the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) **115,** 1486–1499 doi: 10.1111/tpj.16351

PERSPECTIVE

Exploring plant *cis*-regulatory elements at single-cell resolution: overcoming biological and computational challenges to advance plant research

John Pablo Mendieta, Ankush Sangra, Haidong Yan, Mark A. A. Minow and Robert J. Schmitz*

Department of Genetics, University of Georgia, Athens 30602, Georgia, USA

Received 20 April 2023; revised 6 June 2023; accepted 8 June 2023; published online 13 June 2023.
*For correspondence (e-mail schmitz@uga.edu).

SUMMARY

***Cis*-regulatory elements (CREs) are important sequences for gene expression and for plant biological processes such as development, evolution, domestication, and stress response. However, studying CREs in plant genomes has been challenging. The totipotent nature of plant cells, coupled with the inability to maintain plant cell types in culture and the inherent technical challenges posed by the cell wall has limited our understanding of how plant cell types acquire and maintain their identities and respond to the environment via CRE usage. Advances in single-cell epigenomics have revolutionized the field of identifying cell-type-specific CREs. These new technologies have the potential to significantly advance our understanding of plant CRE biology, and shed light on how the regulatory genome gives rise to diverse plant phenomena. However, there are significant biological and computational challenges associated with analyzing single-cell epigenomic datasets. In this review, we discuss the historical and foundational underpinnings of plant single-cell research, challenges, and common pitfalls in the analysis of plant single-cell epigenomic data, and highlight biological challenges unique to plants. Additionally, we discuss how the application of single-cell epigenomic data in various contexts stands to transform our understanding of the importance of CREs in plant genomes.**

Keywords: gene regulation, epigenomics, cis-regulation, enhancers, silencers, single-cell genomics.

INTRODUCTION

Multicellular eukaryotes arose due to evolutionary pressures driving the sub-functionalization of cells into dedicated roles, allowing organisms to have functions that are more advanced than its cellular components. Cellular specialization results from differential use of the genome between cell types, which is partly driven by variable use of *cis*-regulatory elements (CREs) that are important for gene transcription and silencing. In plants, cell types have evolved specialized metabolisms and unique cell wall morphologies that link form to function enabling cells to fill their structural and physiological roles *in planta* (Alberts et al., 2002). Plant cell structures fascinated early plant scientists; the observation of microscopic cell wall 'cages' within onion leaves led Robert Hooke to develop the term 'cell' and variable cell wall morphologies were first used to classify plant cell types (Hooke, 1665). However, plant cell-type definitions have been continually refined by sequential scientific breakthroughs such as the increased resolution of microscopy and advances in molecular genetic techniques. Advances in single-cell genomics allow measurement of cell-type-specific transcripts and CREs, which is a bettering our understanding of the gene regulatory networks present in cells, and how they impact all manner of phenomena in *planta*. However, there are numerous technical and biological challenges associated with single-cell genomics data that must be overcome before these questions can be addressed.

In this perspective, we discuss the historical ways plant cell types have been described and how cell-type definitions have evolved. We examine how cell types have been defined genetically, and how this identified marker genes critical for cell-type function. Additionally, we discuss current biological and technical challenges associated with the single-cell genomics identification of cell-type-specific regulatory sequences. Lastly, we highlight how emerging technologies will overcome some of these challenges, improving the ability to study the

cellular context in which molecular processes affect plant phenotypes.

PLANT CELL TYPES – A HISTORICAL PERSPECTIVE

Scientists have been describing the cellular makeup of plants for centuries. Plant cell biology began with the advent of microscopy and histology, with early descriptions of stomata and guard cells dating back to 1671 ('Anatomie des plantes', n.d.). This research laid the groundwork of plant anatomy and established early models of plant cell types. Cells were classified based on the structure of their cell walls, with parenchyma having thin non-lignified cell walls, collenchyma having thick non-lignified cell walls, and sclerenchyma having lignified cell walls (Imperatorskaia͡ akademiia͡ nauk (Russia) et al., 1868). Although critical, these early descriptions of plant 'cell types' had limited resolution and overlooked cells with unique structure and function. Advances in microscopy in the following centuries facilitated more accurate descriptions of plant cell types. Increasing microscopy resolution produced descriptions of cell-type subclasses within the classical definitions of parenchyma, collenchyma, and sclerenchyma (Leroux, 2012). This led to the first described companion cells, sieve tube elements, and bundle sheath cells (Strasburger, 1888; Wilhelm, 1880). These newly described cell types were not just categorized but were described in their developmental and gross anatomical contexts within the plant. Foundational work by Esau and Sharman combined microscopy with serial sectioning experiments, describing vascular development in multiple plant species (Esau, 1939, 1954; Sharman, 1942). This combination of techniques revealed the cellular patterns in mature tissue, and how these arrangements emerge from their cellular precursors (Esau, 1943; Sharman, 1942). Further work focused on the meristem, a collection of plant stem cells that divide to produce new growth. Tracking cellular division and maturation from meristems provided an early understanding of plant cell-type differentiation, revealing how anatomical patterns are established by development (Evert et al., 2006).

Early on it was understood that DNA encoded the genetic instructions which give rise to plant form, but our understanding of the genetic processes that controlled cell-fate decisions were limited. Initial genetic analysis exploited the clonal development of mutant sectors with visible phenotypes. In brief, these studies used mutagens, like X-rays, to induce somatic mutations in progenitor cells to determine the cell's contributions to organismal phenotype. In plants, mutant-based studies demonstrated that manipulation of DNA sequence could radically change plant phenotypes and cell fates (Hake & Freeling, 1986; Sinha & Hake, 1990). For instance, stable mutagenesis gave rise to *liguleless-1* mutants that have radically different leaf morphology with a misplaced ligule on the margin of the leaf blade (Becraft & Freeling, 1991). However, these studies were limited in their capacity to identify the sequence causing these morphological alterations. This quickly changed with advances in molecular genetic techniques that allowed for pinpointed manipulation of plant DNA.

In the 1990s, molecular genetic techniques allowed for fine-scale alteration of DNA and inquiry into the genetic processes driving the emergence of specific cell types. Early genetic screens found that cell identity could be ablated by single-gene knockouts. One excellent example is *shortroot* (*shr)* in *Arabidopsis thaliana* (Benfey et al., 1993). In *shr* mutant plants, root endodermis cells fail to form, resulting in significantly stunted root growth and illustrating that *SHR* is indispensable for endodermis cell-type identity. Further analysis of *SHR* revealed it is a mobile transcription factor critical for cell fate differentiation (Helariutta et al., 2000). The identification of *SHR* and other transcription factors that defined cell identity generated questions aimed at how cell fates were encoded within the genome. These questions remain the subject of active investigation, with ongoing experiments continually offering deeper insights into the molecular events that drive plant cell-fate decisions.

THE GENETIC UNDERPINNINGS OF PLANT CELL IDENTITY

Plant cell development and function result from a complex interplay of genes responsible for determining cell fate and maintaining cellular identity. Identification of key developmental regulators, like *SHR*, demonstrated that the development of entire cell lineages depended on the expression of a few genes. Determining how and where these essential 'marker' genes of cell identity were expressed became a central question in plant genetics. Subsequently, molecular genetic approaches such as mutagenesis screens, and reporter gene assays, were developed to assess the cellular context in which these marker genes were expressed. These advancements resulted in the identification of many other genes critical in cell-type identity. For example, *GLABRA1* (*GL1*) in *A. thaliana* controls trichome fate, as *gl1* null plants generated by T-DNA insertion had no trichomes on the leaf and stem (Figure 1a) (Herman & Marks, 1989; Oppenheimer et al., 1991). Despite establishing the necessity of *GL1* for trichome formation, this finding did not elucidate its expression pattern or how *GL1* facilitated trichome development. This knowledge gap led to the creation of promoter-reporter lines, in which a gene's transcriptional regulatory sequences (promoters and CREs) are fused to a reporter (e.g., GUS, GFP) to illuminate where and when the gene is expressed (Birnbaum et al., 2003; Brady et al., 2003; Helariutta et al., 2000; Stadler et al., 2005). In *GL1* reporters, expression was found to change throughout development; in early development, *GL1* is expressed throughout the early leaf primordia, but, as the epidermis matures, only

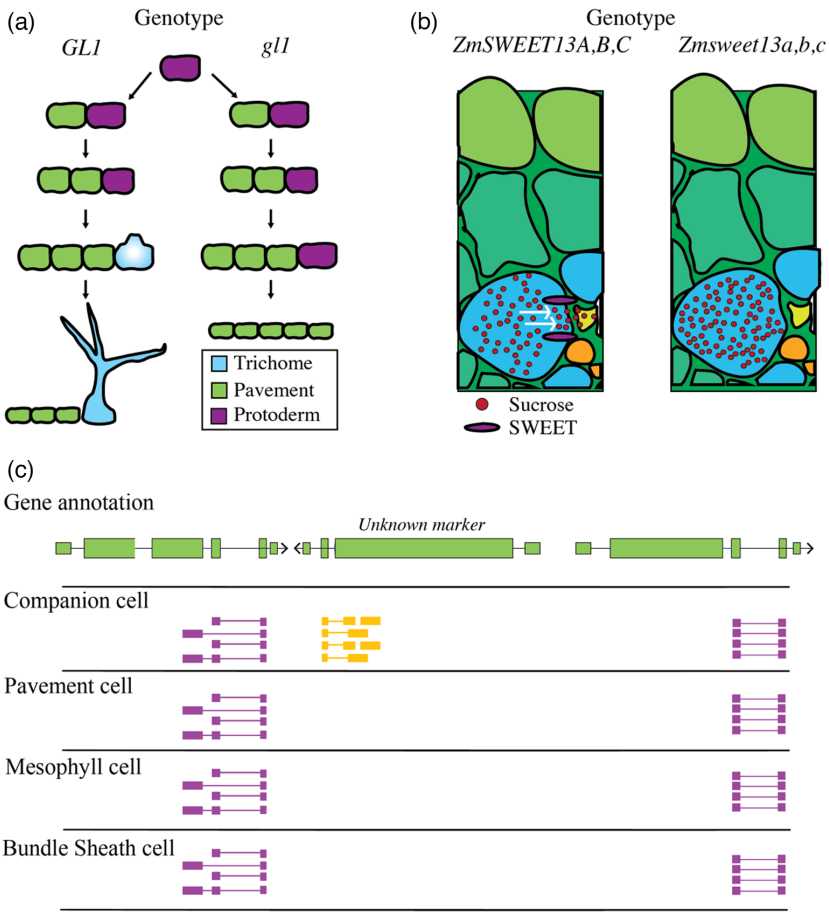

Figure 1. Plant cell-type markers define either unique developmental, metabolic, or physiological states.
(a) Model for proper function of *GLABRA1* in *Arabidopsis thaliana (GL1)*, which promotes trichome development (left). Knockouts of *gl1* remove the capacity for protoderm cells to differentiate into trichomes, generating additional pavement cells (right).
(b) *ZmSWEET13s* (purple) are required for the transport of sucrose from bundle sheath cells into the vasculature in *Zea mays*. *zmsweet13* knockouts raise sucrose concentrations in bundle sheath cells.
(c) Hypothetical example of a *de novo* discovered marker gene identified by single-cell RNA-seq. Expression of the *de novo* discovered marker, *Unknown*, is limited to companion cells, as opposed to pavement cells, mesophyll cells, and bundle sheath cells. Single-cell RNA-seq reads are colored by their strand, with purple reads representing the positive strand and yellow reads representing the negative strands.

trichomes precursors maintain high *GL1* expression (Kirik et al., 2001; Larkin et al., 1994; Oppenheimer et al., 1991). Research into genes crucial for cell development expanded reporter methods by combining cell-type reporters with protoplast isolation to isolate cell populations and conduct genome-wide identification of transcription factors associated with specific cell types (Birnbaum et al., 2005; Toufighi et al., 2005). Application of these genome-wide assays identified genes crucial for the development of particular cell types. Presently, cell-type-specific genetic inquiry in plants has the potential to be significantly enhanced through single-cell methodologies, allowing for refined discrete measurement from individual cells and empowering our understanding of plant cell fate decisions.

While genes important in the development of cell types are critical to our understanding of how cell types differentiate, they do not reveal much about plant cell-type function. This has led to researchers looking for genes that are important to the function of mature cell types. For instance, genes such as *SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTER 13* (*ZmSWEET13*), a sucrose transporter, is expressed specifically in bundle sheath cells and phloem parenchyma (Bezrutczyk et al., 2018, 2021). Knockouts of *ZmSWEET13* impair phloem loading increasing sucrose concentrations in leaves (Figure 1b). Although not required for abaxial bundle sheath cell development, *ZmSWEET13* represents a key gene required for cell-type-specific function. Similarly, Arabidopsis *SUCROSE-PROTON SYMPORTER 2* (*SUC2*) drives sieve element sucrose loading through companion cell-specific expression (Stadler & Sauer, 1996). *suc2* knockout plants are stunted due to impaired sucrose transport, but companion cell identity is unaffected (Gottwald et al., 2000). Genes such as *SUC2* and *ZmSWEET* further our understanding of the genetic partitioning of

functions between plant cell types. This genetic division enhances our perspective of what constitutes a plant cell type, transitioning from definitions based on histology to those based on gene expression and the function of discrete genetic loci. Although finding cell-type-specific functional genes is valuable, their identification is generally done by investigating a single gene at a time, requiring a significant investment of time and resources. Single-cell genomics provides an opportunity to discover additional genes important to the function of specific cell types on a genome-wide scale, across all cell types sampled at a single time. This influx of information will quickly evolve our understanding of cell types from a few key loci to combinations of genes critical for both development and function.

Single-cell genomics allows the measurement of chromatin states and mRNAs in thousands of individual cells (Buenrostro, Wu, Litzenburger, et al., 2015; Cusanovich et al., 2015; Jaitin et al., 2014). Plant single-cell genomics is especially exciting given the lack of cell-type-specific genomic measurements outside of model plant systems. The information-richness and high complexity of single-cell datasets are useful because it allows for a detailed understanding of how different cell types utilize the genome. However, single-cell technologies remove cells from the sampled tissue, erasing any knowledge about position or identity, and complicating the identification of each cell's cell type. Therefore, annotation relies on molecular marker genes to reveal cell identity *post hoc*. This annotation is confounded by the gradient of transitional cell identities that underpin differentiation. For instance, in *A. thaliana* guard cell differentiation from protoderm involves five state transitions, necessitating additional markers to accurately delineate cellular states (Chen et al., 2020). These transitory states make having well-established developmental marker genes critical to the accurate annotation of single-cell datasets. For this reason, the first plant single-cell RNA-seq (scRNA-seq) analysis was conducted on *A. thaliana* roots because root cell types have well-described genes associated with specific cell types and developmental stages (Ryu et al., 2019; Shulse et al., 2019). Once single-cell datasets are accurately annotated, they can be leveraged in powerful ways. Testing for differentially expressed genes in annotated cell types allows for the identification of novel genes potentially critical in proper cell-type function. One scRNA-seq *A. thaliana* study used annotated root cell types to discover 50 genes with cell-type-specific expression patterns (Figure 1c) (Zhang et al., 2019). This *de novo* discovery of cell-type-specific genes provides a wealth of candidate genes to target and study, which will further reveal their importance in specific cell types of interest. Single-cell genomics will increase the speed of cell-type-specific gene identification, improving our understanding of which loci are critical for proper cell-type function and development in plants.

THE REGULATORY GENOME SPECIFIES HOW CELL TYPES ARE ESTABLISHED

Although scRNA-seq will improve our knowledge of cell-type-specific gene expression, our understanding of the processes driving these expression patterns remains poor. Pairing of single-cell technologies with assays identifying the regulatory genome stands to greatly enhance our understanding of how the genome can regulate the expression of both developmental and functionally important genes. Cell-type-specific expression is the result of different cells using the same genetic blueprint encoded in the genome in different ways. Cell fates are determined by the interpretation, enhancement, or silencing of instructions encoded in DNA which are driven by CREs (Andersson et al., 2015). CREs are non-coding sequences of DNA composed of transcription factor (TF) binding sites. TFs bind CREs within nucleosome-depleted sequences, to recruit co-factors, remodel chromatin, and regulate gene transcription (Lai et al., 2019). This *cis*-regulation has implications for plant development, environmental response, and evolution (Cramer, 2019).

CREs often work in concert and are then referred to as *cis*-regulatory modules (CRMs) (Figure 2) (Schmitz et al., 2022; Shlyueva et al., 2014). CRMs are further subdivided as 'enhancers', or 'silencers', based on the ability to recruit co-activators or co-repressors to genes (Gisselbrecht et al., 2020; Pang & Snyder, 2020; Shlyueva et al., 2014). Identification of CRMs genome-wide is routinely performed with assays that measure accessible chromatin environments, as these are the regions that are open to TF binding. Methods such as DNase-seq, MNase-seq, as well as FAIRE-seq have been used to study CRMs genome-wide (Boyle et al., 2008; Giresi et al., 2007; Johnson et al., 2006). Currently, the most widely adopted method to investigate accessible chromatin is Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq) (Buenrostro, Wu, Chang, & Greenleaf, 2015). In brief, ATAC-seq works by utilizing a hyperactive Tn5 transposase to directly insert sequencing adapters into accessible chromatin regions of DNA (Figure 2). The fragments generated are then amplified, sequenced, aligned to the genome, and areas more accessible than the genomic background are computationally identified (Figure 2) (Yan et al., 2020). These peaks, named accessible chromatin regions (ACRs), are well-accepted proxies for CRMs, and thus collections of CREs (Bajic et al., 2018; Lu et al., 2017).

With the widespread adoption of ACR identification, numerous discoveries have been made about the regulatory nature of DNA in plant genomes. For instance, it has recently been revealed that ACRs frequently operate on genes >50 kilobases away in plants with large genomes (Ricci et al., 2019). Additionally, variable ACR usage has been implicated in biotic and abiotic stress responses, providing more insights into how the genome tunes

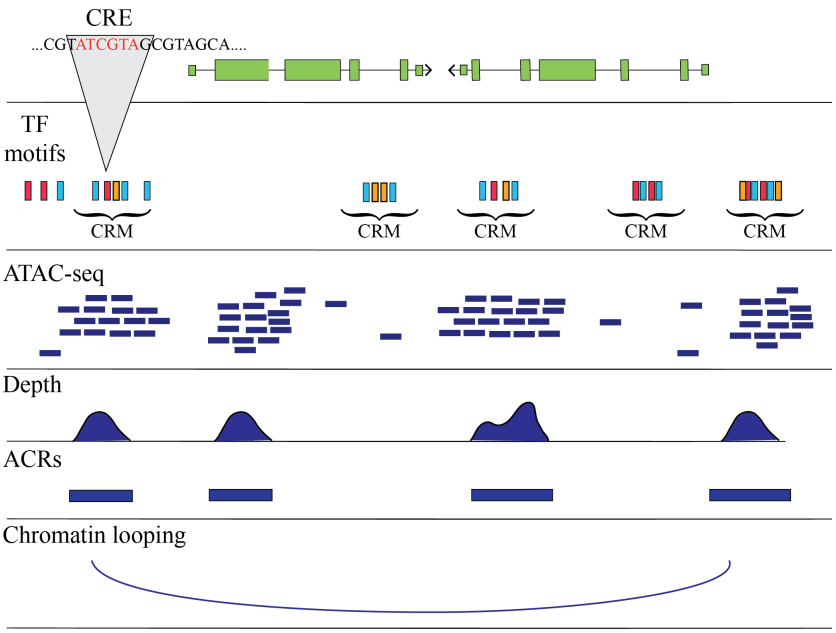

Figure 2. Deciphering the regulatory genome with ATAC-seq. Top: Gene annotation **TF Motifs**: Transcription factor binding motifs. Colors represent different motifs. Motifs occur in clusters referred to as *cis*-regulatory modules (CRMs). Gray arrow highlights a specific TF motif acting as a CRE. **ATAC-seq**: Example of aligned reads from an ATAC-seq experiment. Note that reads align heavily with CRM regions. **Depth**: Histograms of the above read depth. **ACRs**: Identified accessible chromatin regions from the read depth above. **Chromatin looping**: An example of how ACRs do not always operate on the closest gene. The line represents the chromatin interaction between two ACRs.

expression to the environment. (Han et al., 2020; Raju, 2020; Zeng et al., 2019; Zhou et al., 2022). However, ACRs provide no information about whether these regions are enhancing or repressing transcription. This can be predicted by overlaying ACRs with ChIP-seq data which measures the histone modifications nearby (Lu et al., 2019; Oka et al., 2017; Ricci et al., 2019). ACRs that are active are flanked by histone acetylation, whereas those that are actively repressing a target gene are flanked by histone methylation and Polycomb silencing (Lu et al., 2019; Ricci et al., 2019). Recently, the ability to apply ATAC-seq to single cells (scATAC-seq) was developed, allowing for cell-type-specific ACR identification and measurement (Buenrostro, Wu, Litzenburger, et al., 2015; Cusanovich et al., 2015).

SINGLE-CELL ATAC-SEQ, EMERGING PARADIGMS, AND TANGIBLE VALUE

ScATAC-seq has revealed differential usage of ACRs in cellular identity and development in plant models (Dorrity et al., 2021; Farmer et al., 2021; Marand et al., 2021). Although the application of scATAC-seq techniques to plants stands to teach us much about *cis*-regulatory biology, implementing these techniques are non-trivial and come with a series of caveats and challenges. The ability to deconvolute cellular heterogeneity in plant tissue allows for the identification of cell-type-specific ACRs, which can be used to identify TFs and CREs important to cell function. (Marand et al., 2021). Intriguingly, scATAC-seq also offers a method to study developmental trajectories within cell types. Key genes or CREs that operate differently through development can be identified by ordering cell lineages from progenitor to mature cell type (Nelms & Walbot, 2019; Trapnell et al., 2014). This 'pseudo time' method was applied to root hair development in *Oryza sativa*, as well as phloem companion cell development in *Z. mays* root. In *O. sativa*, pseudotime analysis found 13 000 ACRs and 3000 genes important in the transition into root hair cell-type identity (Zhang et al., 2021). In *Z. mays*, it was found that as cells differentiate from quiescent center cells to phloem companion cells the fractions of ACRs that were accessible decreased significantly (Marand et al., 2021). Pseudotime analyses is just one powerful example of the usage of scATAC-seq. Additional application in plants will reveal the *importance cis*-regulation plays in evolution, stress responses, and adaptation. While exciting, the analysis and annotation of these datasets are computationally challenging and require awareness of current limitations.

The computational challenges associated with scATAC-seq data analysis are primarily due to the low number of Tn5 integration events per cell. For example, upwards of 99% of the chromatin accessibility measurements genome-wide are often missing from any particular cell (Buenrostro, Wu, Litzenburger, et al., 2015). This data scarcity has significant ramifications in scATAC-seq analysis. The first step of scATAC-seq analysis is isolating high-quality cells. One way of doing this is by 'pseudo-bulking',

which mimics bulk ATAC-seq by aggregating the reads from all nuclei, to identify peaks (ACRs) (Chen et al., 2019). Then broken nuclei are removed by estimating the Fraction of Reads in Peaks (FRiP) per nucleus, and removing nuclei with Tn5 integration events below a FRiP threshold (generally >0.25). Next, doublets, which are instances where two cells are mistakenly sequenced as one are removed by comparing them against an *in silico* generated doublet set of cells (Wolock et al., 2019). Based on the single-cell technologies used, the top 5–10% of cells with the highest 'doublet score' are removed. The next steps annotate similar cells into cell types. Annotation starts by generating a binary matrix of Tn5 insertions in ACRs by cells, which is fed into dimensionality reduction algorithms. These algorithms, such as singular value decomposition (SVD) or principal component analysis (PCA), cluster cells into similar groups by identifying correlated features (ACR presence/absence), which reveal underlying patterns and relationships among the cells (Figure 3a,b). The resulting principal components, or meta-features, represent high-dimensional data (ACR accessibility) in low-dimensional space. Cell proximity in this low dimensional is a proxy for cell relatedness, either biological or technical (Figure 3b). Presently, either Uniform Manifold Approximation and Projection (UMAP) or t-Distributed Stochastic Neighbor Embedding (t-SNE) are used to visualize scATAC-seq data. These techniques plot cells in 2D while trying to preserve the high-dimensional space computed above (Figure 3b) (Maaten & Hinton, 2008; McInnes et al., 2018). One should not make biological conclusions about relative distance and space between cells, as recent evidence points to the inherent flaws in this approach (Chari et al., 2021). For instance when using three-dimensional datasets with known spatial relationships between points, t-SNE, or UMAP processing scrambles the relative distance between points, indicating that the 2D distances generated are artifactual (Chari et al., 2021). From this embedding, discrete cell clusters are assigned using community detection methods such as Louvain or Leiden algorithms (Blondel et al., 2008; Waltman & van Eck, 2013). In brief, these methods work by trying to identify clusters of cells in high-dimensional space, which maximizes the differences between groups and minimizes the differences within groups based on a given parameter set. Clusters are then analyzed with the assumption that they are representative of roughly homogenous cell types. Annotating clusters to a cell type involves approximating gene expression of cell-type markers by summing gene body and promoter chromatin accessibility (Cusanovich, Hill, et al., 2018). This approximation, while valuable, is imperfect, as chromatin accessibility does not always correlate with gene expression (Figure 3d). Based on the specific chromatin accessibility patterns of known marker genes, clusters are assigned cell types (Figure 3c). The clustering and annotation of cell types remain one of the most time-consuming and difficult steps in scATAC-seq analysis. Current heuristic methods rely on user-based decisions that are often difficult to replicate (Gibson, 2022). As the field matures, more consistent annotation metrics are needed to ensure proper and timely cell-type assignment.

ScATAC-seq analysis is a deeply iterative process. For instance, selecting different ACRs to include in dimensionality reduction can drastically alter cluster membership and generate different results. This requires researchers to try different selections of ACRs to find a set that reduces technical artifacts but maximizes biological interpretation. Technical artifacts can have significant effects on annotations and interpretation of results. For example, cells with a high density of Tn5 insertion events per cell can cluster together, thus the underlying embedding does not represent one of biological variation, but of technical variation (Figure 3b). Technical artifacts can be even more misleading, with cells being assigned specific clusters due to the lack of data, rather than the presence of genuine differences.

Once annotations are finalized, cell-type-specific ACRs are identified. Combining cells of the same cell type via '-pseudo-bulking' allows for the robust identification of ACRs for individual cell types (Cusanovich, Reddington, et al., 2018; Domcke et al., 2020). This deconvolution of tissue-level chromatin accessibility to cell-type resolved accessibility is where the power of a single cell lies. While identifying ACRs from cell-type-level data is straightforward, classifying these ACRs as cell-type-specific or broadly accessible is challenging and is heavily impacted by the statistical approach chosen (Figure 3e). Making this categorization more opaque is cell types that share developmental origins often have similar chromatin accessibility patterns (Cusanovich, Hill, et al., 2018; Domcke et al., 2020). This leads to an additional class of ACRs that are cell-type-restricted or limited in their chromatin accessibility to a few cell types, but not truly cell–type specific. Recent plant scATAC-seq studies have found between 23% and 27% cell-type-restricted ACRs in given species (Marand et al., 2021; Zhang et al., 2021). However, the number and proportion of ACRs that are cell-type specific is unknown, and being established in model systems with more exhaustive sampling (Chen et al., 2018; Domcke et al., 2020). Whether these cell-type-specific ACRs are critical to cell-type function is uncertain and requires follow-up molecular genetics studies. Finally, scATAC-seq provides exciting opportunities to begin deciphering both *cis* and *trans* regulators of the genome. Recent studies have shown the ability to link TFs with their likely binding sites in a cell-type-specific manner, by correlating the chromatin accessibility of transcription factor gene bodies with the accessibility of their corresponding binding sites (Marand et al., 2021). This allows for the identification of cell-type-specific gene regulatory networks which have been long

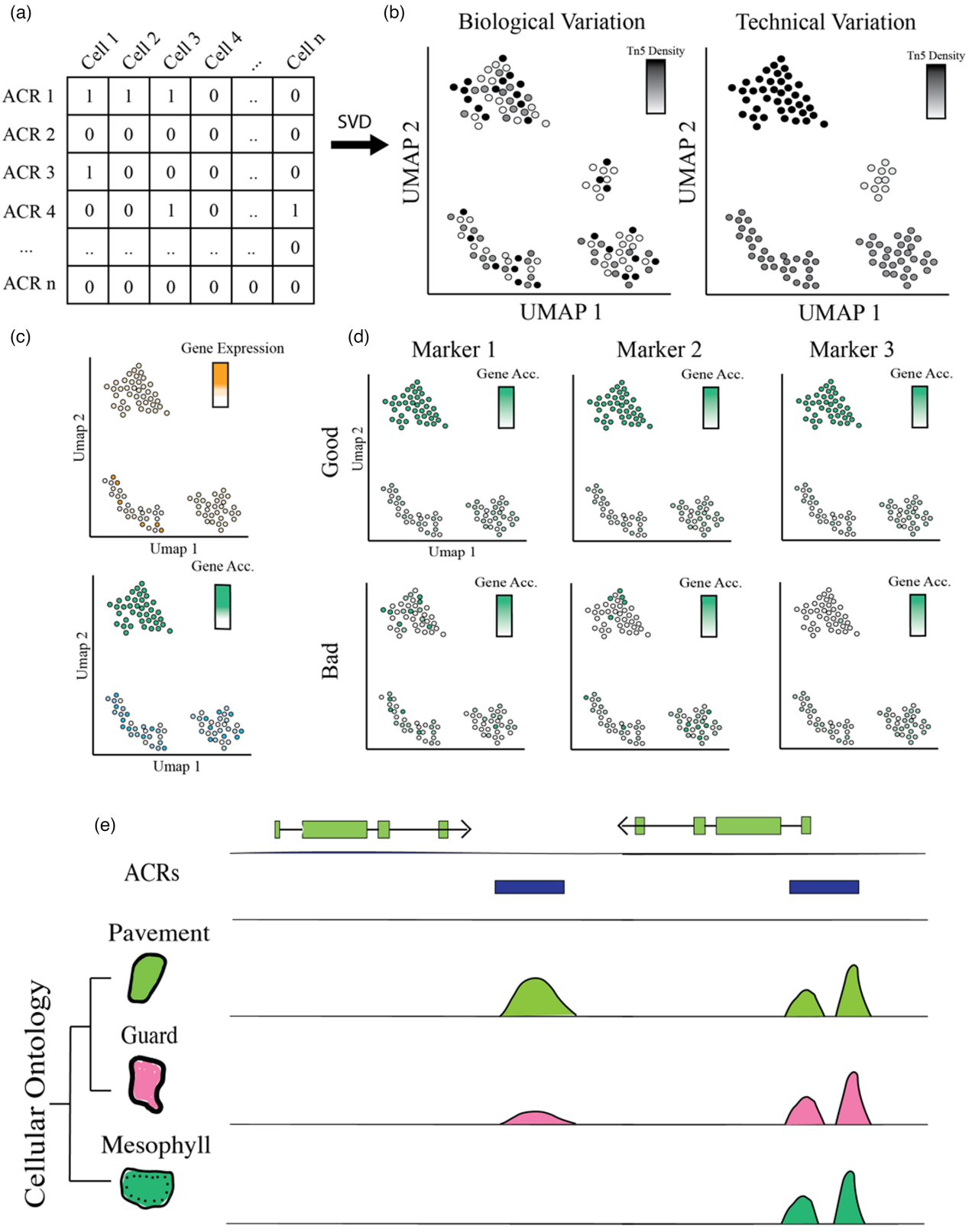

Figure 3. Schematic of analysis paradigms and challenges of single-cell ATAC-seq data.
(a) An example binary matrix being transformed into a UMAP by means of SVD where rows are ACRs identified from a bulked dataset and columns are cells. The input matrix is binary if a cell either has a Tn5 integration event in that ACR or not.
(b) (Left) An example of a good UMAP embedding where Tn5 insertion density is not driving the clustering of cells. (Right) An example of a poor UMAP embedding where technical artifacts due to Tn5 insertion density are driving the clustering.
(c) An instance where chromatin accessibility of a gene (Gene Acc.) does not align with expectations based on gene expression (Gene Expression). Here, the upstream region of the gene depicted could harbor a silencer that recruits TFs to repress its target gene.
(d) (Top) An example of where the UMAP embedding correlates well with the chromatin accessibility of multiple marker genes for a specific cell type. (Bottom) Example of a poor UMAP embedding is where markers do not agree with each other. Whether this is due to poor markers or a poor selection of features to generate the embedding requires additional inquiry.
(e) Challenges associated with the assignment of single-cell ACRs. Whether the middle ACR is cell-type specific or restricted to a few key cell types is challenging to determine. This potentially reflects issues in annotation, or that related cell types share similar chromatin environments. The assignment of cell-type specific ACRs is non-trivial and requires careful considerations by the researcher of both biological and technical challenges.

elusive. While the computational workflow and challenges labeled here may seem daunting, rigorous data analysis avoids many of these pitfalls. However, it should be noted that these specific computational challenges aren't the only issue. Quirks associated with evolution, genome structure, and the unique ways plant cell-type identity can be modified also need to be considered.

BIOLOGICAL CHALLENGES ASSOCIATED WITH THE ANALYSIS OF SCATAC-SEQ IN PLANTS

While the computational challenges associated with scATAC-seq are laid out and addressable, certain unique features of plant biology complicate analysis. Variable genome sizes, rapid changes in gene function caused by molecular evolution, and the totipotent nature of plant cells all alter the interpretation of plant scATAC-seq data. However, while analytically challenging, these features offer unique opportunities to study CREs and their relationship to plant biology generally.

Significant variations in genome size can add additional hurdles to analyzing scATAC-seq data. For instance, genome size affects the use of gene proximal chromatin accessibility as a proxy for gene expression. In compact genomes, the reduced proximity between transcriptional start sites (TSSs) means ACRs often encompass two promoters, which convolutes correlating chromatin accessibility with gene expression (Figure 4a). This is in stark contrast to larger genomes which result from the expansion of intergenic and intronic gene space often as a result of increased transposon load (Lee & Kim, 2014; Lu et al., 2019). This size expansion moves CRMs important for gene expression further upstream of the TSS, increasing the prevalence of gene-distal ACRs (Lu et al., 2019; Oka et al., 2017; Ricci et al., 2019; Zhao et al., 2018). Linking these distal ACRs to their target genes is challenging and requires additional data from proximity-ligation-based methods. Hi-C is the most commonly used, as it captures chromatin interactions ranging from 1 kb to >100 kb depending on the experimental setup and sequencing depth. (Figure 2) (Eagen et al., 2015; Lieberman-Aiden et al., 2009; Mifsud et al., 2015). However, Hi-C remains restricted to bulk tissues, limiting the detection or confirmation of gene-ACR interactions in rare cell types. Predictions about chromatin contacts can be made from scATAC-seq itself but requires further experimentation for validation (Marand et al., 2021; Pliner et al., 2018). Genome size variation necessitates adapting analysis strategies on a per-genome basis, impeding the standardization of scATAC-seq analysis between species.

Although variation in plant genome size complicates scATAC-seq, the lack of high-quality markers for plant species remains the biggest challenge in the annotation and analysis of single-cell ATAC-seq datasets. At present, no species has a comprehensive list of genetic markers specific to each cell type. Consequently, researchers often use markers from one species to aid in the identification of cell types in another species. In the case of non model plants, cell types are identified using gene orthologs borrowed from known cell–type–specific markers in model plant species. However, it is known that gene expression changes rapidly due to molecular evolution (Hill et al., 2020). Gene duplication followed by neo-functionalization, whole genome duplications rewiring large-scale expression patterns, and rapid gene family expansion is a few of the many ways molecular evolution can reshape gene function, and expression (Birchler & Yang, 2022; Hughes et al., 2014; Panchy et al., 2016). Even in a relatively short evolutionary period of 65–70 million years key developmental genes can shift their cell-type expression context, complicating their use in a cross-species context (Hughes & Langdale, 2022). For example, *SHR* has different cell-type specificity in *Z. mays* and *O. sativa* leaves; in *Z. mays*, expression of *ZmSHR1* is limited to the vasculature and bundle sheath cells, whereas the *O. sativa* ortholog, *OsSHR2*, has limited vasculature expression, and is absent from bundle sheath cells (Figure 4b) (Schuler et al., 2018). Due to the unreliability of individual markers, non-model systems need to use sets of markers to annotate cell types in order to Minimize incorrect cell annotation. However, to date the number of markers per cell type is limited, restricting the ability to apply single-cell techniques to non-model plants.

Plants cells have a unique relationship between cell identity by descent, and cell position in the plant. Plant cell fates are not genetically hardwired based off precursors. For instance, although plant cell types generally develop in

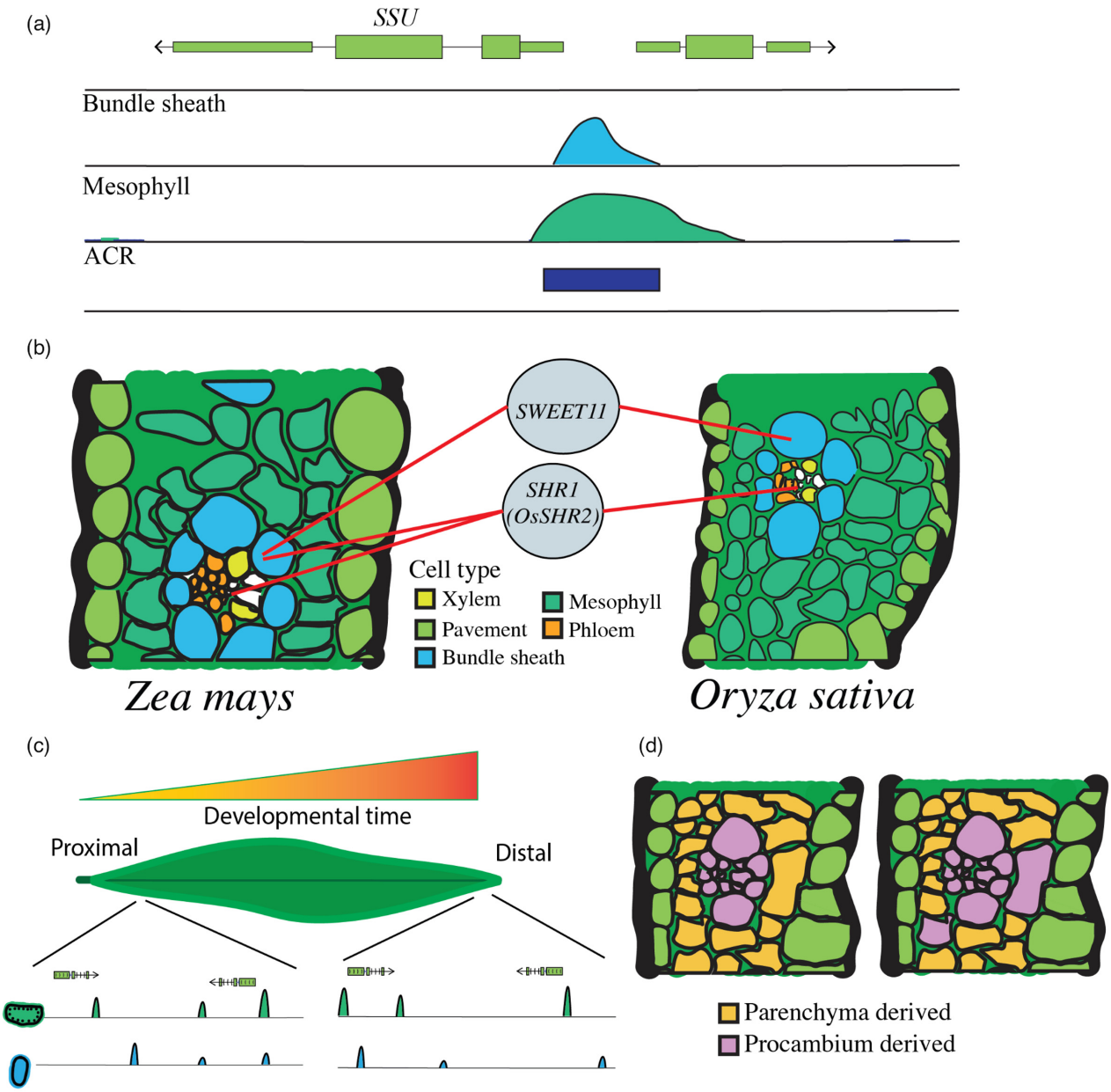

Figure 4. Biological challenges in single-cell ATAC-seq data take many forms and unique situations.
(a) The gene *SHORT SUBUNIT (SSU)* is known to have a specific function in bundle sheath cells in *Zea mays*. However, SSU may not behave as a useful marker for annotation via scATAC-seq due to the proximity of its promoter to the start of an uncharacterized gene.
(b) Leaf cross sections of two species, *Z. mays* and *Oryza sativa*. Cell types are color-coded, and important marker genes are labeled in gray circles. Red lines point to the cell types these genes are active in. Although *SHR1* is expressed in bundle sheath and vascular cells in *Z. mays*, it is not found in *O. sativa* bundle sheath cells.
(c) Leaf laid out from proximal to distal ends with developmental gradient overlaid on top, with the oldest cells being at the tip and the newest cells at the base. It is currently unknown whether different regions of the leaf might have different regulatory chromatin environments and subfunctionalization (bottom).
(d) Leaf cross sections where colors indicate cells originating from either parenchyma (ground tissue) or procambium. (Left) Normal development and (Right) abnormal development with position-dependent effects altering the origin of a mesophyll cell derived from the procambium.

predictable lineages, such as the procambium giving rise to the vasculature, exceptions can occur. Instead, the location of a plant cell during development can have greater impact on cell fate than their stem cell niche of origin (Reinhardt et al., 2003). This has been shown in *Z. mays*, where the mesophyll cells neighboring the bundle sheath lineage may be derived from procambial cells, or ground meristem cells (Figure 4d) (Esau, 1943; Langdale et al., 1989; Sharman, 1942). This position-dependent effect is well documented for vasculature and epidermal cell types

in species including cotton, tobacco, and sunflower (Dolan & Poethig, 1998; Esau, 1954; Hung et al., 1998; Jegla & Sussex, 1989; Poethig & Sussex, 1985a, 1985b). This poses a challenge, as these cells which have undergone position-dependent effects are likely to cluster with cells that share the same precursor, and not with cells with the same terminal identity. This makes annotation more difficult and increases the heterogeneity in identified cell types. Finally, since these events are rare, isolating, and studying these populations is challenging, but could pose a valuable study system to understand the role of how variable precursors alter the chromatin environment of differentiated cells.

Finally, the gradient nature of plant growth provides additional challenges. Plant organs grow in a gradient of development, with younger cells found closer to the dividing meristem, and older cells further away. Continual organogenesis and development result in gene expression profiles that are dependent on the section sampled within an organ (Figure 4c). In *Z. mays* this developmental progression yields differences in the expression of key carbon metabolism enzymes at different sections of the leaf (Li et al., 2010; Pick et al., 2011; Wang et al., 2013). Whether these different sections of the leaf, and the cell types found within, constitute different cell types or specific sub-functionalization is up for debate, and further complicates placing cells into discrete categories (Zeng, 2022). This heterogeneity has already been hinted at in some studies. A combinatorial scATAC-seq and scRNA study of *A. thaliana* roots, found unique genetic and epigenomic markers in three different clusters of endodermal cells, illustrating that discrete sub-functionalization may happen within previously described cell types (Dorrity et al., 2021). The extent to which these clusters represent unique sub-functional cell types remains open and requires further exploration.

THE AGE OF SINGLE-CELL REGULATORY GENOMICS

scATAC-seq enables the genome-wide investigation into the function and importance of plant cell-type-specific CREs. Although we can now identify cell-type-specific CREs in plant genomes, our understanding of how these regions interact with the coding genome is still quite limited. Leveraging intra- and inter-genetic diversity, along with treatment conditions, stands to greatly improve our understanding of CREs in plant biology and their role in responding to environmental stimuli, population adaptation, and diversity, as well as reveal their importance over evolutionary time.

Performing scATAC-seq on a phenotypically diverse intra-specific population will clarify the influence of genetic CRE variation on phenotypes with cell-type resolution. Genetic variation in regulatory sequences can result in species adaptation to novel environments in both plant and animal systems (Cleves et al., 2014; Studer et al., 2011; van der Burg et al., 2020; Wucherpfennig et al., 2022). In plants, CRE variation in the flowering time gene CONSTANS underlies flowering time diversity in natural accessions of *A. thaliana* (Rosas et al., 2014). However, most studies addressing CRE genetic variation lack cell-type resolved data and therefore may overlook genetic variance in rare cell-type CREs that underpins local adaptation. Combining quantitative genetic approaches, like genome-wide association (GWA), with scATAC-seq, phenotypic associations, and chromatin accessibility variation can be correlated, potentially identifying the CREs, and cell types, underpinning trait variation within distinct populations (Das et al., 2022). Although a nascent area of study, the combination of scATAC-seq and population diversity may reveal how CRE genetic diversity alters the regulatory epigenome to shape species adaptation.

Beyond applying scATAC-seq to single species populations, comparative genomics focused on diverse species offers the opportunity to examine plant CRE evolution at a deeper timescale. Plant genomes exhibit a high rate of structural variation and sequence turnover as compared to animal genomes, causing rapid CRE turnover (Paterson et al., 2010). Highlighting the high rate of CRE change between even closely related species, a study comparing distal CREs between sister species *Z. mays* and *S. bicolor*, found approximately one-third of CREs were shared and accessible in the same tissue, one-third were novel to each lineage, and one-third shared sequence similarity but were not within accessible chromatin in the tissue examined (Lu et al., 2019). While CREs sequences change quickly, the gene regulatory networks they influence may be more stable. Investigating root hair cell type development in four eudicots found that although few orthologous CREs were conserved across all species, TF binding at key genes was preserved (Maher et al., 2018). Pairing comparative genomics analyses with scATAC-seq will allow investigation into the pace of CRE sequence changes in specific cell types within individual plant lineages. This approach will enhance both our understanding of plant CRE evolution and uncover conserved mechanisms underpinning plant adaptation and resilience to environmental changes.

Finally, CREs drive responses to environmental stimuli. Differential CRE usage is vital in response to disease, cold, drought, and hormonal signals (Azodi et al., 2020; Moore et al., 2022; Reynoso et al., 2019; Zou et al., 2011). One comparative genomics study examined CRE usage with a flooding treatment and identified root-specific CREs associated with flooding, which revealed shared motifs within flood-responsive CREs across four species studied, representing 123 million years of evolutionary divergence (Reynoso et al., 2019). This flooding research suggests that regulatory networks behind abiotic stress responses may be conserved for millions of years. Integrating scATAC-seq with environmental treatments will

identify the CREs crucial for cell-type-specific environmental responses. Beyond discovering environmentally dynamic CREs, this approach will find the cell types with the most responsive CRE usage in different conditions, revealing which cell types drive stress adaptation. This focus on cell-type responses could have far-reaching implications for our understanding of environmental response in plants, as a previous study has traditionally been restricted to organismal response.

PLANT CELL TYPES – DEFINITIONS IN FLUX

While the age of single-cell genomics stands to alter our understanding of plant cell biology, it is important to acknowledge that the definition of a cell type is in flux. In this perspective, we define a 'cell type' as a cell with unique molecular signatures, and that alteration of this signature modifies the form or function of a given cell type. However, although valuable, this definition has limitations. For instance, what is the threshold of molecular changes needed to separate related cells into distinct cell types? How many differentially expressed genes or differentially accessible CREs are needed to constitute a novel cell type? This problem becomes especially acute when trying to delineate plant cells in transitional identities, as developing plant cells exist along a continuum of maturity with few discrete stages. While the discussions surrounding plant cell-type classifications may appear semantical, it underpins real biological questions. How we define 'cell types' will have real implications for biologist moving forward (Efroni, 2018; Zeng, 2022).

Despite their immense development, maturity, and anatomical differences, inevitably, knowledge of plant cell types will be compared to what we know about animal cell types. In plants, there is wide variation in the number of identified cell types, with 55 cell types being identified in a recent *Z. mays* single–cell atlas and 180 cell types in *A. thaliana* (Lee et al., 2023; Marand et al., 2021). This contrasts significantly with animals, as in mouse brains alone there exists 45 types of inhibitory neurons (Hodge et al., 2019). The existence of fewer plant cell types could be explained by technological limitations and less intensive study than that found in animal models. Alternatively, the paucity of plant cell types may reflect real biological differences between plants and animals. Unlike mammals, plant cell divisions result in the incomplete separation of nuclei; cytokinesis ends with the deposition of a new cell wall (cell plate) that contains plasmodesmata pores that retain cytoplasmic, and endoplasmic reticulum, connections between the daughter cells (Burch-Smith & Zambryski, 2012). The interconnectedness of plant cells through plasmodesmata has large implications in plant biology and may fuel the differences between plant and animal cell types, as plant cells exist as a connected community, not individuals. This interconnectedness has led some to propose a more holistic 'organism-level' view. Instead of focusing on cells or cell types as the biologically meaningful units of study, the organismal theory proposes to focus on the entire organism, as plant cells rarely work in isolation (Kaplan & Hagemann, 1991). However, this organism-level perspective conflicts with the severe phenotypic alterations caused by mutants that eliminate specific cell types as detailed above. In either case, single-cell (epi)genomics will reveal more about why plant cell types are less numerous than their mammalian counterparts. These techniques provide unprecedented cellular resolution, and if the lack of plant cell types is driven by past technical limitations, single-cell genomics will usher in an era of discovery wherein many new discrete plant cell types will be unveiled. Alternatively, if these new techniques confirm a relatively small number of more homogenous plant cell types, it may provide credence to the notion that plant cells should be studied as a physiological unit, highlighting the importance of intercellular cooperation in plant biology. Single-cell regulatory genomics stands to enliven plant research and provides the toolset to address these basic questions about the cell-type composition of plants.

ACKNOWLEDGMENTS

We would like to thank KD for feedback on this perspective. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institute of Health under award number T32GM007103. Additionally, this work was supported by the National Science Foundation (IOS-1856627, IOS-2026554, and MCB-2120132) to RJS. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number DE-SC0023338 also to RJS.

REFERENCES

Alberts, B., **Johnson, A.**, **Lewis, J.**, **Raff, M.**, **Roberts, K.** & **Walter, P.** (2002) The plant cell wall. In: *Molecular biology of the cell*, 4th edition. New York: Garland Science.

Anatomie des plantes [WWW Document], n.d. URL https://bibdigital.rjb.csic.es/records/item/13576-redirection [Accessed 12.19.22].

Andersson, R., **Sandelin, A.** & **Danko, C.G.** (2015) A unified architecture of transcriptional regulatory elements. *Trends in Genetics*, **31**, 426–433. Available from: https://doi.org/10.1016/j.tig.2015.05.007

Azodi, C.B., **Lloyd, J.P.** & **Shiu, S.-H.** (2020) The cis-regulatory codes of response to combined heat and drought stress in *Arabidopsis thaliana*. *NAR Genomics Bioinforma*, **2**, lqaa049. Available from: https://doi.org/10.1093/nargab/lqaa049

Bajic, M., **Maher, K.A.** & **Deal, R.B.** (2018) Identification of open chromatin regions in plant genomes using ATAC-seq. In: Bemer, M. & Baroux, C. (Eds.) *Plant chromatin dynamics*. New York, New York, NY: Springer, pp. 183–201. Available from: https://doi.org/10.1007/978-1-4939-7318-7_12

Becraft, P.W. & **Freeling, M.** (1991) Sectors of liguleless-1 tissue interrupt an inductive signal during maize leaf development. *Plant Cell*, **3**, 801–807. Available from: https://doi.org/10.1105/tpc.3.8.801

Benfey, P.N., **Linstead, P.J.**, **Roberts, K.**, **Schiefelbein, J.W.**, **Hauser, M.-T.** & **Aeschbacher, R.A.** (1993) Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development*, **119**, 57–70. Available from: https://doi.org/10.1242/dev.119.1.57

Bezrutczyk, M., **Hartwig, T.**, **Horschman, M.**, **Char, S.N.**, **Yang, J.**, **Yang, B.** ***et al.*** (2018) Impaired phloem loading in zmsweet13a,b,c sucrose transporter triple knock-out mutants in *Zea mays*. *The New Phytologist*, **218**, 594–603. Available from: https://doi.org/10.1111/nph.15021

Bezrutczyk, M., **Zöllner, N.R.**, **Kruse, C.P.S.**, **Hartwig, T.**, **Lautwein, T.**, **Köhrer, K.** ***et al.*** (2021) Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. *Plant Cell*, **33**, 531–547. Available from: https://doi.org/10.1093/plcell/koaa055

Birchler, J.A. & **Yang, H.** (2022) The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell*, **34**, 2466–2474. Available from: https://doi.org/10.1093/plcell/koac076

Birnbaum, K., **Jung, J.W.**, **Wang, J.Y.**, **Lambert, G.M.**, **Hirst, J.A.**, **Galbraith, D.W.** ***et al.*** (2005) Cell type–specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nature Methods*, **2**, 615–619. Available from: https://doi.org/10.1038/nmeth0805-615

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

Blondel, V.D., **Guillaume, J.-L.**, **Lambiotte, R.** & **Lefebvre, E.** (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, **2008**, P10008. Available from: https://doi.org/10.1088/1742-5468/2008/10/P10008

Boyle, A.P., **Davis, S.**, **Shulha, H.P.**, **Meltzer, P.**, **Margulies, E.H.**, **Weng, Z.** ***et al.*** (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell*, **132**, 311–322. Available from: https://doi.org/10.1016/j.cell.2007.12.014

Brady, S.M., **Sarkar, S.F.**, **Bonetta, D.** & **McCourt, P.** (2003) The ABSCISIC ACID INSENSITIVE 3 (ABI3) gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in Arabidopsis. *The Plant Journal*, **34**, 67–75. Available from: https://doi.org/10.1046/j.1365-313X.2003.01707.x

Buenrostro, J.D., **Wu, B.**, **Chang, H.Y.** & **Greenleaf, W.J.** (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide: ATAC-seq for assaying chromatin accessibility. In: Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. et al. (Eds.) *Current protocols in molecular biology*, Vol. **109**. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 21.29.1–21.29.9. Available from: https://doi.org/10.1002/0471142727.mb2129s109

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

Burch-Smith, T.M. & **Zambryski, P.C.** (2012) Plasmodesmata paradigm shift: regulation from without versus within. *Annual Review of Plant Biology*, **63**, 239–260. Available from: https://doi.org/10.1146/annurev-arplant-042811-105453

Chari, T., **Banerjee, J.** & **Pachter, L.** (2021) The specious art of single-cell genomics (preprint). *Genomics*. Available from: https://doi.org/10.1101/2021.08.25.457696

Chen, H., **Lareau, C.**, **Andreani, T.**, **Vinyard, M.E.**, **Garcia, S.P.**, **Clement, K.** ***et al.*** (2019) Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biology*, **20**, 241. Available from: https://doi.org/10.1186/s13059-019-1854-5

Chen, L., **Wu, Z.** & **Hou, S.** (2020) SPEECHLESS speaks loudly in stomatal development. *Frontiers in Plant Science*, **11**, 114.

Chen, X., **Miragaia, R.J.**, **Natarajan, K.N.** & **Teichmann, S.A.** (2018) A rapid and robust method for single cell chromatin accessibility profiling. *Nature Communications*, **9**, 5345. Available from: https://doi.org/10.1038/s41467-018-07771-0

Cleves, P.A., **Ellis, N.A.**, **Jimenez, M.T.**, **Nunez, S.M.**, **Schluter, D.**, **Kingsley, D.M.** ***et al.*** (2014) Evolved tooth gain in sticklebacks is associated with a *cis* -regulatory allele of *Bmp6*. *Proceedings of the National Academy of Sciences*, **111**, 13912–13917. Available from: https://doi.org/10.1073/pnas.1407567111

Cramer, P. (2019) Organization and regulation of gene transcription. *Nature*, **573**, 45–54. Available from: https://doi.org/10.1038/s41586-019-1517-4

Cusanovich, D.A., **Daza, R.**, **Adey, A.**, **Pliner, H.A.**, **Christiansen, L.**, **Gunderson, K.L.** ***et al.*** (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, **348**, 910–914. Available from: https://doi.org/10.1126/science.aab1601

Cusanovich, D.A., **Hill, A.J.**, **Aghamirzaie, D.**, **Daza, R.M.**, **Pliner, H.A.**, **Berletch, J.B.** ***et al.*** (2018) A single-cell atlas of In vivo mammalian chromatin accessibility. *Cell*, **174**, 1309–1324.e18. Available from: https://doi.org/10.1016/j.cell.2018.06.052

Cusanovich, D.A., **Reddington, J.P.**, **Garfield, D.A.**, **Daza, R.**, **Aghamirzaie, D.**, **Marco-Ferreres, R.** ***et al.*** (2018) The cis-regulatory dynamics of embryonic development at single cell resolution. *Nature*, **555**, 538–542. Available from: https://doi.org/10.1038/nature25981

Das, A.C., **Foroutan, A.**, **Qian, B.**, **Hosseini Naghavi, N.**, **Shabani, K.** & **Shooshtari, P.** (2022) Single-cell chromatin accessibility data combined with GWAS improves detection of relevant cell types in 59 complex phenotypes. *International Journal of Molecular Sciences*, **23**, 11456. Available from: https://doi.org/10.3390/ijms231911456

Dolan, L. & **Poethig, R.S.** (1998) Clonal analysis of leaf development in cotton. *American Journal of Botany*, **85**, 315–321. Available from: https://doi.org/10.2307/2446322

Domcke, S., **Hill, A.J.**, **Daza, R.M.**, **Cao, J.**, **O'Day, D.R.**, **Pliner, H.A.** ***et al.*** (2020) A human cell atlas of fetal chromatin accessibility. *Science*, **370**, eaba7612. Available from: https://doi.org/10.1126/science.aba7612

Dorrity, M.W., **Alexandre, C.M.**, **Hamm, M.O.**, **Vigil, A.-L.**, **Fields, S.**, **Queitsch, C.** ***et al.*** (2021) The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nature Communications*. **12**, 3334. Available from: https://doi.org/10.1038/s41467-021-23675-y

Eagen, K.P., **Hartl, T.A.** & **Kornberg, R.D.** (2015) Stable chromosome condensation revealed by chromosome conformation capture. *Cell*, **163**, 934–946. Available from: https://doi.org/10.1016/j.cell.2015.10.026

Efroni, I. (2018) A conceptual framework for cell identity transitions in plants. *Plant & Cell Physiology*, **59**, 696–706. Available from: https://doi.org/10.1093/pcp/pcx172

Esau, K. (1939) Development and structure of the phloem tissue. *The Botanical Review*, **5**, 373–432.

Esau, K. (1943) Ontogeny of the vascular bundle in *Zea Mays*. *Hilgardia*, **15**, 325–368.

Esau, K. (1954) Primary vascular differentiation in plants. *Biological Reviews*, **29**, 46–86. Available from: https://doi.org/10.1111/j.1469-185X.1954.tb01397.x

Evert, R.F., **Esau, K.** & **Esau, K.** (2006) *Esau's Plant anatomy: meristems, cells, and tissues of the plant body: their structure, function, and development*, 3rd edition. Hoboken, N.J: Wiley-Interscience.

Farmer, A., **Thibivilliers, S.**, **Ryu, K.H.**, **Schiefelbein, J.** & **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. *Molecular Plant*, **14**, 372–383. Available from: https://doi.org/10.1016/j.molp.2021.01.001

Gibson, G. (2022) Perspectives on rigor and reproducibility in single cell genomics. *PLoS Genetics*, **18**, e1010210. Available from: https://doi.org/10.1371/journal.pgen.1010210

Giresi, P.G., **Kim, J.**, **McDaniell, R.M.**, **Iyer, V.R.** & **Lieb, J.D.** (2007) FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Research*, **17**, 877–885. Available from: https://doi.org/10.1101/gr.5533506

Gisselbrecht, S.S., **Palagi, A.**, **Kurland, J.V.**, **Rogers, J.M.**, **Ozadam, H.**, **Zhan, Y.** ***et al.*** (2020) Transcriptional silencers in drosophila serve a dual role as transcriptional enhancers in alternate cellular contexts. *Molecular Cell*, **77**, 324–337.e8. Available from: https://doi.org/10.1016/j.molcel.2019.10.004

Gottwald, J.R., **Krysan, P.J.**, **Young, J.C.**, **Evert, R.F.** & **Sussman, M.R.** (2000) Genetic evidence for the in planta role of phloem-specific plasma membrane sucrose transporters. *Proceedings of the National Academy of Sciences*, **97**, 13979–13984. Available from: https://doi.org/10.1073/pnas.250473797

Hake, S. & **Freeling, M.** (1986) Analysis of genetic mosaics shows that the extra epidermal cell divisions in Knotted mutant maize plants are induced by adjacent mesophyll cells. *Nature*, **320**, 621–623. Available from: https://doi.org/10.1038/320621a0

Han, J., **Wang, P.**, **Wang, Q.**, **Lin, Q.**, **Chen, Z.**, **Yu, G.** ***et al.*** (2020) Genome-wide characterization of DNase I-hypersensitive sites and cold response regulatory landscapes in grasses. *The Plant Cell*, **32**, 2457–2473. Available from: https://doi.org/10.1105/tpc.19.00716

Helariutta, Y., **Fukaki, H.**, **Wysocka-Diller, J.**, **Nakajima, K.**, **Jung, J.**, **Sena, G.** ***et al.*** (2000) The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. *Cell*, **101**, 555–567. Available from: https://doi.org/10.1016/S0092-8674(00)80865-X

Herman, P.L. & **Marks, M.D.** (1989) Trichome development in *Arabidopsis thaliana*. II. Isolation and complementation of the GLABROUS1 gene.

Plant Cell, **1**, 1051–1055. Available from: https://doi.org/10.1105/tpc.1.11.1051

Hill, M.S., **Vande Zande, P.** & **Wittkopp, P.J.** (2020) Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics*, **1–13**, 203–215. Available from: https://doi.org/10.1038/s41576-020-00304-w

Hodge, R.D., **Bakken, T.E.**, **Miller, J.A.**, **Smith, K.A.**, **Barkan, E.R.**, **Graybuck, L.T.** ***et al.*** (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature*, **573**, 61–68. Available from: https://doi.org/10.1038/s41586-019-1506-7

Hooke, R. (1665) *Micrographia: or, some physiological descriptions of minute bodies made by magnifying glasses. With observations and inquiries thereupon*. London: Jo. Martyn, and Ja. Allestry, pp. 1635–1703.

Hughes, T.E. & **Langdale, J.A.** (2022) SCARECROW is deployed in distinct contexts during rice and maize leaf development. *Development*, **149**, dev200410. Available from: https://doi.org/10.1242/dev.200410

Hughes, T.E., **Langdale, J.A.** & **Kelly, S.** (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Research*, **24**, 1348–1355. Available from: https://doi.org/10.1101/gr.172684.114

Hung, C.-Y., **Lin, Y.**, **Zhang, M.**, **Pollock, S.**, **David Marks, M.** & **Schiefelbein, J.** (1998) A common position-dependent mechanism controls cell-type patterning and GLABRA2 regulation in the root and hypocotyl epidermis of Arabidopsis. *Plant Physiology*, **117**, 73–84.

Imperatorskaia͡ akademiia͡ nauk (Russia), nauk (Russia), I. akademiia͡, nauk (Russia), I. akademiia͡. (1868) *Mémoires de l'Académie impériale des sciences de St.-Pétersbourg*. St.-Petersburg: L'Académie.

Jaitin, D.A., **Kenigsberg, E.**, **Keren-Shaul, H.**, **Elefant, N.**, **Paul, F.**, **Zaretsky, I.** ***et al.*** (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science*, **343**, 776–779. Available from: https://doi.org/10.1126/science.1247651

Jegla, D.E. & **Sussex, I.M.** (1989) Cell lineage patterns in the shoot meristem of the sunflower embryo in the dry seed. *Developmental Biology*, **131**, 215–225. Available from: https://doi.org/10.1016/S0012-1606(89)80053-3

Johnson, S.M., **Tan, F.J.**, **McCullough, H.L.**, **Riordan, D.P.** & **Fire, A.Z.** (2006) Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. *Genome Research*, **16**, 1505–1516. Available from: https://doi.org/10.1101/gr.5560806

Kaplan, D.R. & **Hagemann, W.** (1991) The relationship of cell and organism in vascular plants. *Bioscience*, **41**, 693–703. Available from: https://doi.org/10.2307/1311764

Kirik, V., **Schnittger, A.**, **Radchuk, V.**, **Adler, K.**, **Hülskamp, M.** & **Bäumlein, H.** (2001) Ectopic expression of the Arabidopsis AtMYB23 gene induces differentiation of trichome cells. *Developmental Biology*, **235**, 366–377. Available from: https://doi.org/10.1006/dbio.2001.0287

Lai, X., **Stigliani, A.**, **Vachon, G.**, **Carles, C.**, **Smaczniak, C.**, **Zubieta, C.** ***et al.*** (2019) Building transcription factor binding site models to understand gene regulation in plants. *Molecular Plant*, **12**, 743–763. Available from: https://doi.org/10.1016/j.molp.2018.10.010

Langdale, J.A., **Lane, B.**, **Freeling, M.** & **Nelson, T.** (1989) Cell lineage analysis of maize bundle sheath and mesophyll cells. *Developmental Biology*, **133**, 128–139. Available from: https://doi.org/10.1016/0012-1606(89)90304-7

Larkin, J.C., **Oppenheimer, D.G.**, **Lloyd, A.M.**, **Paparozzi, E.T.** & **Marks, M.D.** (1994) Roles of the GLABROUS1 and TRANSPARENT TESTA GLABRA genes in Arabidopsis trichome development. *Plant Cell*, **6**, 1065–1076. Available from: https://doi.org/10.1105/tpc.6.8.1065

Lee, S.-I. & **Kim, N.-S.** (2014) Transposable elements and genome size variations in plants. *Genomics & Informatics*, **12**, 87–97. Available from: https://doi.org/10.5808/GI.2014.12.3.87

Lee, T.A., **Nobori, T.**, **Illouz-Eliaz, N.**, **Xu, J.**, **Jow, B.**, **Nery, J.R.** ***et al.*** (2023) A single-nucleus atlas of seed-to-seed development in Arabidopsis (preprint). *Plant Biology*. Available from: https://doi.org/10.1101/2023.03.23.533992

Leroux, O. (2012) Collenchyma: a versatile mechanical tissue with dynamic cell walls. *Annals of Botany*, **110**, 1083–1098. Available from: https://doi.org/10.1093/aob/mcs186

Li, P., **Ponnala, L.**, **Gandotra, N.**, **Wang, L.**, **Si, Y.**, **Tausta, S.L.** ***et al.*** (2010) The developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, **42**, 1060–1067. Available from: https://doi.org/10.1038/ng.703

Lieberman-Aiden, E., **van Berkum, N.L.**, **Williams, L.**, **Imakaev, M.**, **Ragoczy, T.**, **Telling, A.** ***et al.*** (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293. Available from: https://doi.org/10.1126/science.1181369

Lu, Z., **Hofmeister, B.T.**, **Vollmers, C.**, **DuBois, R.M.** & **Schmitz, R.J.** (2017) Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Research*, **45**, e41. Available from: https://doi.org/10.1093/nar/gkw1179

Lu, Z., **Marand, A.P.**, **Ricci, W.A.**, **Ethridge, C.L.**, **Zhang, X.** & **Schmitz, R.J.** (2019) The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nature Plants*, **5**, 1250–1259. Available from: https://doi.org/10.1038/s41477-019-0548-z

Maaten, L.v.d. & **Hinton, G.** (2008) Visualizing data using t-SNE. *Journal of Machine Learning Research*, **9**, 2579–2605.

Maher, K.A., **Bajic, M.**, **Kajala, K.**, **Reynoso, M.**, **Pauluzzi, G.**, **West, D.A.** ***et al.*** (2018) Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell*, **30**, 15–36. Available from: https://doi.org/10.1105/tpc.17.00581

Marand, A.P., **Chen, Z.**, **Gallavotti, A.** & **Schmitz, R.J.** (2021) A cis-regulatory atlas in maize at single-cell resolution. *Cell*, **184**, 3041–3055.e21. Available from: https://doi.org/10.1016/j.cell.2021.04.014

McInnes, L., **Healy, J.** & **Melville, J.** (2018) UMAP: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Mifsud, B., **Tavares-Cadete, F.**, **Young, A.N.**, **Sugar, R.**, **Schoenfelder, S.**, **Ferreira, L.** ***et al.*** (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, **47**, 598–606. Available from: https://doi.org/10.1038/ng.3286

Moore, B.M., **Lee, Y.S.**, **Wang, P.**, **Azodi, C.**, **Grotewold, E.** & **Shiu, S.-H.** (2022) Modeling temporal and hormonal regulation of plant transcriptional response to wounding. *Plant Cell*, **34**, 867–888. Available from: https://doi.org/10.1093/plcell/koab287

Nelms, B. & **Walbot, V.** (2019) Defining the developmental program leading to meiosis in maize. *Science*, **364**, 52–56. Available from: https://doi.org/10.1126/science.aav6428

Oka, R., **Zicola, J.**, **Weber, B.**, **Anderson, S.N.**, **Hodgman, C.**, **Gent, J.I.** ***et al.*** (2017) Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biology*, **18**, 137. Available from: https://doi.org/10.1186/s13059-017-1273-4

Oppenheimer, D.G., **Herman, P.L.**, **Sivakumaran, S.**, **Esch, J.** & **Marks, M.D.** (1991) A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. *Cell*, **67**, 483–493. Available from: https://doi.org/10.1016/0092-8674(91)90523-2

Panchy, N., **Lehti-Shiu, M.** & **Shiu, S.-H.** (2016) Evolution of gene duplication in plants. *Plant Physiology*, **171**, 2294–2316. Available from: https://doi.org/10.1104/pp.16.00523

Pang, B. & **Snyder, M.P.** (2020) Systematic identification of silencers in human cells. *Nature Genetics*, **52**, 254–263. Available from: https://doi.org/10.1038/s41588-020-0578-5

Paterson, A.H., **Freeling, M.**, **Tang, H.** & **Wang, X.** (2010) Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*, **61**, 349–372. Available from: https://doi.org/10.1146/annurev-arplant-042809-112235

Pick, T.R., **Bräutigam, A.**, **Schlüter, U.**, **Denton, A.K.**, **Colmsee, C.**, **Scholz, U.** ***et al.*** (2011) Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. *Plant Cell*, **23**, 4208–4220. Available from: https://doi.org/10.1105/tpc.111.090324

Pliner, H.A., **Packer, J.S.**, **McFaline-Figueroa, J.L.**, **Cusanovich, D.A.**, **Daza, R.M.**, **Aghamirzaie, D.** ***et al.*** (2018) Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular Cell*, **71**, 858–871.e8. Available from: https://doi.org/10.1016/j.molcel.2018.06.044

Poethig, R.S. & **Sussex, I.M.** (1985a) The developmental morphology and growth dynamics of the tobacco leaf. *Planta*, **165**, 158–169. Available from: https://doi.org/10.1007/BF00395038

Poethig, R.S. & **Sussex, I.M.** (1985b) The cellular parameters of leaf development in tobacco: a clonal analysis. *Planta*, **165**, 170–184.

Raju, S.K.K. (2020) Comparative profiling examines roles of DNA regulatory sequences and accessible chromatin during cold stress response in grasses. *The Plant Cell*, **32**, 2451–2452. Available from: https://doi.org/10.1105/tpc.20.00471

Reinhardt, D., **Frenz, M.**, **Mandel, T.** & **Kuhlemeier, C.** (2003) Microsurgical and laser ablation analysis of interactions between the zones and layers of the tomato shoot apical meristem. *Development*, **130**, 4073–4083. Available from: https://doi.org/10.1242/dev.00596

Reynoso, M.A., **Kajala, K.**, **Bajic, M.**, **West, D.A.**, **Pauluzzi, G.**, **Yao, A.I.** ***et al.*** (2019) Evolutionary flexibility in flooding response circuitry in angiosperms. *Science*, **365**, 1291–1295. Available from: https://doi.org/10.1126/science.aax8862

Ricci, W.A., **Lu, Z.**, **Ji, L.**, **Marand, A.P.**, **Ethridge, C.L.**, **Murphy, N.G.** ***et al.*** (2019) Widespread long-range cis -regulatory elements in the maize genome. *Nature Plants*, **5**, 1237–1249. Available from: https://doi.org/10.1038/s41477-019-0547-0

Rosas, U., **Mei, Y.**, **Xie, Q.**, **Banta, J.A.**, **Zhou, R.W.**, **Seufferheld, G.** ***et al.*** (2014) Variation in Arabidopsis flowering time associated with cis-regulatory variation in CONSTANS. *Nature Communications*, **5**, 3651. Available from: https://doi.org/10.1038/ncomms4651

Ryu, K.H., **Huang, L.**, **Kang, H.M.** & **Schiefelbein, J.** (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiology*, **179**, 1444–1456. Available from: https://doi.org/10.1104/pp.18.01482

Schmitz, R.J., **Grotewold, E.** & **Stam, M.** (2022) Cis-regulatory sequences in plants: their importance, discovery, and future challenges. *Plant Cell*, **34**, 718–741. Available from: https://doi.org/10.1093/plcell/koab281

Schuler, M.L., **Sedelnikova, O.V.**, **Walker, B.J.**, **Westhoff, P.** & **Langdale, J.A.** (2018) SHORTROOT-mediated increase in stomatal density has No impact on photosynthetic efficiency. *Plant Physiology*, **176**, 757–772. Available from: https://doi.org/10.1104/pp.17.01005

Sharman, B.C. (1942) Developmental anatomy of the shoot of *Zea mays* L. *Annals of Botany*, **6**, 245–282. Available from: https://doi.org/10.1093/oxfordjournals.aob.a088407

Shlyueva, D., **Stampfel, G.** & **Stark, A.** (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews. Genetics*, **15**, 272–286. Available from: https://doi.org/10.1038/nrg3682

Shulse, C.N., **Cole, B.J.**, **Ciobanu, D.**, **Lin, J.**, **Yoshinaga, Y.**, **Gouran, M.** ***et al.*** (2019) High-throughput single-cell transcriptome profiling of plant cell types. *Cell Reports*, **27**, 2241–2247.e4. Available from: https://doi.org/10.1016/j.celrep.2019.04.054

Sinha, N. & **Hake, S.** (1990) Mutant characters of knotted maize leaves are determined in the innermost tissue layers. *Developmental Biology*, **141**, 203–210. Available from: https://doi.org/10.1016/0012-1606(90)90115-y

Stadler, R. & **Sauer, N.** (1996) The *Arabidopsis thaliana* AtSUC2 gene is specifically expressed in companion cells. *Botanica Acta: Journal of the German Botanical Society*, **109**, 299–306. Available from: https://doi.org/10.1111/j.1438-8677.1996.tb00577.x

Stadler, R., **Wright, K.M.**, **Lauterbach, C.**, **Amon, G.**, **Gahrtz, M.**, **Feuerstein, A.** ***et al.*** (2005) Expression of GFP-fusions in Arabidopsis companion cells reveals non-specific protein trafficking into sieve elements and identifies a novel post-phloem domain in roots. *The Plant Journal*, **41**, 319–331. Available from: https://doi.org/10.1111/j.1365-313X.2004.02298.x

Strasburger, E. (1888) *Histologische Beiträge*. Jena: G. Fischer. Available from: https://doi.org/10.5962/bhl.title.24451

Studer, A., **Zhao, Q.**, **Ross-Ibarra, J.** & **Doebley, J.** (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*, **43**, 1160–1163. Available from: https://doi.org/10.1038/ng.942

Toufighi, K., **Brady, S.M.**, **Austin, R.**, **Ly, E.** & **Provart, N.J.** (2005) The Botany Array resource: e-Northerns, expression angling, and promoter analyses. *The Plant Journal*, **43**, 153–163. Available from: https://doi.org/10.1111/j.1365-313X.2005.02437.x

Trapnell, C., **Cacchiarelli, D.**, **Grimsby, J.**, **Pokharel, P.**, **Li, S.**, **Morse, M.** ***et al.*** (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, **32**, 381–386. Available from: https://doi.org/10.1038/nbt.2859

van der Burg, K.R.L., **Lewis, J.J.**, **Brack, B.J.**, **Fandino, R.A.**, **Mazo-Vargas, A.** & **Reed, R.D.** (2020) Genomic architecture of a genetically assimilated seasonal color pattern. *Science*, **370**, 721–725. Available from: https://doi.org/10.1126/science.aaz3017

Waltman, L. & **van Eck, N.J.** (2013) A smart local moving algorithm for large-scale modularity-based community detection. *The European physical journal B*, **86**, 471. Available from: https://doi.org/10.1140/epjb/e2013-40829-0

Wang, P., **Kelly, S.**, **Fouracre, J.P.** & **Langdale, J.A.** (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *The Plant Journal*, **75**, 656–670. Available from: https://doi.org/10.1111/tpj.12229

Wilhelm, K. (1880) *Beiträge zur Kenntniss des Siebröhrenapparates dicotyler Pflanzen*. Germany: Verlag von Wilhelm Engelmann. Available from: https://www.google.com/books/edition/Beitr%C3%A4ge_zur_Kenntniss_des_Siebr%C3%B6hrena/YbAXAAAAYAAJ?hl=en&gbpv=0 [Accessed 8th February 2023].

Wolock, S.L., **Lopez, R.** & **Klein, A.M.** (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Systems*, **8**, 281–291.e9. Available from: https://doi.org/10.1016/j.cels.2018.11.005

Wucherpfennig, J.I., **Howes, T.R.**, **Au, J.N.**, **Au, E.H.**, **Roberts Kingman, G.A.**, **Brady, S.D.** ***et al.*** (2022) Evolution of stickleback spines through independent cis-regulatory changes at HOXDB. *Nature Ecology & Evolution*, **6**, 1537–1552. Available from: https://doi.org/10.1038/s41559-022-01855-3

Yan, F., **Powell, D.R.**, **Curtis, D.J.** & **Wong, N.C.** (2020) From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*, **21**, 22. Available from: https://doi.org/10.1186/s13059-020-1929-3

Zeng, H. (2022) What is a cell type and how to define it? *Cell*, **185**, 2739–2755. Available from: https://doi.org/10.1016/j.cell.2022.06.031

Zeng, Z., **Zhang, W.**, **Marand, A.P.**, **Zhu, B.**, **Buell, C.R.** & **Jiang, J.** (2019) Cold stress induces enhanced chromatin accessibility and bivalent histone modifications H3K4me3 and H3K27me3 of active genes in potato. *Genome Biology*, **20**, 123. Available from: https://doi.org/10.1186/s13059-019-1731-2

Zhang, T.-Q., **Chen, Y.**, **Liu, Y.**, **Lin, W.-H.** & **Wang, J.-W.** (2021) Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nature Communications*, **12**, 2053. Available from: https://doi.org/10.1038/s41467-021-22352-4

Zhang, T.-Q., **Xu, Z.-G.**, **Shang, G.-D.** & **Wang, J.-W.** (2019) A single-cell RNA sequencing profiles the developmental landscape of Arabidopsis root. *Molecular Plant*, **12**, 648–660. Available from: https://doi.org/10.1016/j.molp.2019.04.004

Zhao, H., **Zhang, W.**, **Chen, L.**, **Wang, L.**, **Marand, A.P.**, **Wu, Y.** ***et al.*** (2018) Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiology*, **176**, 2789–2803. Available from: https://doi.org/10.1104/pp.17.01467

Zhou, P., **Enders, T.A.**, **Myers, Z.A.**, **Magnusson, E.**, **Crisp, P.A.**, **Noshay, J.M.** ***et al.*** (2022) Prediction of conserved and variable heat and cold stress response in maize using cis-regulatory information. *Plant Cell*, **34**, 514–534. Available from: https://doi.org/10.1093/plcell/koab267

Zou, C., **Sun, K.**, **Mackaluso, J.D.**, **Seddon, A.E.**, **Jin, R.**, **Thomashow, M.F.** ***et al.*** (2011) Cis-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 14992–14997. Available from: https://doi.org/10.1073/pnas.1103202108