and Food, La Trobe University, AgriBio Building, Bundoora, VIC 3086, Australia
²Australian Research Council Research Hub for Medicinal Agriculture, La Trobe University, AgriBio Building, Bundoora, VIC 3086, Australia
³BizData, Level 9/278, Collins Street, Melbourne, VIC 3000, Australia
⁴Department of Animal, Plant and Soil Sciences, School of Life Science, La Trobe University, Bundoora, VIC 3086, Australia
⁵Research and Advanced Computing, BizData, Level 31, 2-6, Gilmer Terrace, Wellington, 6011, New Zealand

*Correspondence: James Whelan (j.whelan@latrobe.edu.au), Felipe Ayora (felipe.ayora@bizdata.co.nz)

https://doi.org/10.1016/j.xplc.2022.100302

ABSTRACT

scCloudMine is a cloud-based application for visualization, comparison, and exploration of single-cell transcriptome data. It does not require an on-site, high-power computing server, installation, or associated expertise and expense. Users upload their own or publicly available scRNA-seq datasets after preprocessing for visualization using a web browser. The data can be viewed in two color modes—Cluster, representing cell identity, and Values, showing levels of expression—and data can be queried using keywords or gene identification number(s). Using the app to compare studies, we determined that some genes frequently used as cell-type markers are in fact study specific. The apparent cell-specific expression of PHO1;H3 differed between GFP-tagging and scRNA-seq studies. Some phosphate transporter genes were induced by protoplasting, but they retained cell specificity, suggesting that cell-specific responses to stress (i.e., protoplasting) can occur. Examination of the cell specificity of hormone response genes revealed that 132 hormone-responsive genes display restricted expression and that the jasmonate response gene TIFY8 is expressed in endodermal cells, in contrast to previous reports. It also appears that JAZ repressors have cell-type-specific functions. These features identified using scCloudMine highlight the need for resources to enable biological researchers to compare their datasets of interest under a variety of parameters. scCloudMine enables researchers to form new hypotheses and perform comparative studies and allows for the easy re-use of data from this emerging technology by a wide variety of users who may not have access or funding for high-performance on-site computing and support.

Key words: single cell, RNA sequencing, visualization, discovery, comparison

Lewsey M.G., Yi C., Berkowitz O., Ayora F., Bernado M., and Whelan J. (2022). scCloudMine: A cloud-based app for visualization, comparison, and exploration of single-cell transcriptomic data. Plant Comm. 3, 100302.

INTRODUCTION

The ever-increasing throughput, declining cost per sample, and growing diversity of gene expression analysis methods have resulted in massive public repositories of freely accessible transcriptome data (Zhang et al., 2020). A variety of tools are available that provide user-friendly data access and interrogation to plant biologists who are not experts in bioinformatics or primary data processing. However, these tools predominantly serve the most common and more established techniques, such as bulk RNA sequencing (RNA-seq) or microarray analysis of whole organs. Single-cell RNA-seq (scRNA-seq) has emerged more recently as an extremely powerful technique to investigate the functions of individual cells and cell types. In recent years scRNA-seq has been applied to several plant species, generating valuable resources for researchers to explore plant genomics at the single-cell level (Denyer et al., 2019; Jean-Baptiste et al., 2020).
Plant Communications

2019; Ryu et al., 2018; Liu et al., 2021; Xu et al., 2021; Zhang et al., 2021). However, these data are typically stored in a text-based format that must be processed and visualized before users can explore it. This creates a significant challenge for plant scientists with limited computational experience and therefore limits data re-use.

Beyond visualization, researchers need to be able to compare different datasets of interest, design new experiments, compare studies, and define cell-specific genes using the parameters of their choice. Publicly available RNA-seq and scRNA-seq repositories provide a valuable resource for re-use of raw data, making them more accessible and enabling their thorough investigation by subject matter experts who have a limited bioinformatics background. For RNA-seq data, this has been achieved by a variety of databases that have emerged in the last 20 years. The Bio-Analytic Resource for Plant Biology (BAR) and Genevestigator are notable examples that provide repositories and analytical tools for the comparison of RNA-seq data (Zimmermann et al., 2004; Toufighi et al., 2005).

However, these visualization tools are not compatible with the highly dimensional and large datasets created by scRNA-seq, in which the expression of thousands of genes is measured in thousands of cells. This issue will only become more challenging as costs decrease over time. Furthermore, most plant biologists are not bioinformatics experts and are unlikely to rapidly learn how to analyze scRNA-seq data with the currently available command line tools. This creates a significant obstacle to accessing and re-using public data. Moreover, substantial computing hardware, which costs tens to hundreds of thousands of dollars to acquire, is required for the storage and processing of scRNA-seq data and needs ongoing infrastructure maintenance and specialized staff for continued use. When re-analysis is likely to be a short-term project and not an ongoing requirement, it is difficult to justify such a substantial capital investment.

Tools have been developed to overcome these obstacles, such as SC1 (Moussa and Mandoiu, 2021), Single Cell Explorer (Feng et al., 2019), and alona (Franzen and Bjorkegren, 2020). However, each of these tools has some disadvantages. For example, SC1 is not currently able to process plant scRNA-seq data. Single Cell Explorer must be installed on a Linux server by the users. Moreover, these tools focus primarily on scRNA-seq data processing while limited data visualization functions. Genevestigator has established a single-cell RNA-seq portal, but it is restricted to animal/human studies.

One of the central questions in single-cell analyses is the identification of marker genes used to determine cell types (Shaw et al., 2021). Stress responses diminish cell-specific signatures, thereby impeding cell-type assignment, as shown for Arabidopsis roots (Jean-Baptiste et al., 2019). Similarly, in a study that examined the tissue specificity of gene expression in Arabidopsis leaves, tissue-specific expression decreased with applications of chemicals that mimicked adverse growth conditions (e.g., oxidative stress) or hormones (Berkowitz et al., 2021). It is therefore difficult for non-expert researchers to test the robustness of marker genes defined by earlier studies and to evaluate how consistent they may be under different conditions. In addition, the increasing application of genomics to non-model systems means that the capability to define marker genes will be needed for a variety of plants species, as will tools that allow researchers to compare markers between studies and species. PlantscRNAdb has compiled scRNA-seq datasets from plants and defined 26,326 marker genes for 128 different cell types from four plant species (Shaw et al., 2021). This apparently high number shows the diversity and difference in approaches to defining marker genes. Researchers must be able to adjust the parameters used to define marker genes in order to include new information and allow for discovery. Parameters set too strictly or too leniently may cause results to be missed or meaningless.

To address this challenge, we developed scCloudMine. This tool focuses on the interactive visualization of user-processed or publicly available scRNA-seq data. All data storage and processing are conducted using a Microsoft Azure cloud-based platform, eliminating the necessity for users to purchase costly computer hardware with associated maintenance costs. The application (app) provides a web-based, user-friendly graphical interface, making it appropriate for plant scientists who are not bioinformatics experts. The provision of these tools within an app allows them to be rapidly and easily deployed and updated within a user’s institutional environment, greatly reducing challenges around installation and software maintenance. Overall, this cloud-based approach can scale readily, depending on the needs of individuals or institutions. scCloudMine can be obtained for local installation at http://Single-Cell-Visualisation.loomesoftware.com.

RESULTS

Overview of the single-cell app

scCloudMine is designed to provide an easily accessible environment in which users new to scRNA-seq can explore data. It enables users to visualize cell expression profiles. The app is deployed within the user’s host institution information technology (IT) infrastructure and, once established, enables users to connect over the Internet using a web browser via a custom web address (URL) that can be associated with any institutional domain. The architecture of the app uses the Microsoft Azure public cloud (Figure 1). It is based on Platform as a Service (PaaS) and Software as a Service (SaaS) solutions. This provides automatic scalability when additional storage or computer resources are required to perform an analysis or visualization; the cloud back end will automatically make those resources available to the app, within specified parameters. This is achieved using the Loom platform (https://www.loomesoftware.com/index) and Microsoft Active Directory. As a cloud-based app, it does not require any on-site computer server or storage with associated maintenance and infrastructure costs.

There are three main screens in the scCloudMine interface when the app has been installed: Upload Experiment, Experiment Analysis, and Experiment Status (Figure 2, Supplemental Video 1). Upload Experiment provides the functionality to upload gene expression data, clustering and gene ontology (GO) information, and metadata associated with the data so that they can be identified as part of an experiment. Experiment Analysis enables visualization of data for experiments that have been uploaded (Figure 2) and has rich controls for searching, filtering, and modifying the visualized data. The user can also
export the visualized data in GO slim or comma-separated values (CSV) format.

Utility of scCloudMine

When the app has been installed on their server, users can upload experiments using the Upload Experiment function. Three data tables in CSV format are required to upload an experiment: (1) a gene expression matrix, (2) a meta information table describing each cell, and (3) a GO slim file for the studied species. In the gene expression matrix, the first column is the gene ID, and the other columns are the unique cell barcodes; the values represent the expression of each gene in each cell. The meta information table provides the three-dimensional coordinates from clustering analysis (e.g., t-distributed stochastic neighbor embedding [T-SNE] and Uniform and Manifold Approximation and Projection [UMAP]), cell identity, and other experimental information for each cell. The GO slim file contains the annotation of each gene. Examples of these data tables have been provided in Supplemental Tables 1–3. Notably, the system is species agnostic as long as all three files relate to the same species and use consistent nomenclature.

The main strength of our app is its ability to provide scRNA-seq data visualization with various interactive functions. Users can select different uploaded experiments from the drop-down list and examine the related meta information. The app provides a search function based on gene ID and/or GO annotation, which allows users to explore and discover genes of interest (Figure 2C). For instance, one can search genes associated with phosphate-related biological processes by providing “phosphate” in the “Search GO_SLIM” box and then selecting an associated gene in the “Search GenelD” box. One can also search for a gene of interest directly from the search box. Cells can be filtered based on the expression level of a searched gene, with cells above or below the user-selected threshold masked. Moreover, the app provides two color modes (Cluster and Values) for displaying the cells. In Cluster mode, the color represents cell identity (Cluster number), whereas in Values mode, color indicates the expression level of cells (Figure 2B). The app also provides a three-dimensional visualization of the cell clustering (Figure 2D) and selection of cells based on datasets (Figure 2E) or cluster identity (Figure 2F).

Characterizing variation in four publicly available Arabidopsis root scRNA-seq datasets

Roots provide an ideal tissue for analyzing cell heterogeneity because high-quality ground-truth datasets exist in which all root cell types have been carefully identified, purified, and transcriptomically analyzed (Drapek et al., 2017; Shahan et al., 2021). Moreover, root development is very well characterized and roots are easily accessible and amenable to single-cell processing methods (Shahan et al., 2021). Several scRNA-seq studies have been performed in Arabidopsis roots using the Columbia-0 (Col-0) genotype and three mutants in the Col-0 background (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019). Heterogeneity of cellular responses to environmental stimuli, such as high temperature and sucrose supply, have also been profiled in two studies (Jean-Baptiste et al., 2019; Shulse et al., 2019). These studies have greatly advanced our understanding of the spatiotemporal trajectories of plant root development. To maximize the use of these publicly available datasets and cross-validate results from different studies, we integrated four public Arabidopsis root scRNA-seq datasets and visualized the results in our single-cell app (Supplemental Table 4; see data processing methods in methods; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019). By integrated analysis and visualization, we were able to compare the results of these studies and understand more about experimental variation when working with a single genotype.

We examined variations among different root single-cell experiments performed on the same genotype under similar conditions. Cell-type (i.e., cluster)-specific marker genes identified in one of
the root experiments were used to assign cell identities to the cell clusters identified in our integrated analysis of the four experiments (Figure 3; Supplemental Table 5; Denyer et al., 2019). Our analysis revealed that some marker genes were specific across studies—for example, pericycle/phloem genes (CLE26, NTL, NAT7, HCA2, and FAF4). However, examination of the same genes in individual studies indicated that there was still variation in the extent of expression (Figure 4). FAF4 was defined as a good marker for the pericycle/phloem based on the data taken from Jean-Baptiste et al. (2019) but not based on the data from Shulse et al. (2019). At1g14190 and UGD1 were markers for protoxylem based on all the studies, and GH9C1 and At1g07795 appeared quite robust for trichoblasts, although notably GH9C1 was also found consistently in the meristematic xylem and was undefined in the study of Shulse et al. (2019). When all datasets were integrated and analyzed together, we observed that previously defined atrichoblast markers were in fact also found in the columella, quiescent center (QC), cortex, meristematic xylem, and trichoblasts. Likewise, the cortex marker TBL41 was found in the meristematic xylem, atrichoblasts, and trichoblasts. Although atrichoblast marker genes were highly expressed in atrichoblast clusters (0 and 12), they were also expressed in other cell types, such as columella cells and trichoblasts, consistent with a previous report (Denyer et al., 2019).

We also discovered three clusters (1, 9, and 11) in our integrated analysis of all datasets that could not be assigned to a cell type because none of the known marker genes was expressed specifically in these clusters. This analysis revealed that although the concept of marker genes is useful for the biological interpretation of individual studies, markers may vary across different experiments. This does not invalidate individual studies; rather, it suggests that subtle differences in growth conditions cause variations that become apparent when a larger set of studies is analyzed. The ability to compare studies and look at them together using our app is valuable for designing future experiments and forming hypotheses. The ability to identify cell-specific promoters benefits from combining many studies, but understanding variation between individual studies is also beneficial before proceeding with experimental studies.

Mining expression profiles of phosphate transporters using scCloudMine

Phosphate (Pi) is an important macronutrient for plant growth and development. Pi transporters (PHTs) play a critical role in Pi uptake from the environment and Pi translocation between organs, cell types, and organelles (Hamburger et al., 2002; Muchhal et al., 1996; Mudge et al., 2002; Shinn et al., 2004). Five PHT families (PHT1–5) with different subcellular localizations have been characterized in plants based on phylogenetic analysis (Irigoyen et al., 2011; Mudge et al., 2002; Versaw and Harrison, 2002; Wang et al., 2017; Zhu et al., 2012). PHOSPHATE1 (PHO1) is a Pi exporter that is responsible for loading Pi into the xylem vessels (Hamburger et al., 2002), and 10 PHO1 homologs (PHO1;H1–10) have been identified in Arabidopsis (Wang et al., 2004). Although many studies have been performed to decipher the functions of the PHTs, a comprehensive spatial profile of their expression has not yet been constructed (Hamburger et al., 2002; Khan et al., 2014; Mudge et al., 2002; Shinn et al., 2004; Stefanovic et al., 2007). We therefore examined the expression of all known genes (32 genes) from the PHT and PHO1 families in the integrated Arabidopsis root dataset (Figure 5; Supplemental Table 6). Out of the 32 genes, only 20 genes passed the quality control checks (see data sources and processing) (Figure 5). Consistent with a previous report (Mudge et al., 2002), PHT1;1 is strongly expressed in trichoblasts, where Pi is taken up from the soil (Figures 5A and 5C; Supplemental Table 6). We also found that PHT1;1 was highly expressed in the cortex, confirming an additional role for PHT1;1 in the transport of Pi from the epidermis into the central cylinder of the root (Karthikeyan et al., 2002) (Figures 5A and 5C; Supplemental Table 6). PHO1 was the first gene in this family to be characterized and was expressed mainly in stelar cells, including pericycle and xylem parenchyma cells, consistent with our results (Figure 5C) (Hamburger et al., 2002). To date, only PHO1;H1 has been shown to complement the PHO1 loss-of-function mutant, and the functions of the other PHO1 homologs remain unknown (Stefanovic et al., 2007). Seven PHO1 family genes (PHO1, PHO1;H1, PHO1;H2, PHO1;H4, PHO1;H5, PHO1;H7, and PHO1;H10) are induced by protoplasting (Denyer et al., 2019), but only PHO1, PHO1;H1,
PHO1;H3, and PHO1;H10 passed the quality control checks in our study. Interestingly, in contrast to a previous study that reported expression of PHO1;H3 in the root vascular cylinder (Khan et al., 2014), we found that PHO1;H3 was expressed mainly in the endodermis (Figures 5B and 5C). Endodermal cells have thick cell walls that restrict water and ion transport through this layer to the symplastic pathway, and PHO1;H3 may therefore also facilitate Pi transport from the cortex into the stelar cells through the endodermis. These significant differences between studies with respect to cell-type-specific PHT expression highlight the potential for comparative approaches enabled by our app to prompt further hypothesis-driven experiments.

Characterizing single-cell expression patterns of plant hormone genes in the Arabidopsis root

Plant hormones regulate plant growth, development, and stress responses (Santner et al., 2009; Kumar, 2013). Essentially, all plant hormones are involved in root development in some manner (Fu and Harberd, 2003; Kumar, 2013; Qin et al., 2019) (Staswick et al., 1992; Ruzicka et al., 2007; Tian et al., 2009; Zhao et al., 2014; McAdam et al., 2016; Yang et al., 2017). Hormonal signaling operates in a cell- and tissue-specific manner, which is critical for enabling the unique functions and responses of individual cell types (Novak et al., 2017). However, there has been little transcriptomic analysis of the expression patterns of plant hormone signaling genes at the level of individual cells. The advent of single-cell gene expression technologies allows us to investigate this subject.

To characterize the single-cell expression profiles of plant hormone-related genes, we examined the expression of 685 marker genes for 7 hormones from previous studies: auxin (63 genes), abscisic acid (311), brassinolide (6), cytokinins (14), ethylene (3), gibberellic acid (9), and jasmonic acid (JA) (279) (Birnbaum et al., 2003; Nemhauser et al., 2006; Zander et al., 2020). We focused our analysis on core components of the JA transcriptional regulatory mechanism because the promoters of the MYC2, -3, and -4 transcription factors drive cell-specific reporter expression in roots (Gasperini et al., 2015). Although we observed some variation between cell types in the scRNA-seq datasets, these transcription factors were expressed relatively broadly, suggesting that differences in experimental conditions or the approaches employed may contribute to changed expression domains. We reasoned that other JA components may behave similarly, and we examined the expression of MYC2, -3, and -4, the JAZ repressors, and other related factors (Figure 6B). NINJA was expressed broadly across root cell types, consistent with the previously reported behavior of its promoter (Gasperini et al., 2015). Only one component, TIFY8, exceeded the tau threshold of 0.85, indicating cluster-specific expression in endodermis cluster 14. TIFY8 interacts with proteins that regulate root meristem initiation, and this expression pattern perhaps reflects the participation of the endodermis in lateral root formation (Cuellar Perez et al., 2014; Torres-Martinez et al., 2019). There was, however, clear variation across clusters in both expression level and proportion of cells expressing other JA transcriptional regulators for those components that did not exceed the tau threshold. For example, expression of JAZ repressors was generally lower in the columella, QC, and meristematic xylem clusters. Expression of JAZ5 was highest in the endodermis, pericycle/phloem, and protoxylem clusters, whereas JAZ6 expression was highest in cortex cluster 16. These results probably indicate that JAZ repressors have cell-type-specific functions in gene expression regulation in the root, potentially explaining why the JAZ/TIFY family is relatively large.

DISCUSSION

scRNA-seq has been used by plant scientists since 2013, and related publications have dramatically increased recently because of the advent of droplet-based scRNA-seq technologies.

We focused our analysis on core components of the JA transcriptional regulatory mechanism because the promoters of the MYC2, -3, and -4 transcription factors drive cell-specific reporter expression in roots (Gasperini et al., 2015). Although we observed some variation between cell types in the scRNA-seq datasets, these transcription factors were expressed relatively broadly, suggesting that differences in experimental conditions or the approaches employed may contribute to changed expression domains. We reasoned that other JA components may behave similarly, and we examined the expression of MYC2, -3, and -4, the JAZ repressors, and other related factors (Figure 6B). NINJA was expressed broadly across root cell types, consistent with the previously reported behavior of its promoter (Gasperini et al., 2015). Only one component, TIFY8, exceeded the tau threshold of 0.85, indicating cluster-specific expression in endodermis cluster 14. TIFY8 interacts with proteins that regulate root meristem initiation, and this expression pattern perhaps reflects the participation of the endodermis in lateral root formation (Cuellar Perez et al., 2014; Torres-Martinez et al., 2019). There was, however, clear variation across clusters in both expression level and proportion of cells expressing other JA transcriptional regulators for those components that did not exceed the tau threshold. For example, expression of JAZ repressors was generally lower in the columella, QC, and meristematic xylem clusters. Expression of JAZ5 was highest in the endodermis, pericycle/phloem, and protoxylem clusters, whereas JAZ6 expression was highest in cortex cluster 16. These results probably indicate that JAZ repressors have cell-type-specific functions in gene expression regulation in the root, potentially explaining why the JAZ/TIFY family is relatively large.

DISCUSSION

scRNA-seq has been used by plant scientists since 2013, and related publications have dramatically increased recently because of the advent of droplet-based scRNA-seq technologies.
Although the increasing number of scRNA-seq studies has advanced our understanding of plant cell heterogeneity, mining public data is challenging for researchers with limited computational expertise. In this study, we developed scCloudMine, a web-based platform for scRNA-seq data visualization. Unlike other online scRNA-seq tools that focus on data processing, our app provides a user-friendly data visualization interface (Feng et al., 2019; Franzen and Bjorkegren, 2020; Madler et al., 2020). In addition to a common gene-based search function, scCloudMine also provides a function-based search option that is useful for examining the expression of genes related to a specific biological process. We also presented an integrated Arabidopsis root dataset from four publications (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019), showing the utility of scCloudMine for defining hypotheses for experimental testing. Together with scCloudMine, publicly available datasets provide a powerful resource for researchers who wish to examine gene expression at a single-cell resolution.

The current version of scCloudMine is designed to enable visualization of processed data that have already undergone mapping and gene quantification (see data sources and processing in methods). Future developments are envisaged that integrate prior processing steps into scCloudMine so that both the secondary analysis (sequencing data processing) and tertiary analysis (visualization) become automated. In this future scenario, the input into scCloudMine would become the primary analysis from a sequencer (FASTQ files prior to mapping), creating an end-to-end solution whereby single-cell gene expression analysis would not require specialized high-performance computing knowledge. This would have the advantage that experimental data could be analyzed immediately by direct end users who have the biological knowledge to interpret the findings.

In our study, we annotated cell clusters from an integrated, multi-experiment root scRNA-seq dataset using well-characterized cell marker genes. We identified patterns of cell identity similar to those reported in previous studies (Figure 3) (Ryu et al., 2019; Shulse et al., 2019; Jean-Baptiste et al., 2019; Denyer et al., 2019). Unexpectedly, we found that the specificity of some marker genes varied between studies, with some genes expressed in different cell types compared with previous publications. PHO1;H3 was reported to be specifically expressed in the stele, but it was expressed mainly in the endodermis in this study (Figures 5B and 4C) (Khan et al., 2014). These differences may result from differences in environmental stimuli or growth stages between studies. For example, stele-specific expression of PHO1;H3 was detected in plants grown in a zinc-deficient environment, whereas data analyzed here were obtained from seedlings supplied with sufficient nutrients (Khan et al., 2014). Indeed, cell heterogeneity in response to environmental stimuli has been reported previously (Berkowitz et al., 2021; Jean-Baptiste et al., 2019; Shulse et al., 2019). We also observed that a variety of genes encoding PHTs were induced by protoplasting and would normally be discarded from analysis. However, these genes were still expressed in a cell-specific manner (Figure 5C). Likewise, the analysis of hormone-response genes demonstrated that many genes are expressed in a cell-enriched manner, indicating that data from whole-organ studies that are used to interpret signaling and functional pathways may need to be re-examined, as all the genes are not expressed in the same cell. Thus, these pathways may not exist
Figure 5. Visualization of the expression of phosphate transporter genes in different cell types.

(A) and (B) Single-cell app screenshot of the expression of PHT1;1 (A) and PHO1;H3 (B) across the four studies. Cells with fewer than five and three reads were filtered out for PHT1;1 and PHO1;H3, respectively. PHT1;1 is expressed mainly in trichoblasts, atrichoblasts, and the cortex; PHO1;H3 is expressed mainly in the endodermis.

(C) Expression of expressed phosphate transporter genes across the four indicated studies. Clusters of cells in which PHT1;1 and PHO1;H3 were highly expressed are represented by light blue boxes. Expression percentage of PHT1;1 in cluster 12 was lower in Jean-Baptiste et al. (2019) than in the other three studies. Protoplasting-induced genes are highlighted in red text. Although induced by protoplast isolation, PHT1;4 and PHO1 are still expressed in a cell-specific manner (red boxes), with PHT1;4 expressed in the cortex and atrichoblasts and PHO1 expressed in the pericycle/phloem.
in any one cell, suggesting heterogeneity in cell transcriptomes. Quantitative imaging of single-cell transcriptional dynamics in plant cells has demonstrated large differences between neighboring cells, consistent with the conclusions reached here through the comparison of cell-specific transcriptomes (Alamos et al., 2021; Hani et al., 2021).

Emerging technologies are producing ever-increasing amounts of data. The next grand challenge in plant science is the construction of a Plant Cell Atlas, envisioned as mapping all cells and tissues of a plant across multiple scales and data modalities (Plant Cell Atlas Consortium, 2021). Although large centralized databases are invaluable for housing datasets for re-use by researchers, individuals must be able to select specific datasets relevant to their research focus and to re-analyze and visualize them to form new hypotheses and assess their own results. This creates a continually growing demand for researchers to become experts in system administration in order to download and install various packages and for institutions to have the computer infrastructure and staff expertise to support such equipment. This traditional approach comes at a considerable cost to institutions and individuals that is paid irrespective of use. The on-site approach also requires users to be trained in informatic analysis to use such packages. However, analysis of scRNA-seq datasets is often intermittent and can be as infrequent as once or twice a year for an individual researcher. Once analyzed, scRNA-seq data generate hypotheses for experimental testing. Therefore, the activation cost for the average individual laboratories to access and analyze such data can be substantial and time consuming, even though the data are readily available for re-use. Thus, scCloudMine, as developed and deployed here, is designed to bypass the large cost necessary to obtain and maintain such facilities. The widespread availability of the Azure cloud platform means that it can be used by individuals and institutions without the capital or facilities to invest in significant computational and storage capacity. The app can also serve as a model for genomics software development on other cloud platforms. Finally, it can be used by the community for educational or other purposes, enabling pre-tertiary students to carry out project-based research.

METHODS

Single-cell app architecture

The architecture of the application uses the Microsoft Azure public cloud (Figure 1; Table 1) and allows easy deployment via the Loome platform. Users connect to the single-cell app over the Internet using a
repositories (Supplemental Table 4). Note that the TAIR10 genome was used by the authors to generate raw count data across the four publicly available datasets (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019). It is first necessary to process the raw data, and best practices for doing so have been reviewed well elsewhere (Shaw et al., 2021). Our approach was to use Seurat (version 3.2.2) for downstream data processing, including quality control, data normalization, identification of highly variable features, dimensional reduction, and cell clustering (Stuart et al., 2019). In brief, to discard low-quality cells, we filtered out cells that expressed fewer than 800 genes or 3,000 unique molecule identifiers. Cells with more than 20% mitochondrial sequences were also removed to exclude dead cells, and genes expressed in fewer than 30 cells were also eliminated. After filtering, 21,557 expressed genes across 16,213 cells were retained. The raw count data were normalized using sctransform, and different datasets were integrated using Seurat (version 3.2.2) (Hafemeister and Satija, 2019; Stuart et al., 2019). Dimensionality reduction was performed on the normalized data using principle component analysis followed by UMAP to visualize the data structure. Cell clusters were characterized using the function FindClusters from the Seurat package with resolution = 0.6, resulting in 21 cell clusters (Figure 3).

Calculation of cell/cluster specificity index (tau)

The metric tau was used as an index of cluster-specific expression and was computed as previously described (Yanai et al., 2005). The average expression of an individual gene was calculated across all cells of the same cell type. The tau value was then computed using these average expression levels from the different cell types.

SUPPLEMENTAL INFORMATION

Supplemental information is available at Plant Communications Online.

FUNDING

This work was supported by grants from the Australian Research Council Discovery to J.W. (DP210103258) and the Australian Research Council Industrial Transformation Research Hub in Medicinal Agriculture (IT180100006) to J.W. and M.G.L.

AUTHOR CONTRIBUTIONS

M.G.L. and J.W. conceived the project, and C.Y. and O.B. carried out the analysis of the biological datasets. F.A. and M.B. designed and implemented the cloud-based architecture for uploading, processing, and visualizing the data. J.W. and F.A. drafted the manuscript with extensive editing by all authors.

ACKNOWLEDGMENTS

F.A. and M.B. are employed by BizData, which hosts the application on the Loome platform. No conflict of interest is declared.

Table 1. The specifications for components are shown.

| Component                  | Type               | Specifications                                      |
|----------------------------|--------------------|----------------------------------------------------|
| App services               | App service plan   | Auto-scaling from 1 to 3 instances based on CPU utilization. S1 pricing tier |
| Visualization web app      | App service        | Managed by app services. Includes custom domain for the single-cell app |
| File uploader web app      | App service        | Managed by app services                              |
| SQL server                 | SQL server         | Logical SQL server. Automatic tuning and transparent data encryption |
| Application database       | SQL database       | Serverless, Gen5, 16 vCores                         |
| Visualization database     | SQL database       | General purpose. S1 pricing tier                    |
| Loome Integrate agent      | Azure Container Instance | Linux, 1 instance, 2 vCores, 8 GB memory           |
| Blob Storage               | Storage account    | General purpose v1                                  |

REFERENCES

Alamos, S., Reimer, A., Ninogi, K.K., and Garcia, H.G. (2021). Quantitative imaging of RNA polymerase II activity in plants reveals the single-cell basis of tissue-wide transcriptional dynamics. Nat. Plants 7:1037–1049. https://doi.org/10.1038/s41477-021-00976-0.

Berkowitz, O., Xu, Y., Liew, L.C., Wang, Y., Zhu, Y., Hurgobin, B., Lewsey, M.G., and Whelan, J. (2021). RNA-seq analysis of laser microdissected Arabidopsis thaliana leaf epidermis, mesophyll and vasculature defines tissue-specific transcriptional responses to...
Plant Communications

Multiple stress treatments. Plant J. 107:938–955. https://doi.org/10.1111/tpj.15314.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science 302:1956–1960. https://doi.org/10.1126/science.1090022.

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. https://doi.org/10.1016/j.molp.2021.05.002.

Cuelar Perez, A., Nagels Durand, A., Vanden Bossche, R., De Clercq, G., Goossens, A., et al. (2014). The non-JAZ TIFY protein TIFY8 from Arabidopsis thaliana is a transcriptional repressor. PLoS ONE 9:e84891. https://doi.org/10.1371/journal.pone.0084891.

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 mRNA sequencing. Dev. Cell 48:840–852.e845. https://doi.org/10.1016/j.devcel.2019.02.022.

Drapek, C., Sparks, E.E., and Benfey, P.N. (2017). Uncovering gene regulatory networks controlling plant cell differentiation. Trends Genet. 33:529–539. https://doi.org/10.1016/j.tig.2017.05.002.

Feng, D., Whitehurst, C.E., Shan, D., Hill, J.D., and Yue, Y.G. (2019). Single Cell Explorer, collaboration-driven tools to leverage large-scale single cell RNA-seq data. BMC Genomics 20:676. https://doi.org/10.1186/s12864-019-6053-y.

Franzen, O., and Bjorkgren, J.L.M. (2020). alona: a web server for single-cell RNA-seq analysis. Bioinformatics 36:3910–3912. https://doi.org/10.1093/bioinformatics/btaa269.

Fu, X., and Harberd, N.P. (2003). Auxin promotes Arabidopsis root growth by modulating gibberellin response. Nature 421:740–743. https://doi.org/10.1038/nature01387.

Gasperini, D., Chetelat, A., Acosta, I.F., Goossens, J., Pauwels, L., Goossens, A., Dreo, S., Alfonso, E., and Farmer, E.E. (2015). Multilayered organization of jasmonate signalling in the regulation of root growth. Plos Genet. 11:e1005300. https://doi.org/10.1371/journal.pgen.1005300.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20:296. https://doi.org/10.1186/s13059-019-1874-1.

Hamburger, D., Rezzonico, E., MacDonald-Comber Petetot, J., Somerville, C., and Poirier, Y. (2002). Identification and characterization of the Arabidopsis PHO1 gene involved in phosphate loading to the xylem. Plant Cell 14:889–902. https://doi.org/10.1105/tpc.000745.

Hani, S., Cuyas, L., David, P., Secco, D., Whelan, J., Thibaud, M.C., Merret, R., Mueller, F., Pochon, N., Javot, H., et al. (2021). Live single-cell RNA sequencing resolves molecular relationships among individual plant cells. Nat. Plants 7:1050–1064. https://doi.org/10.1038/s41477-021-00981-3.

Irigoyen, S., Karlsson, P.M., Kuruvilla, J., Spetea, C., and Versaw, W.K. (2011). The sink-specific plastidic phosphate transporter PHT4.2 influences starch accumulation and leaf size in Arabidopsis. Plant Physiol. 157:1765–1777. https://doi.org/10.1104/pp.111.181925.

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. https://doi.org/10.1105/tpc.18.00785.

Karthikeyan, A.S., Varadarajan, D.K., Mukatira, U.T., D’Urzo, M.P., Damsz, B., and Raghothama, K.G. (2002). Regulated expression of Arabidopsis phosphate transporters. Plant Physiol. 130:221–233. https://doi.org/10.1104/pp.020007.

Khan, G.A., Bouraine, S., Wege, S., Li, Y., de Carbonnel, M., Berthomieu, P., Poirier, Y., and Rouached, H. (2014). Coordination between zinc and phosphate homeostasis involves the transcription factor PHR1, the phosphate exporter PHO1, and its homologue PHO1;H3 in Arabidopsis. J. Exp. Bot. 65:871–884. https://doi.org/10.1093/jxb/eru444.

Kumar, P.P. (2013). Regulation of biotic and abiotic stress responses by plant hormones. Plant Cell Rep 32:943. https://doi.org/10.1007/s00299-013-1460-z.

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. https://doi.org/10.1016/j.molp.2020.12.014.

Madler, S.C., Julien-Laferriere, A., Wyss, L., Phan, M., Kang, A.S.W., Ulrich, E., Schmucki, R., Zhang, J.D., Ebeling, M., Badi, L., et al. (2020). Besca, a single-cell transcriptomics analysis toolkit to accelerate translational research. BioRxiv https://doi.org/10.1101.2020.08.11.245795.

McAdam, S.A., Brodribb, T.J., and Ross, J.J. (2016). Shoot-derived abscisic acid promotes root growth. Plant Cell Environ 39:652–659. https://doi.org/10.1111/pce.12668.

Moussa, M., and Mandoiu, I.I. (2021). SC1: a tool for interactive web-based single-cell RNA-Seq data analysis. J. Comput. Biol. 28:820–841. https://doi.org/10.1089/cmb.2021.0051.

Muchhal, U.S., Pardo, J.M., and Raghothama, K.G. (1996). Phosphate transporters from the higher plant Arabidopsis thaliana. Proc. Natl. Acad. Sci. U S A 93:10519–10523. https://doi.org/10.1073/pnas.93.19.10519.

Mudge, S.R., Rae, A.L., Diatloff, E., and Smith, F.W. (2002). Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in Arabidopsis. Plant J. 31:341–353. https://doi.org/10.1046/j.1365-313x.2002.01356.x.

Nemhauser, J.L., Hong, F., and Chory, J. (2006). Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. Cell 126:467–475. https://doi.org/10.1016/j.cell.2006.05.050.

Novak, O., Napier, R., and Ljung, K. (2017). Zooming in on plant hormone analysis: tissue- and cell-specific approaches. Annu. Rev. Plant Biol. 68:323–348. https://doi.org/10.1146/annurev-arplant-042916-040812.

Plant Cell Atlas Consortium. (2021). Science forum: vision, challenges and opportunities for a plant cell atlas. eLife 10:e66877. https://doi.org/10.7554/eLife.66877.

Qiu, H., He, L., and Huang, R. (2019). The coordination of ethylene and other hormones in primary root development. Front Plant Sci. 10:874. https://doi.org/10.3389/fpls.2019.00874.

Ruzicka, K., Ljung, K., Vanneste, S., Podhorska, R., Beeckman, T., Friml, J., and Benkova, E. (2007). Ethylene regulates root growth through effects on auxin biosynthesis and transport-dependent auxin distribution. Plant Cell 19:2197–2212. https://doi.org/10.1105/tpc.107.052126.

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. https://doi.org/10.1104/pp.18.01482.

Santer, A., Calderon-Villalobos, L.I., and Estelle, M. (2009). Plant hormones are versatile chemical regulators of plant growth. Nat. Chem. Biol. 5:301–307. https://doi.org/10.1038/nchembio.165.
Shah, R., Nolan, T.M., and Benfey, P.N. (2021). Single-cell analysis of cell identity in the Arabidopsis root apical meristem: insights and opportunities. J. Exp. Bot. erab228. https://doi.org/10.1093/jxb/erab228.

Shaw, R., Tian, X., and Xu, J. (2021). Single-cell transcriptome analysis in plants: advances and challenges. Mol. Plant 14:115–126. https://doi.org/10.1016/j.molp.2020.10.012.

Shin, H., Shin, H.S., Dewbre, G.R., and Harrison, M.J. (2004). Phosphate transport in Arabidopsis: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-phosphate environments. Plant J. 39:629–642. https://doi.org/10.1111/j.1365-313X.2004.02161.x.

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.e2244. https://doi.org/10.1016/j.celrep.2019.04.054.

Staswick, P.E., Su, W., and Howell, S.H. (1992). Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an Arabidopsis thaliana mutant. Proc. Natl. Acad. Sci. U S A 89:6837–6840. https://doi.org/10.1073/pnas.89.15.6837.

Stefanovic, A., Ribot, C., Rouached, H., Wang, Y., Chong, J., Belbahri, L., Delessert, S., and Poirier, Y. (2007). Members of the PHO1 gene family show limited functional redundancy in phosphate transfer to the shoot, and are regulated by phosphate deficiency via distinct pathways. Plant J. 50:982–994. https://doi.org/10.1111/j.1365-313X.2007.03108.x.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stockeius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell 177:1888–1902.e1821. https://doi.org/10.1016/j.cell.2019.05.031.

Tian, Q.Y., Sun, P., and Zhang, W.H. (2009). Ethylene is involved in nitrate-dependent root growth and branching in Arabidopsis thaliana. New Phytol. 184:918–931. https://doi.org/10.1111/j.1469-8137.2009.03004.x.

Torres-Martinez, H.H., Rodriguez-Alonso, G., Shishkova, S., and Dubrovsky, J.G. (2019). Lateral root primordium morphogenesis in angiosperms. Front Plant Sci. 10:206. https://doi.org/10.3389/fpls.2019.00206.

Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J. (2005). The botany array resource: e-northerns, expression angling, and promoter analyses. Plant J. 43:153–163. https://doi.org/10.1111/j.1365-313X.2005.02437.x.

Versaw, W.K., and Harrison, M.J. (2002). A chloroplast phosphate transporter, PHT2;1, influences allocation of phosphate within the plant and phosphate-starvation responses. Plant Cell 14:1751–1766. https://doi.org/10.1105/tpc.002220.

Wang, D., Lv, S., Jiang, P., and Li, Y. (2017). Roles, regulation, and agricultural application of plant phosphate transporters. Front Plant Sci. 8:817. https://doi.org/10.3389/fpls.2017.00817.

Wang, Y., Ribot, C., Rezzonico, E., and Poirier, Y. (2004). Structure and expression profile of the Arabidopsis PHO1 gene family indicates a broad role in inorganic phosphate homeostasis. Plant Physiol. 135:400–411. https://doi.org/10.1104/pp.103.037945.

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.e556. https://doi.org/10.1016/j.devcel.2020.12.015.

Yanai, I., Benjamini, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650–659. https://doi.org/10.1093/bioinformatics/bti042.

Yang, Z.B., He, C., Ma, Y., Herde, M., and Ding, Z. (2017). Jasmonic acid enhances al-induced root growth inhibition. Plant Physiol. 173:1420–1433. https://doi.org/10.1104/pp.16.01756.

Zander, M., Lewsey, M.G., Clark, N.M., Yin, L., Bartlett, A., Saldierna Guzman, J.P., Hann, E., Langford, A.E., Jow, B., Wise, A., et al. (2020). Integrated multi-omics framework of the plant response to jasmonic acid. Nat. Plants 6:290–302. https://doi.org/10.1038/s41467-020-0605-7.

Zhang, H., Zhang, F., Yu, Y., Feng, L., Jia, J., Liu, B., Li, B., Guo, H., and Zhai, J. (2020). A comprehensive online database for exploring approximately 20,000 public Arabidopsis RNA-Seq libraries. Mol. Plant 13:1231–1233. https://doi.org/10.1016/j.molp.2020.08.001.

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. https://doi.org/10.1038/s41467-021-22362-4.

Zhao, Y., Xing, L., Wang, X., Hou, Y.J., Gao, J., Wang, P., Duan, C.G., Zhu, X., and Zhu, J.K. (2014). The ABA receptor PYL8 promotes lateral root growth by enhancing MYB77-dependent transcription of auxin-responsive genes. Sci. Signal 7:ra53. https://doi.org/10.1126/scisignal.2005051.

Zhu, W., Miao, Q., Sun, D., Yang, G., Wu, C., Huang, J., and Zheng, C. (2012). The mitochondrial phosphate transporters modulate plant responses to salt stress via affecting ATP and gibberellin metabolism in Arabidopsis thaliana. PLoS ONE 7:e43530. https://doi.org/10.1371/journal.pone.0043530.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol. 136:2621–2632. https://doi.org/10.1104/pp.104.046367.