

Anno genominis XX: 20 years of Arabidopsis genomics

Nicholas J. Provart ^{1,*†} Siobhan M. Brady ² Geraint Parry ³ Robert J. Schmitz ⁴
 Christine Queitsch ^{5,6} Dario Bonetta ⁷ Jamie Waese ¹ Korbinian Schneeberger ^{8,9} and
 Ann E. Loraine ^{10,*†}

- 1 Department of Cell & Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, M5S 3B2, Canada
- 2 Department of Plant Biology and Genome Center, University of California, Davis, California, 95616, USA
- 3 GARNet, School of Biosciences, Cardiff University, Cardiff, CF10 3AX, UK
- 4 Department of Genetics, University of Georgia, Georgia, 30602, USA
- 5 Department of Genome Sciences, School of Medicine, University of Washington, Seattle, Washington, 98195, USA
- 6 Brotman Baty Institute for Precision Medicine, Seattle, Washington, 98195, USA
- 7 Faculty of Science, Ontario Tech University, Oshawa, Ontario, L1G 0C5, Canada
- 8 Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, D-50829, Cologne, Germany
- 9 Faculty of Biology, LMU Munich, 82152 Munich, Germany
- 10 Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, 28223, USA

*Author for correspondence: nicholas.provart@utoronto.ca (N.J.P.), ann.loraine@uncc.edu (A.E.L.)

†Senior authors

N.J.P. outlined the review with input from all authors and *The Plant Cell*. N.J.P. and G.P. wrote the Introduction, and G.P. conducted the MASC survey for this review. S.M.B. covered single-cell genomics, while R.J.S. and C.Q. reviewed epigenomics articles. D.B. covered forward genomics and mapping by sequencing. J.W., N.J.P., and A.E.L. wrote the Visualizing the Future section. K.S. summarized articles covering 1001 genomes. N.J.P. prepared the figures and assembled the manuscript. All authors helped to edit the manuscript.

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instruction for Authors (www.plantcell.org) are: Nicholas J. Provart (nicholas.provart@utoronto.ca) and Ann E. Loraine (ann.loraine@uncc.edu).

Abstract

Twenty years ago, the *Arabidopsis thaliana* genome sequence was published. This was an important moment as it was the first sequenced plant genome and explicitly brought plant science into the genomics era. At the time, this was not only an outstanding technological achievement, but it was characterized by a superb global collaboration. The Arabidopsis genome was the seed for plant genomic research. Here, we review the development of numerous resources based on the genome that have enabled discoveries across plant species, which has enhanced our understanding of how plants function and interact with their environments.

Introduction

The Internet was just a small network called NSFNET, the Sony Walkman cassette player was the choice of music delivery for teenagers and lab workers, and it was possible to read around 1,000 nucleotides of DNA sequence a day after running radioactively labeled Sanger sequencing reactions on a vertical polyacrylamide gel and waiting a couple of

days for a film to be exposed. No doubt inspired by the nascent effort to sequence the human genome (workshops for which were held in 1985 in Santa Fe, New Mexico, and Santa Cruz, California), earlier adopters of *Arabidopsis thaliana* (Arabidopsis) as a research organism (Provart et al., 2016) met at four workshops sponsored by the US National Science Foundation (NSF) in 1989 and early 1990. This led

to the release of a draft document in June 1990 at that year's International Conference on Arabidopsis Research in Vienna called "A Long-Range Plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project" (http://arabidopsisresearch.org/images/publications/mascreports/1990_MASCPlan.pdf).

This project was developed "based on the recognition that a profound understanding of plant biology is essential in order to meet the immediate and future challenges facing world agriculture and the global environment." The mission statement of this project was "to identify all of the genes by using a functional biological approach leading to the determination of the complete sequence of the Arabidopsis genome by the end of this [i.e. the 20th] century." At the outset, the authors of the report recognized the importance of international coordination for "rapid and efficient advances" in Arabidopsis genome research. Furthermore, they foresaw the need for biological resources centers (at least two, "advisable for security reasons, if nothing else") and an informatics program to enable the sharing of data. Somerville and Koornneef (2002) describe the process by which the Arabidopsis Genome Initiative came together in the 1990s to sequence the genome—lots of cooperation and coordination were involved!

In this review, we cover the initial efforts to sequence the genome, and how the genome has been updated over time. We discuss how the genome became a platform for many other omics-based approaches, from early functional genomics efforts to document the expression pattern of all Arabidopsis genes and create T-DNA knockout collections, to more recent interactomic, epigenomic, and single-cell RNA-seq-based approaches. We emphasize how international cooperation has led to breakthroughs in our understanding of plant biology, and we finish by looking to the future in which the Arabidopsis genome will continue to play an important role.

Creating a platform for genomics resources from raw nucleotide sequences

From BACs and TACs to Araport11

The sequences of chromosomes 2 and 4 from Arabidopsis were published in December 1999 (Lin et al., 1999; Mayer et al., 1999), a year before an article presenting the sequences of the final three chromosomes and an overall description of genome analysis was published in *Nature* in December 2000 (Arabidopsis Genome Initiative, 2000). These manuscripts described an elaborate strategy consisting of physical mapping via fingerprinting (either by restriction fragment analysis, hybridization, or PCR), integration with genetic maps, and end sequence analysis of 47,788 BAC (bacterial artificial chromosome) clones to create 10 contigs covering the chromosome arms and centromeric heterochromatin (assembled from 1,569 BAC, TAC (transformation-competent artificial chromosome), cosmid, and P1 (bacteriophage P1 clones). Telomeric sequences were assembled from separate yeast artificial chromosome (YAC) and

phage clones. The Arabidopsis Genome Initiative article reported 25,498 genes and ancient whole genome duplications. In the two decades since then, there have been 11 genome annotation revisions, from J. Craig Venter's The Institute for Genomic Research iteration 1 released in August of 2001 through The Arabidopsis Information Resource's 6th update (TAIR6) in November 2005, to the current Araport11 version containing 27,655 genes in June 2016 (Cheng et al., 2017). The history of the 11 genome annotation versions may be explored online at https://www.arabidopsis.org/portals/genAnnotation/genome_snapshot.jsp. The original article is one of the most impactful articles in plant research and has been cited 5,960 times in the scientific literature across a broad range of disciplines and by researchers around the world (see Figure 1).

The Arabidopsis genome as a seed for plant genomic research

Over the past 30 years, the Multinational Arabidopsis Steering Committee (MASC) has advised on the activities of the Arabidopsis community, documenting its continuing evolution via an annual progress report (Parry et al., 2020) and facilitating community integration through oversight of the annual International Conference on Arabidopsis Research (<http://arabidopsisresearch.org>). In the lead up to the preparation of this manuscript, MASC circulated a survey to understand what the sequencing of the Arabidopsis genome has meant to members of the global community.

Publication of the Arabidopsis genomic sequence sent a message to the wider scientific community about the legitimacy of using plant experimental systems as a tool for fundamental discoveries in cell and molecular biology. Professor Klaus Mayer at the Helmholtz Center in Munich was part of the sequencing team and recalls, "The experience of a truly visionary and ambitious international collaboration and approach the new challenge in plant research changed my life and scientific career forever. The size and complexity of this project (remember no next generation sequencing [NGS] at that time!) as well as coordination over US, Europe and Japan was a pioneering challenge but opened lots of opportunities, funding-wise, science-wise etc. Probably the alliance of some remarkable personalities and talents, organizations, infrastructures (sequencing and computing) and the overall spirit of pioneering a game changer in plant research was the basis for the remarkable success."

Lessons learned from assembly and analysis of the Arabidopsis genome was a prelude for the use of genomic techniques in plants with more complex genomes. As Mayer summarizes "...the experiences in Arabidopsis were extremely important and [were crucial in] developing strategies to approach the size- and complexity-wise holy grails in plant genomics (e.g. maize & Triticeae)." This is corroborated by Dr. Miriam Gifford at the University of Warwick: "[the Arabidopsis Genome Initiative]...pushed ahead the sequencing of other plant genomes, set a standard for genome publication and access, and highlighted the depth and



Figure 1 Breakdown of the top 25 Web of Science categories (top panel) and top 25 countries (bottom panel) for the 5,960 articles that have cited the original Arabidopsis genome article (AGI, 2000) as of August 1, 2020, illustrating the impact that the genome sequence of Arabidopsis has had across scientific disciplines and around the world.

complexity of even a simple plant genome to researchers studying non-plant organisms.”

The multinational coordinated Arabidopsis functional genomics project and AtGenExpress

The Multinational Coordinated Arabidopsis Functional Genomics Project was an idea that developed from a workshop funded by the NSF that was held in early 2000 at the Salk Institute in California entitled “Functional Genomics and the Virtual Plant: A blueprint for understanding how plants are built and how to improve them” (text available online at <https://www.arabidopsis.org/portals/masc/workshop2010.jsp>). This workshop proposed “to determine the function of all Arabidopsis genes during the next decade, using a systems approach” and was intended to build on the success of the genome sequence. More than 60 projects were funded under the “Arabidopsis 2010” designation, ranging from studies to elucidate the ionome and translational regulons, to many others using mutational and over-expression approaches on subsets of Arabidopsis genes to understand their functions. One of the most useful aspects of the 2010 Project was the generation of expression profiles for (almost) all Arabidopsis genes.

The idea of developing AtGenExpress, an effort to document the expression patterns of all Arabidopsis genes using high-throughput expression profiling technologies, first appeared in the 2003 MASC report (http://arabidopsisresearch.org/images/publications/mascreports/2003_MASCReport.pdf) and was in fact hinted at in the 2000 publication of the Arabidopsis genome: “at a biochemical level, the specificity conferred by nucleotide sequence, and the completeness of the survey allow complex mixtures of RNA and protein to be resolved into their individual components using microarrays and mass spectrometry.”

Several funding agencies, notably the NSF in the USA, the DFG in Germany, RIKEN in Japan, and BBSRC in the United Kingdom, awarded grants to generate large expression data sets using both custom cDNA microarrays and the Affymetrix ATH1 array. The generation of a developmental atlas (Schmid et al., 2005) and an abiotic response map (Kilian et al., 2007), along with unpublished—but publicly available—transcriptomic data sets for hormone responses, chemical inhibitors, plant–pathogen interactions, and others, plus easy access provided by tools like the Arabidopsis eFP Browser (Winter et al., 2007) and Genevestigator (Zimmermann et al., 2004; Hruz et al., 2008) provided

thousands of plant researchers with instantaneous access to expression profiles of their genes of interest, saving each hundreds of hours of research time by obviating the need to perform RNA gel blot analyses in their individual laboratories. These data would help tease apart “functional redundancy” brought about by “extensive gene duplications” seen through careful analysis of the Arabidopsis genome (Arabidopsis Genome Initiative, 2000). In addition, these data sets act as databases of not only positive results but also negative results, letting researchers know where genes are not expressed. The articles describing these data sets and tools, although now superseded by RNA-seq technology (which too benefits from a high-quality reference genome), are among the most highly cited in Arabidopsis research. From the perspective of technological breakthroughs, the first microarray profiling experiment on any organism was conducted using 45 cDNAs from Arabidopsis (Schena et al., 1995—with a whopping 13,000 citations!), as was the first strand-specific RNA-seq experiment (Lister et al., 2008).

The rise of NGS, 1001 genomes, and “next” NGS

NGS: get your cheap genomes here!

The availability of a mature genome annotation (then at TAIR6) together with its small genome made Arabidopsis the optimal plant system for the rising next-generation sequencing (NGS) technologies in first decade of this century (see Figure 2 for a depiction of the types of research and technologies the genome sequence has enabled). Such technologies dramatically reduced the cost of sequencing a genome. It was no surprise that soon after initial pilot studies (Nordborg et al., 2005; Ossowski et al., 2008), an ambitious 1001 Genomes Project for Arabidopsis was launched in 2009 (Weigel and Mott, 2009). Following multiple, individual genome analyses of larger collections of different Arabidopsis populations (Cao et al., 2011; Gan et al., 2011; Long et al., 2013; Schmitz et al., 2013), the 1001 Genomes Consortium

published their integrative analysis of 1,135 Eurasian *A. thaliana* genomes in 2016 (1001 Genomes Consortium, 2016).

One of the major findings of this analysis was that the vast majority of the Eurasian accessions were derived from a recent expansion of a single clade, which spread along the Eurasian east–west axis, possibly supported by the rapid expansion of agriculture, and finally reached northern America and central Asia, where the youngest populations could be observed. This post-glacial spread into new ecosystems required adaptation to new environmental challenges, introducing geographical changes in allele frequency and allele distribution across the species range (1001 Genomes Consortium, 2016). In addition to the Eurasian clade, a few samples revealed the existence of five highly diverged “relict” lineages all located in either African or southern European regions, most prominently on the Iberian Peninsula. A more recent study on the genomic diversity in Africa revealed that Arabidopsis is actually native not only to Eurasia and North Africa but also to the Afro-alpine regions of sub-Saharan Africa and South Africa, including the greatest variation among Arabidopsis accessions identified so far (Durvasula et al., 2017).

Besides population demography, a main motivation for the 1001 Genomes Project was to build genetic resources that would enable the global community of Arabidopsis researchers to perform genome-wide association studies (GWAS). Since Arabidopsis accessions are inbred, they have stable homozygous genomes. Thus, individual accessions included in the 1001 Genomes Consortium can be distributed as seeds and readily used in association mapping without the need to sequence them over and over again. Seeds from all these accessions are available at the Arabidopsis Biological Resource Center (ABRC), and the genome sequences can be downloaded from the 1001genomes.org web portal. It should be pointed out that there are companion methylomes and transcriptomes available for these accessions, too (Kawakatsu et al., 2016).

The density of genetic markers in the 1001 Genomes data set is sufficient to find associations across the entire genome. However, with the improvement of next generation sequencing from short to long read technologies, it is becoming increasingly clear that genome resequencing misses large and complex genomic variation (Zapata et al., 2016). A recent comparison of eight de novo assembled genomes revealed a huge amount (up to 6 Mb or 5% of the genome) of nonreference sequences in each of the individual accessions’ genomes (Jiao and Schneeberger, 2020). This additional sequence introduced copy-number changes in approximately 5,000 genes, including approximately 1,900 genes that were not part of the reference annotation. Together, the assemblies revealed a pan-genome size of 135 Mb, including a total of approximately 30,000 genes across the entire population of Arabidopsis, which is likely to be an underestimation, as most of the African diversity was not included in this set of assembled accessions. In addition to sequence differences, differences in collinearity

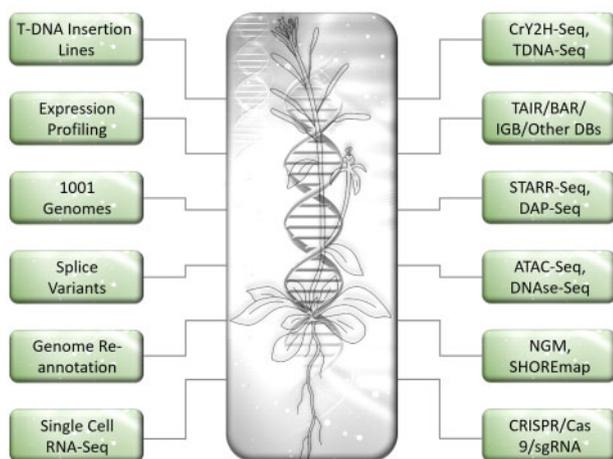


Figure 2 The Arabidopsis genome sequence as a platform. The Arabidopsis genome sequence has allowed a vast ecosystem of research areas and technologies to flourish.

could be observed between the genomes. In some genomic regions, all eight genomes were completely rearranged, forming eight distinct haplotypes. These hotspots of rearrangements were enriched for genes implicated in biotic stress responses. This suggested that such hotspots of rearrangements undergo different evolutionary dynamics, including the rapid accumulation of new mutations to generate quick responses in the interminable battle against biotic stressors.

“Next” NGS

However, genomics has not yet reached its limits. The latest revolution in genome sequencing introduced high-quality long reads (e.g. Pacific Biosciences’ HiFi reads) without the high sequencing error rate that used to be the unifying shortcoming of all long-read technologies. Assemblies based on HiFi reads outcompete the quality of the widely used reference sequence, including the reconstruction of telomere-to-telomere sequences (Miga et al., 2020). The unlimited access to genomic variation at all levels will help deepen our understanding of the molecular basis of natural variation and help unravel the genomic basis of adaptation to new environments. While we will soon have access to many reference-quality genome sequences, the data will require new bioinformatics solutions to make full use of all the new information. Thus, 20 years after the release of the reference sequence, the peak of plant genomics is yet to come.

RNA-seq and splice variants, reannotating the genome

Assembling a genomic sequence into chromosomes or pan-genomes is only the first step toward defining the genome of a species. In addition, we also need notations on the sequence that indicate the locations of genes and details about how gene sequences are copied into RNA. These details, called “annotations” and “gene models,” delineate the start and stop positions of introns, exons, and open reading frames, which define how transcription, RNA processing, and translation (in the case of protein-coding genes) occur at a locus. NGS has also enabled unprecedented understanding of such details.

In the early days, these annotations came from two sources: ab initio gene prediction programs and high-throughput cDNA sequencing projects. The cDNA sequencing projects produced expressed sequence tags (ESTs) that genome project scientists aligned onto the genome and used to define gene models. Later, assembled RNA-seq data served much the same goal, and in fact, 113 RNA-seq data sets were used to reannotate the most recent Arabidopsis genome annotation release, Araport11 (Cheng et al., 2017).

Early on, it became clear that many genes produce more than one mature transcript species due to alternative splicing, alternative promoters, and alternative polyadenylation. At first, we were limited to simply observing and documenting these variants. Now, thanks to the large number of RNA-seq data sets available in the public databases, we are beginning to quantify how often and where these transcript variants occur.

Bioinformatics analysis of ESTs (Ner-Gaon et al., 2004; English et al., 2010) discovered, and RNA-seq studies later confirmed, that the retention of introns in spliced transcripts is unusually common in Arabidopsis relative to mammals. In plants, introns are smaller and lack the polypyrimidine tracts present in introns in other species. The serine- and arginine-rich family of splicing regulator proteins is larger in plants than in other species, with several plant-specific members (Barta et al., 2010). Taken together, these observations suggest that splicing biochemistry has plant-specific features. Alternative splicing is especially common in splicing-related genes and in genes involved in regulating circadian cycling. The type B response regulator family illustrates this phenomenon. One branch of this family undergoes normal levels of alternative splicing and is involved in cytokinin signal transduction. Members of a related branch, first designated the “pseudo-response regulators,” regulate clock pathways and are highly alternatively spliced (Matsushika et al., 2000). Another key finding is that temperature changes trigger changes in splicing patterns, particularly among genes involved in splicing (Gulledge et al., 2012; Calixto et al., 2018). These observations have led to speculation that the splicing machinery remodels itself via alternative splicing in response to temperature changes and that the temperature-dependent regulation of splicing may in turn intersect with the regulation of circadian rhythms. Finally, most alternatively spliced genes appear to produce the same or similar proportions of splice variants in diverse tissues and sample types (Loraine et al., 2013). When viewed through the lens of bulk RNA-seq data, splicing patterns appear remarkably stable across diverse treatments and conditions.

RNA-seq of individual cells

The most recent transformative technology to arise in biology is single-cell sequencing. This methodology was first applied in plants in studies of the Arabidopsis root. The first version of single-cell sequencing sorted single cells into a well plate and carried out subsequent miniaturized synthesis of RNA-seq libraries (Efroni et al., 2016). These libraries were used to elucidate the developmental trajectories that individual cells undertake in root regeneration. If cut within a particular distance from the root tip, cells remaining at the cut end are able to form a re-organized stem cell niche within several days. This sequencing determined that this regeneration follows a developmental program similar to the program that occurs during embryogenesis (Efroni et al., 2016).

Several years later, DropSeq and 10× technologies, both involving microfluidic devices, were used in a flurry of publications at the end of 2018 and early 2019 (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). These approaches utilized the well-deduced spatiotemporal reference maps of the Arabidopsis root in order to ascribe cell identity. While these studies differed in the numbers of cells sequenced, they were able to obtain generally similar groups of cell types, as partitioned

by transcriptome variation. Novel biological advances included inferences made from mapping the trajectory of root epidermal cells into hair cell and nonhair cell identity, as well as multiple states in the endodermis developmental trajectory.

Single-cell studies have also elucidated the influence of heat shock or sucrose on cell identity (Jean-Baptiste et al., 2019; Shulse et al., 2019). Heat shock results in subtle changes in cell identity, while sucrose changes in the proportion of cell types but not in their identity per se: instead, many cell type- or tissue-specific responses were observed. Finally, the use of genetics or inducible lines highlights the power of single-cell RNA-sequencing in revealing the complex regulation of cell type identity.

More recently, single-cell profiles of the female gametophyte (Song et al., 2020) and stomata (Liu et al., 2020) have been published. It is likely that these profiles will soon be joined by profiles of a diversity of cell types in Arabidopsis found throughout the plant body and in more recalcitrant tissues (Rodriguez-Villalon and Brady, 2019). Single-cell chromatin accessibility profiles have also been published as a preprint, which further illustrate the complexity of gene regulation at cellular resolution (Dorrity et al., 2020; Farmer et al., 2020). It is very exciting that we can now examine the genomes of single cells, which is sure to provide more insight into plant biology.

CrY2H-Seq me a river (of data)

Given the rapidly decreasing costs and vastly improved capacity of NGS platforms, making an assay “sequenceable” is one way to dramatically increase its throughput. Mike Snyder and colleagues have documented the proliferation of “-seq”-based methods over the first part of the past decade, plotting these methods by year of publication and by the magnitude of impact in terms of number of citations of the method (Reuter et al., 2015). While RNA-seq has had the greatest impact of these high-throughput sequencing technologies, the CrY2H-seq method (Trigg et al., 2017) is sure to have a large influence in our ability to decipher the interactome of Arabidopsis and other plants in the coming years. In this method, the coding sequences for proteins of interest are cloned into activation-domain and DNA-binding domain “bait” and “prey” vectors using specially designed plasmids. These vectors are brought together in a yeast strain with a Cre reporter. If two proteins interact, a Cre recombinase is produced in this reporter strain, such that a new plasmid is formed to create fused fragments of the coding sequences of the interacting pairs. This fused product can be rapidly sequenced using NGS technologies. Reverse genetic methods (Charloteaux et al., 2011) can be used to determine exactly how two proteins interact to in a rapid and efficient manner. That said, good old Y2H performed using standard cloning and sequencing procedures has identified hundreds of thousands of protein–protein interactions. Some of these interactions were identified in small-scale screens, while others were uncovered in massive efforts (more on this

later). We also touch on other notable “-seq” methods below.

Epigenomics and chromatin accessibility surveys

The discovery of gene silencing in plants mediated by small RNAs in the late 1990s by Sir David Baulcombe and colleagues (Hamilton and Baulcombe, 1999; Dalmay et al., 2000) spurred the growth of the plant epigenetics field. Modifications to histones and to the genome itself via methylation are referred to as the epigenome (Bernstein et al., 2010); their discovery has been instrumental during the last two decades for helping to decipher the functional elements of plant genomes.

Studies in Arabidopsis have led the way in decoding plant epigenomes. Many of the original plant epigenome studies mapped the location of small RNAs, DNA methylation, and histone modifications using massively parallel signature sequencing (Meyers et al., 2004) or chromatin–immunoprecipitation followed by hybridization to tiling microarrays (Lu et al., 2005; Zhang et al., 2006, 2007; Yazaki et al., 2007; Zilberman et al., 2007; Bernatavichute et al., 2008; Zhang et al., 2009; Roudier et al., 2011; Coleman-Derr and Zilberman, 2012; Li et al., 2015), both enabled by a high-quality genome assembly and annotation. These epigenomics studies uncovered highly distinct properties that demarcate euchromatin (gene-rich) and heterochromatin (gene-poor, transposon, and repeat rich) based on small RNA and chromatin modification patterns. Since these original studies, numerous epigenomic studies have been carried out in a variety of Arabidopsis accessions (Vaughn et al., 2007; Schmitz et al., 2013; Dubin et al., 2015; Hagemann et al., 2015; Kawakatsu et al., 2016) and in numerous plant species (Gent et al., 2013; West et al., 2014; Niederhuth et al., 2016; Oka et al., 2017; Li et al., 2019; Lu et al., 2019; Ricci et al., 2019; Zhao et al., 2020), indicating that the patterns and distributions originally discovered in Arabidopsis Col-0 are generally found throughout the eukaryotes.

“Active” genomic regions were originally identified based on sensitivity to endonuclease cleavage (Gottesfeld et al., 1975; Weintraub and Groudine, 1976; Wu et al., 1979; Keene et al., 1981; Feng and Villeponteau, 1992). With the advent of high-throughput sequencing, endonuclease hypersensitivity (DNase-seq, MNase-seq) and transposase-mediated insertions (ATAC-seq) have been used to delineate regulatory DNA (i.e. accessible chromatin) genome-wide in hundreds of human cell types, animals, and several plant species (Thomas et al., 2011; Neph et al., 2012; Thurman et al., 2012; Sullivan et al., 2014; Yue et al., 2014; Rodgers-Melnick et al., 2016; Lu et al., 2017; Oka et al., 2017; Ricci et al., 2019), including Arabidopsis and maize (*Zea mays*), among others. Although most data sets have been generated in Arabidopsis, with its unusually compact, repeat-poor genome, some general features of plant regulatory landscapes have emerged. As in animals, the regulatory compartment in plants is small, scales with genome size (ranging from 4%

in Arabidopsis to 0.6% in maize), and is depleted for DNA methylation (Lu et al., 2019; Crisp et al., 2020). Unlike in animals, the majority of accessible chromatin sites in Arabidopsis and other plants tend to be closely associated with genes, localizing just upstream of transcription start sites, in addition to residing in intergenic regions and 5'-UTRs; however, as expected, the number of distal accessible sites increases with genome size (Maher et al., 2018; Lu et al., 2019). A subset of distal accessible sites, i.e. putative long-range enhancers, do share some of the histone modifications associated with enhancers in animals; however, clear distinctions are emerging (Oka et al., 2017; Lu et al., 2019; Ricci et al., 2019).

In both plants and animals, trait-associated variants are enriched in accessible chromatin. In humans, of the 5,654 noncoding variants associated with 207 diseases and 447 quantitative phenotypes, ~80% reside either within accessible chromatin sites or in linkage disequilibrium with variants at these sites (Maurano et al., 2012). Although GWAS in Arabidopsis have only been performed on a few strains (< 200, ~100 quantitative traits), genetic variants associated with over 70 traits tend to reside in accessible chromatin sites (Sullivan et al., 2014).

Somewhat paradoxically, the vast majority of differentially accessible sites in divergent Arabidopsis strains show no underlying genetic variation or differences in chromatin modifications (Alexandre et al., 2018), implying indirect effects on chromatin accessibility at many individual loci. In general, chromatin accessibility is only weakly correlated with the expression of nearby genes; this correlation improves when considering sites with dynamically changing accessibility across different tissues, developmental stages, or in response to treatments (Sullivan et al., 2014, 2019; Maher et al., 2018). However, even if an accessible site is fully deleted, as frequently found among diverse Arabidopsis strains, only 25% of nearby genes show significant changes in gene expression (Alexandre et al., 2018). This weak correlation is observed in both ways: despite the massive changes in gene expression, the majority of accessible chromatin sites in Arabidopsis show few changes across tissues, developmental stages, and in response to treatments (Sullivan et al., 2014, 2019; Maher et al., 2018). The relative stasis of the Arabidopsis regulatory landscape compared to animals suggests that cell and tissue identity might be less rigidly epigenetically encoded in plants; alternatively, tissue heterogeneity in bulk studies may contribute to this effect. The latter interpretation is supported by results of a single-cell ATAC-seq study of Arabidopsis and maize roots, in which ~30% of all accessible sites showed cell type-specific patterns (Dorrity et al., 2020; Farmer et al., 2020; Marand et al., 2020), greatly exceeding the estimates of 5%–10% of dynamic sites in bulk studies (Sullivan et al., 2019). Although single-cell approaches discovered many more differentially accessible sites, they did not resolve the weak correlation between chromatin accessibility and gene expression at individual loci. This is consistent with the relevance of indirect effects, such as the

binding of transcription factors (TFs) that poise a gene for activation and/or the binding of TFs that mediate gene repression. Further confounding these results is the lack of direct measurements of mRNA abundance and chromatin accessibility from the exact same cell. Future efforts to utilize multi-omic methods will no doubt resolve these questions. Nevertheless, a cell's entire regulatory landscape or its transcriptome independently can capture a cell's identity, arguing against simplistic single-locus models to explain regulatory output (Dorrity et al., 2020).

A promising strategy to understand regulatory elements at nucleotide resolution is STARR-seq, a massively parallel plasmid-based assay that determines the activity and strength of putative promoters and enhancers by testing large libraries of fragments for their ability to enhance transcription (Arnold et al., 2013). Recent efforts in plants (Ricci et al., 2019; Sun et al., 2019; Jores et al., 2020) showcase this method's potential for identifying distal enhancers and using saturation mutagenesis to define functional residues, which commonly overlap with clusters of TF motifs. The comprehensive enumeration of Arabidopsis TF motifs has been a major step toward interpreting accessible chromatin and STARR-seq data. Numerous groups have contributed to this effort with ChIP-seq data for specific TFs and with protein-binding microarrays for multiple TFs (Weirauch et al., 2014). However, DNA affinity purification sequencing, a high-throughput assay that uses in vitro-expressed TFs to interrogate naked genomic DNA, was a true game changer (O'Malley et al., 2016). When applied to all 1,725 Arabidopsis TFs, this approach identified high-confidence motifs for 529 TFs, representing all major TF families. However, the typically short binding motifs often do not suffice to resolve TF identity beyond TF families. That said, it is hard to imagine that this method would have been developed without a robust genome sequence.

How will we resolve the complexity of gene regulation? We posit that the existing motif information, together with integrated single-cell ATAC-seq and single-cell RNA-seq data, will ultimately allow us to resolve the direct and indirect effects in gene regulation. Single-cell ATAC-seq can identify cell type-specific TF family motif enrichments. In turn, single-cell RNA-seq will identify the specific TF family member whose expression changes across cells can explain accessibility changes in sites containing the respective TF motif (Dorrity et al., 2020). However, building these anticipated models of gene regulatory networks will require many more cells than have currently been sampled to fully capture the range of possible cell states.

Genomics for my research: resources for identifying mutations, functions, interactions, and networks

Mapping by sequencing

It goes without saying that the ability to conduct exploratory genetics has underpinned the success of Arabidopsis among model plants. The linking of genes with a biological

process in an unbiased manner remains an unparalleled approach for understanding gene function. Since this approach is typically based on mutagenesis with a chemical mutagen, which induces a large number of mutations per genome, it increases the chance of identifying plants with a relevant phenotype and of isolating a wide spectrum of mutations ranging from amorphs to neomorphs. The process of identifying a causative mutation via positional cloning through the association of phenotype with genotype has changed dramatically over the past few decades. In the 1990s, mapping causative mutations was a laborious, time-consuming process that involved chromosome walking, whereby a physical map was assembled from YACs and markers had to be identified one by one (Goodman et al., 1995). Once a map position was established, significant work remained to pinpoint the gene containing the causative mutation. It is perhaps not surprising that this process often consumed all of a graduate student's time at the bench.

A decade later, when the Arabidopsis whole genome sequence became available, researchers could easily identify mapping markers, which sped up the process of gene cloning from a multiyear process to a year or less (Arabidopsis Genome Initiative, 2000; Lukowitz et al., 2000; Jander et al., 2002). Although faster, this process was still tedious since, after initial rough mapping using bulk-segregant analysis, fine mapping required the researcher to follow markers in approximately 1,000 segregating plants. Another decade later, a third generation of mapping, spurred on by next-generation whole-genome sequencing technology, offered an even faster trajectory from phenotype to gene (Schneeberger et al., 2009; Cuperus et al., 2010; Austin et al., 2011). In addition, third-generation mapping or mapping-by-sequencing afforded a number of different options for cloning a causative mutation (for a more extensive review, see Schneeberger, 2014). With mapping-by-sequencing, it is also possible to sequence mutant genomes directly (Ashelford et al., 2011; Nordström et al., 2013). The advantage of this approach is that it has the potential to capture multiple mutations responsible for a specific phenotype. Although sorting and filtering the large number mutagen-induced mutations is not trivial, it is possible to improve the odds of successfully identifying the causative mutation(s) if multiple mutant alleles are available from the same mutant pool.

Apart from deciding which crossing scheme to use, mapping-by-sequencing experiments must also consider the number of recombinant plants to sequence, the sequencing coverage, and the type of sequencing (single- or paired-end). These practical aspects are important, since they will not only influence the overall success of the mapping experiment but also the overall cost associated with the mapping experiment (James et al., 2013). Wilson-Sánchez et al. (2019) used computer simulations to assess different mapping scenarios, with the goal of creating a guide for better experimental design. Here, they considered whether different sequencing technologies are better suited to mapping experiments and the sequencing depth required for calling

single nucleotide variants at high confidence. As with James et al. (2013), the authors also considered outcross versus backcross schemes, the number of genomes that should be sequenced for ultimate accuracy, and the best ways to differentiate between background mutations versus induced mutations. In addition, they also considered what they call "pseudo-backcrossing," where two mutants with additive phenotypes are combined to produce the F2 mapping population, which can then be used to simultaneously clone the causative genes of both mutants.

Another important practical aspect of mapping-by-sequencing is of course data analysis. In practice, this means sorting through mutations that were induced by the mutagen as well as those that are naturally occurring between diverged strains. For example, commonly used mutagens such as ethyl-methanesulfonate, depending on the dose applied, can cause in excess of a 1,000 mutations per genome (Jander et al., 2002). In addition, polymorphisms between the two most commonly used ecotypes of Arabidopsis, Col-0, and Ler, are on the order of 55,000 (Jander et al., 2002). The best available tools to conduct this type of analysis in Arabidopsis vary in their requirement for the researcher to have some coding knowledge or to prepare data prior to their implementation (Schneeberger et al., 2009; Austin et al., 2011; Wachsman et al., 2017). For example, while all the tools support variant calling, mutation mapping, and filtering of mutations for their effects, the SHOREmap tool (Schneeberger et al., 2009; Sun and Schneeberger, 2015) and the SIMPLE tool (Wachsman et al., 2017) are both command line pipelines, while the next-generation mapping (NGM) tool (Austin et al., 2011) is a web-based tool which, in contrast to the other tools, looks for homozygosity islands linked to the causative mutation rather than allele frequencies. This means that the NGM tool is only viable for mapping outcrossed populations, while the SHOREmap and SIMPLE tools can be used to analyze backcrossed populations. The choice of tool will therefore depend on the starting mapping population and on the confidence of the researcher in implementing the mapping tool. These tools are a testament to how far we have come from the early days of positional cloning and to the power of a mature genome sequence: rather than a chromosome walk, we now conduct a digital walk.

The aforementioned forward genetic approaches, while transformative in their own way, have been complemented by reverse genetic approaches enabled in the past 20 years by the sequencing of large T-DNA collections, notably the SALK, SAIL, WiscDsLox, and GABI-KAT lines (Sessions et al., 2002; Alonso et al., 2003; Woody et al., 2007; Kleinboelting et al., 2012). The availability of these lines, especially those made available in an open manner from their inception, has permitted knockout mutations for almost any desired gene to be ordered (from the two stock centers that had been set up as recommended in 1990, the ABRC in Ohio and the Nottingham Arabidopsis Stock Centre in the UK) at the click of a mouse, through websites like the Ecker

Laboratory's SIGnAL T-DNA Express site. These mutants can easily be examined for phenotypes to support presumed biological roles for the affected gene (O'Malley and Ecker, 2010). NGS (with mapping done to a high-quality reference genome) has amplified the power of these lines by allowing the identification of multiple insertions, sometimes with complex architectures (O'Malley et al., 2007; Jupe et al., 2019). The availability of a high-quality reference sequence also allows targeted genetic modifications to be made for reverse genetics with the CRISPR/Cas9/sgRNA system (Jiang et al., 2013).

Genomic databases for hypothesis generation

How do we know which genes to focus on for reverse genetic approaches? Databases provide many leads (Brady and Provart, 2009). The first database for Arabidopsis genomics to come online was the *Arabidopsis thaliana* Database (Flanders et al., 1998; Rhee et al., 1999), which provided a link between physical maps and sequences as they became available and provided a visualization of the AGI's sequencing progress. This was followed soon after by TAIR (Garcia-Hernandez et al., 2002; Rhee et al., 2003), which enables the exploration of other gene sequences in a gene's neighborhood, identification of similar sequences via BLAST (Altschul et al., 1990), gene functional classification using Gene Ontology (Ashburner et al., 2000), gene family membership, and more. Early gene expression databases, such as Genevestigator (Zimmermann et al., 2004) and the Bio-Analytic Resource (BAR, originally published at the "Botany Array Resource"; Toufighi et al., 2005), provided access to gene expression data sets that were being generated and published as part of the AtGenExpress effort.

The Arabidopsis eFP Browser (Winter et al., 2007) at the BAR displays a selected gene's expression pattern by dynamically coloring the tissues in a pictographic representation of a plant based on gene transcript levels from multiple experiments. This tool is deceptively simple, but it provides a powerful interface for exploring and visualizing early and more recent atlases of development (Schmid et al., 2005; Klepikova et al., 2016), abiotic stress (Kilian et al., 2007), biotic stress (AtGenExpress initiative), chemical experiments (Goda et al., 2008), and many tissue-specific experiments. These data sets contain more than 35 million records, representing "big data" exploration in under five clicks.

While gene expression data can provide useful ideas for narrowing down the phenotypic search space, other types of data, such as protein-protein interaction data and network-based data, are also increasingly being used for hypothesis generation. The Arabidopsis Interactome 1 (Dreze et al., 2011) measured approximately 6,200 interactions between approximately 2,700 Arabidopsis proteins. Further large- and meso-scale studies (e.g. Lumba et al., 2014; Smakowska-Luzan et al., 2018; Cao et al., 2019; Carianopol et al., 2020) and hundreds of small-scale experiments, collated in tools such as the Arabidopsis Interactions Viewer 2 (Dong et al., 2019), are also valuable resources. Many of these data are collected in unbiased ways, and thus, being

able to identify interaction partners can provide high-quality candidate genes for a researcher's biological system. Likewise, networks based on coexpression, functional association, or gene regulation (Lee et al., 2010; Bassel et al., 2011; Taylor-Teeple et al., 2015 as a few examples) can provide avenues for hypothesis generation.

Visualizing the future

Being able to explore data in a unified manner helps leverage the incredible genomic data that have been generated for Arabidopsis over the past 20 years. ePlant (Waese et al., 2017) introduces the concept of a zoomable user interface to help users explore Arabidopsis data from the kilometer level down to the nanometer level of data using a combination of chart types. ePlant connects to several publicly available web services to download genome, proteome, interactome, transcriptome, and 3D molecular structure data for one or more genes or gene products of interest. The molecule viewer module is especially interesting because it maps information from four separate databases onto a 3D model of the selected protein's molecular structure: complete protein sequences (Krishnakumar et al., 2015); nonsynonymous single nucleotide polymorphism locations in the underlying gene sequence (Joshi et al., 2012) with a list of ecotypes in which they are found; Pfam domains (Bateman et al., 2002 and updates); and CDD feature hits (Marchler-Bauer et al., 2002 and updates). These features make it easy to see where polymorphisms occur within a protein molecule and speculate how they might interact with binding sites and other domains of interest. Zhang et al. (2020) just released ARS, an Arabidopsis RNA-seq database for exploring expression levels in approximately 20,000 RNA-seq data sets.

A genome browser is also essential for exploring Arabidopsis genome data. This graphical tool allows users to visualize data mapped to the genomic sequence axis alongside gene model annotations and data from other laboratories' experiments. Currently, there are three major genome browsers available for plant scientists, which vary with respect to interactivity, number of features, speed, and usability: Integrated Genome Browser from BioViz.org (Freese et al., 2016), Integrative Genome Viewer from the Broad Institute (Robinson et al., 2011), and JBrowse, a web-based tool data providers must incorporate into their web sites (Buels et al., 2016). TAIR has most recently rescued Araport's extensive collection of JBrowse tracks for display in its own version of JBrowse running on its locus pages (Pasha et al., 2020). This provides a centralized framework for adding new tracks showing non-coding transcripts (Kindgren et al., 2020), epigenomic data (Hofmeister and Schmitz, 2018), and other emerging data sets.

IGB and IGV are both stand-alone desktop tools that users download, install, and run on their local machines. These tools can open files stored locally or on the internet via URLs. IGV from the Broad Institute is better known, but IGB has more features, such as ProtAnnot for exploring the effects of alternative splicing on protein-coding genes

(Mall et al., 2016), and it provides access to RNA-seq, ChIP-seq, T-DNA, and other Arabidopsis data sets. This reflects IGB's early funding from the National Science Foundation's Arabidopsis 2010 program (https://www.nsf.gov/awardsearch/showAward?AWD_ID=0820371).

The impact of the publication of the Arabidopsis genome continues to this day with the sequence and associated tools still guiding the daily activities of researchers. Dr Sara Farrona from the National University of Ireland, Galway recalls, "I still remember when the publication of the Arabidopsis genome came out in 2000. I had started my PhD project focused on chromatin remodeling proteins in Arabidopsis just a few months prior and having access to its genome completely shaped the way I tackled my research. In the following years, the Arabidopsis genome, its browser and all the information publicly available for each of its genes would have and still has an extraordinary impact. I and members of my lab still use it on a daily basis."

Overall, the impact of the genome is perfectly summed up by Dr Piers Hemsley from the University of Dundee "Almost every aspect of my research, from cloning and expression analysis to proteomics and EvoDevo work, would be next to impossible without it. As an enabling resource it has yet to be surpassed in its application to almost every aspect of my work." The impact is also evident outside of the Arabidopsis research community in the myriad of plant genome articles that use the Arabidopsis genome sequence to help with assembly or to annotate genes, most recently ones for eggplant (*Solanum melongena*; Wei et al., 2020) and tea (*Camellia sinensis* var. *sinensis*; Xia et al., 2020). The generation of other types of omics data from agronomically important plants will benefit from landmark methods and data sets first generated in Arabidopsis, all predicated on a high-quality genome sequence.

The "ultimate expression" of the researchers who met at the "Functional Genomics and the Virtual Plant: A blueprint for understanding how plants are built and how to improve them" workshop that led to the Multinational Coordinated Arabidopsis Functional Genomics Project was "nothing short of a virtual plant which one could observe growing on a computer screen, stopping this process at any point in that development, and with the click of a computer mouse, accessing all the genetic information expressed in any organ or cell under a variety of environmental conditions." Now, 10 years after the Arabidopsis 2010 projects wrapped up, are we there yet?

The answer is that we are getting closer (e.g. Shapiro et al., 2015; Banwarth-Kuhn et al., 2019; Maheshwari et al., 2020), but there is still a long way to go. Further documenting other types of molecules and their modifications in different Arabidopsis tissues will increase our knowledge of plant biology beyond transcripts and genomes: atlases of lipids, proteins, hormones, and SUMOylation anyone? Single-cell methods are sure to provide vast amounts of new data, but understanding how cells, tissues and whole plants respond to environmental cues and perturbations, let alone

being able to model this at a multi-scale level, is perhaps another decade away. The Plant Cell Atlas project (Rhee et al., 2019; <http://www.plantcellatlas.org/>) will map "molecular machineries to cellular and subcellular domains, follow their dynamic movements, and describe their interactions [to] accelerate discovery in plant science and help to solve imminent societal problems." Undoubtedly, Arabidopsis will be one of the Plant Cell Atlas's subject species. A recent virtual meeting attracted over 300 participants from around the world interested in this project, upholding the spirit of cooperation that kicked off the Arabidopsis genome project more than 30 years ago.

Acknowledgments

We are grateful to David Galbraith from the University of Arizona for Latin declension guidance for the title of this article, to the reviewers for suggestions, and to the editors at The Plant Cell for wordsmithing.

Funding

C.Q. was funded by NSF RESEARCH-PGR 1748843. S.M.B. was partially funded by an HHMI Faculty Scholar Fellowship, NSF-PGRP1856749, and USDA BTT-EAGER 2019-67013-29012. N.J.P. was funded by grants from NSERC and Genome Canada/Ontario Genomics. G.P. received support from the UKRI-BBSRC grant BB/M004376/1, GARNet2020. A.E.L. was funded by NIH NIGMS R01GM103463 (Integrated Genome Browser and IGB Apps) and NSF 0820371 (Arabidopsis 2010: Visualization Software and Data Server for Arabidopsis). R.J.S. was supported by the National Science Foundation (MCB-1856143) and the National Institutes of Health (R01-GM134682). K.S. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2048/1–390686111, and the European Research Council (ERC) Grant "INTERACT" (802629).

Conflict of interest statement. None declared.

References

- 1001 Genomes Consortium** (2016) 1,135 Genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell* **166**: 481–491
- Alexandre CM, Urton JR, Jean-Baptiste K, Huddleston J, Dorriety MW, Cuperus JT, Sullivan AM, Bemm F, Jolic D, Arsovski AA, et al.** (2018) Complex relationships between chromatin accessibility, sequence divergence, and gene expression in Arabidopsis thaliana. *Mol Biol Evol* **35**: 837–854
- Alonso JM, Stepanova AN, Lisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al.** (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301**: 653–657
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ** (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410
- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796–815

- Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A** (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**: 1074–1077
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.** (2000) Gene ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29
- Ashelford K, Eriksson ME, Allen CM, D'Amore R, Johansson M, Gould P, Kay S, Millar AJ, Hall N, Hall A** (2011) Full genome re-sequencing reveals a novel circadian clock mutation in *Arabidopsis*. *Genome Biol* **12**: R28
- Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, McCourt P, Guttman DS** (2011) Next-generation mapping of *Arabidopsis* genes. *Plant J* **67**: 715–725
- Banwarth-Kuhn M, Nematbakhsh A, Rodriguez KW, Snipes S, Rasmussen CG, Reddy GV, Alber M** (2019) Cell-based model of the generation and maintenance of the shape and structure of the multilayered shoot apical meristem of *Arabidopsis thaliana*. *Bull Math Biol* **81**: 3245–3281
- Barta A, Kalyna M, Reddy ASN** (2010) Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR Proteins) in plants. *Plant Cell* **22**: 2926–2929
- Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provart NJ** (2011) Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc Natl Acad Sci* **108**: 9709–9714
- Bateman A, Birney E, Cerruti L, Durbin R, Ewinger L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL** (2002) The Pfam protein families database. *Nucleic Acids Res* **30**: 276–280
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE** (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* **3**: e3156
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al.** (2010) The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* **28**: 1045–1048
- Brady SM, Provart NJ** (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* **21**: 1034–1051
- Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsiek CG, Lewis SE, Stein L, et al.** (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol* **17**: 66
- Calixto CPG, Guo W, James AB, Tzioutziou NA, Entizne JC, Panter PE, Knight H, Nimmo HG, Zhang R, Brown JWS** (2018) Rapid and dynamic alternative splicing impacts the *Arabidopsis* cold response transcriptome. *Plant Cell* **30**: 1424–1444
- Cao FY, Khan M, Taniguchi M, Mirmiran A, Moeder W, Lumba S, Yoshioka K, Desveaux D** (2019) A host–pathogen interactome uncovers phytopathogenic strategies to manipulate plant ABA responses. *Plant J* **100**: 187–198
- Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, et al.** (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* **43**: 956–963
- Carianopol CS, Chan AL, Dong S, Provart NJ, Lumba S, Gazzarrini S** (2020) An abscisic acid-responsive protein interaction network for sucrose non-fermenting related kinase1 in abiotic stress response. *Commun Biol* **3**: 145
- Charloteaux B, Zhong Q, Dreze M, Cusick ME, Hill DE, Vidal M** (2011) Protein–protein interactions and networks: forward and reverse edgecuts. In **Jl Castrillo, SG Oliver**, eds, *Yeast Systems Biology: Methods and Protocols*, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 197–213
- Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD** (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* **89**: 789–804
- Coleman-Derr D, Zilberman D** (2012) Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet* **8**
- Crisp PA, Marand AP, Noshay JM, Zhou P, Lu Z, Schmitz RJ, Springer NM** (2020) Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc Natl Acad Sci U S A* **117**: 23991–24000
- Cuperus JT, Montgomery TA, Fahlgren N, Burke RT, Townsend T, Sullivan CM, Carrington JC** (2010) Identification of MIR390a precursor processing-defective mutants in *Arabidopsis* by direct genome sequencing. *Proc Natl Acad Sci U S A* **107**: 466–471
- Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC** (2000) An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**: 543–553
- Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MCP** (2019) Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev Cell* **48**: 840–852.e5
- Dong S, Lau V, Song R, Ierullo M, Esteban E, Wu Y, Sivieng T, Nahal H, Gaudinier A, Pasha A, et al.** (2019) Proteome-wide, structure-based prediction of protein-protein interactions/new molecular interactions viewer. *Plant Physiol* **179**: 1893–1907
- Dorrity MW, Alexandre C, Hamm M, Vigil A-L, Fields S, Queitsch C, Cuperus J** (2020) The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *bioRxiv*, doi: 2020.07.17.204792 (July 27, 2020 accessed)
- Dreze M, Byrdsong D, Dricot A, Duarte M, Gebreab F, Gutierrez BJ, MacWilliams A, Monachello D, Mukhtar MS, Poulin MM, et al.** (2011) Evidence for network evolution in an *Arabidopsis* interactome map. *Science* **333**: 601–607
- Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Casale FP, Drewe P, Kahles A, Jean G, Vilhjálmsson B, et al.** (2015) DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife* **4**: e05255
- Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Picó FX, Alonso-Blanco C, Hancock AM** (2017) African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **114**: 5213–5218
- Efroni I, Mello A, Nawy T, Ip P-L, Rahni R, DelRose N, Powers A, Satija R, Birnbaum KD** (2016) Root regeneration triggers an embryo-like sequence guided by hormonal interactions. *Cell* **165**: 1721–1733
- English AC, Patel KS, Loraine AE** (2010) Prevalence of alternative splicing choices in *Arabidopsis thaliana*. *BMC Plant Biol* **10**: 102
- Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M** (2020) The impact of chromatin remodeling on gene expression at the single cell level in *Arabidopsis thaliana*. *bioRxiv*: 2020.07.27.223156
- Feng J, Villeponteau B** (1992) High-resolution analysis of c-fos chromatin accessibility using a novel DNase I-PCR assay. *Biochim Biophys Acta* **1130**: 253–258
- Flanders DJ, Weng S, Petel FX, Cherry JM** (1998) AtDB, the *Arabidopsis thaliana* database, and graphical-web-display of progress by the *Arabidopsis* genome initiative. *Nucleic Acids Res* **26**: 80–84
- Freese NH, Norris DC, Loraine AE** (2016) Integrated genome browser: visual analytics platform for genomics. *Bioinforma Oxf Engl* **32**: 2089–2095
- Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, et al.** (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **477**: 419–423
- Garcia-Hernandez M, Berardini T, Chen G, Crist D, Doyle A, Huala E, Knee E, Lambrecht M, Miller N, Mueller LA, et al.** (2002) TAIR: a resource for integrated *Arabidopsis* data. *Funct Integr Genomics* **2**: 239–253
- Gent JJ, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK** (2013) CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res* **23**: 628–637

- Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, et al.** (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J Cell Mol Biol* **55**: 526–542
- Goodman HM, Ecker JR, Dean C** (1995) The genome of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **92**: 10831–10835
- Gottesfeld JM, Murphy RF, Bonner J** (1975) Structure of transcriptionally active chromatin. *Proc Natl Acad Sci U S A* **72**: 4404–4408
- Gulledge AA, Roberts AD, Vora H, Patel K, Loraine AE** (2012) Mining *Arabidopsis thaliana* RNA-seq data with Integrated Genome Browser reveals stress-induced alternative splicing of the putative splicing regulator SR45a. *Am J Bot* **99**: 219–231
- Hagmann J, Becker C, Müller J, Stegle O, Meyer RC, Wang G, Schneeberger K, Fitz J, Altmann T, Bergelson J, et al.** (2015) Century-scale methylation stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genet* **11**: e1004920
- Hamilton AJ, Baulcombe DC** (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950–952
- Hofmeister BT, Schmitz RJ** (2018) Enhanced JBrowse plugins for epigenomics data visualization. *BMC Bioinformatics* **19**: 159
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P** (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* **2008**: 420747
- James GV, Patel V, Nordström KJ, Klases JR, Salomé PA, Weigel D, Schneeberger K** (2013) User guide for mapping-by-sequencing in *Arabidopsis*. *Genome Biol* **14**: R61
- Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL** (2002) *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol* **129**: 440–450
- Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT** (2019) Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**: 993–1011
- Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP** (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res* **41**: e188
- Jiao W-B, Schneeberger K** (2020) Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat Commun* **11**: 989
- Jores T, Tonnie J, Dorrity MW, Cuperus J, Fields S, Queitsch C** (2020) Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell* **32**: 2120–2131
- Joshi HJ, Christiansen KM, Fitz J, Cao J, Lipzen A, Martin J, Smith-Moritz AM, Pennacchio LA, Schackwitz WS, Weigel D, et al.** (2012) 1001 Proteomes: a functional proteomics portal for the analysis of *Arabidopsis thaliana* accessions. *Bioinformatics* **28**: 1303–1306
- Jupe F, Rivkin AC, Michael TP, Zander M, Motley ST, Sandoval JP, Slotkin RK, Chen H, Castanon R, Nery JR, et al.** (2019) The complex architecture and epigenomic impact of plant T-DNA insertions. *PLOS Genet* **15**: e1007819
- Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urlich MA, Castanon R, Nery JR, Barragan C, He Y, et al.** (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**: 492–505
- Keene MA, Corces V, Lowenhaupt K, Elgin SC** (1981) DNase I hypersensitive sites in *Drosophila* chromatin occur at the 5' ends of regions of transcription. *Proc Natl Acad Sci U S A* **78**: 143–146
- Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K** (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J Cell Mol Biol* **50**: 347–363
- Kindgren P, Ivanov M, Marquardt S** (2020) Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in *Arabidopsis*. *Nucleic Acids Res* **48**: 2332–2347
- Kleinboelting N, Huep G, Kloetgen A, Viehoveer P, Weisshaar B** (2012) GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Res* **40**: D1211–D1215
- Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA** (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J* **88**: 1058–1070
- Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, Rosen BD, Cheng C-Y, Moreira W, Mock SA, et al.** (2015) Araport: the *Arabidopsis* Information Portal. *Nucleic Acids Res* **43**: D1003–D1009
- Lee J, Ambaru B, Thakkar P, Marcotte EM, Rhee SY** (2010) Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nat Biotechnol* **28**: 149–156
- Li Y, Mukherjee I, Thum KE, Tanurdzic M, Katari MS, Obertello M, Edwards MB, McCombie WR, Martienssen RA, Coruzzi GM** (2015) The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants. *Genome Biol* **16**: 79–79
- Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, et al.** (2019) The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol* **20**: 139–139
- Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M, et al.** (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**: 761–768
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR** (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536
- Liu Z, Zhou Y, Guo J, Li J, Tian Z, Zhu Z, Wang J, Wu R, Zhang B, Hu Y, et al.** (2020) Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol Plant* **13**: 1178–1193
- Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, Zhang Q, Vilhjálmsson BJ, Korte A, Nizhynska V, et al.** (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat Genet* **45**: 884–890
- Loraine AE, McCormick S, Estrada A, Patel K, Qin P** (2013) RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiol* **162**: 1092–1109
- Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ** (2005) Elucidation of the small RNA component of the transcriptome. *Science* **309**: 1567–1569
- Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ** (2017) Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res* **45**: e41
- Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ** (2019) The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants* **5**: 1250–1259
- Lukowitz W, Gillmor CS, Scheible W-R** (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol* **123**: 795–806
- Lumba S, Toh S, Handfield L-F, Swan M, Liu R, Youn J-Y, Cutler SR, Subramaniam R, Provart N, Moses A, et al.** (2014) A meso-scale abscisic acid hormone interactome reveals a dynamic signaling landscape in *Arabidopsis*. *Dev Cell* **29**: 360–372
- Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, Zumstein K, Woodhouse M, Bubb K, Dorrity MW, 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

- Maheshwari P, Assmann SM, Albert R** (2020) A guard cell abscisic acid (ABA) network model that captures the stomatal resting state. *Front Physiol* **11**
- Mall T, Eckstein J, Norris D, Vora H, Freese NH, Loraine AE** (2016) ProtAnnot: an App for Integrated Genome Browser to display how alternative splicing and transcription affect proteins. *Bioinformatics* **32**: 2499–2501
- Marand AP, Chen Z, Gallavotti A, Schmitz RJ** (2020) A cis-regulatory atlas in maize at single-cell resolution. *bioRxiv*: 2020.09.27.315499
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH** (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* **30**: 281–283
- Matsushika A, Makino S, Kojima M, Mizuno T** (2000) Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in *Arabidopsis thaliana*: insight into the plant circadian clock. *Plant Cell Physiol* **41**: 1002–1012
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al.** (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**: 1190–1195
- Mayer K, Schüller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Düsterhöft A, Stiekema W, Entian KD, Terryn N, et al.** (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402**: 769–777
- Meyers BC, Lee DK, Vu TH, Tej SS, Edberg SB, Matvienko M, Tindell LD** (2004) *Arabidopsis* MPSS. An online resource for quantitative expression analysis. *Plant Physiol* **135**: 801–813
- Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, Brooks S, Howe E, Porubsky D, Logsdon GA, et al.** (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**: 79–84
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, et al.** (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**: 83–90
- Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R** (2004) Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J Cell Mol Biol* **39**: 877–885
- Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, et al.** (2016) Widespread natural variation of DNA methylation within angiosperms. *Genome Biol* **17**: 194
- Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, et al.** (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* **3**: 1289
- Nordström KJV, Albani MC, James GV, Gutjahr C, Hartwig B, Turck F, Paszkowski U, Coupland G, Schneeberger K** (2013) Mutation identification by direct comparison of whole-genome sequencing data from mutant and wild-type individuals using k-mers. *Nat Biotechnol* **31**: 325–330
- Oka R, Zicola J, Weber B, Anderson SN, Hodgman C, Gent JI, Wesselink J-J, Springer NM, Hoefsloot HCJ, Turck F, Stam M** (2017) Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol* **18**: 137–137
- O'Malley RC, Alonso JM, Kim CJ, Leisse TJ, Ecker JR** (2007) An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the *Arabidopsis* genome. *Nat Protoc* **2**: 2910–2917
- O'Malley RC, Ecker JR** (2010) Linking genotype to phenotype using the *Arabidopsis* unimutant collection. *Plant J Cell Mol Biol* **61**: 928–940
- O'Malley RC, Huang S-SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR** (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**: 1280–1292
- Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D** (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res* **18**: 2024–2033
- Parry G, Provart NJ, Brady SM, Uzilday B** (2020) Current status of the multinational *Arabidopsis* community. *Plant Direct* **4**: e00248
- Pasha A, Subramaniam S, Cleary A, Chen X, Berardini TZ, Farmer A, Town C, Provart NJ** (2020) Araport lives: an updated framework for *Arabidopsis* bioinformatics. *Plant Cell* **32**: 2683–2686
- Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S, et al.** (2016) 50 years of *Arabidopsis* research: highlights and future directions. *New Phytol* **209**: 921–944
- Reuter JA, Spacek DV, Snyder MP** (2015) High-throughput sequencing technologies. *Mol Cell* **58**: 586–597
- Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al.** (2003) The *Arabidopsis* information resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res* **31**: 224–228
- Rhee SY, Birnbaum KD, Ehrhardt DW** (2019) Towards building a plant cell atlas. *Trends Plant Sci* **24**: 303–310
- Rhee SY, Weng S, Bongard-Pierce DK, García-Hernández M, Malekian A, Flanders DJ, Cherry JM** (1999) Unified display of *Arabidopsis thaliana* physical maps from AtDB, the A.thaliana database. *Nucleic Acids Res* **27**: 79–84
- Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, et al.** (2019) Widespread long-range cis-regulatory elements in the maize genome. *Nat Plants* **5**: 1237–1249
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP** (2011) Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26
- Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES** (2016) Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci U S A* **113**: E3177–E3184
- Rodriguez-Villalon A, Brady SM** (2019) Single cell RNA sequencing and its promise in reconstructing plant vascular cell lineages. *Curr Opin Plant Biol* **48**: 47–56
- Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, et al.** (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J* **30**: 1928–1938
- Ryu KH, Huang L, Kang HM, Schiefelbein J** (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol* **179**: 1444
- Schena M, Shalon D, Davis RW, Brown PO** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU** (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* **37**: 501–506
- Schmitz RJ, Schultz MD, Ulrich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, Ecker JR** (2013) Patterns of population epigenomic diversity. *Nature* **495**: 193–198
- Schneeberger K** (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat Rev Genet* **15**: 662–676
- Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jørgensen J-E, Weigel D, Andersen SU** (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat Methods* **6**: 550–551
- Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, et al.** (2002) A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* **14**: 2985–2994

- Shapiro BE, Tobin C, Mjolsness E, Meyerowitz EM (2015) Analysis of cell division patterns in the Arabidopsis shoot apical meristem. *Proc Natl Acad Sci U S A* **112**: 4815–4820
- Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, Gouran M, Turco GM, Zhu Y, O'Malley RC, Brady SM, Dickel DE (2019) High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep* **27**: 2241–2247.e4
- Smakowska-Luzan E, Mott GA, Parys K, Stegmann M, Howton TC, Layeghifard M, Neuhold J, Lehner A, Kong J, Grünwald K, et al. (2018) An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. *Nature* **553**: 342–346
- Somerville C, Koornneef M (2002) A fortunate choice: the history of Arabidopsis as a model plant. *Nat Rev Genet* **3**: 883–889
- Song Q, Ando A, Jiang N, Ikeda Y, Chen ZJ (2020) Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. *Genome Biol* **21**: 178
- Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, et al. (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep* **8**: 2015–2030
- Sullivan AM, Arsovski AA, Thompson A, Sandstrom R, Thurman RE, Neph S, Johnson AK, Sullivan ST, Sabo PJ, Neri FV III, et al. (2019) Mapping and dynamics of regulatory DNA in maturing Arabidopsis thaliana siliques. *Front Plant Sci* **10**
- Sun H, Schneeberger K (2015) SHOREmap v3.0: fast and accurate identification of causal mutations from forward genetic screens. *Methods Mol Biol* **1284**: 381–395
- Sun J, He N, Niu L, Huang Y, Shen W, Zhang Y, Li L, Hou C (2019) Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics Proteomics Bioinformatics* **17**: 140–153
- Taylor-Teeple M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, et al. (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**: 571–575
- Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, et al. (2011) Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome Biol* **12**: R43
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. (2012) The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82
- Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The botany array resource: e-Northerns, expression angling, and promoter analyses. *Plant J* **43**: 153–163
- Trigg SA, Garza RM, MacWilliams A, Nery JR, Bartlett A, Castanon R, Goubil A, Feeny J, O'Malley R, Huang S-SC, et al. (2017) CrY2H-seq: a massively-multiplexed assay for deep coverage interactome mapping. *Nat Methods* **14**: 819–825
- Vaughn MW, Tanurdzić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al. (2007) Epigenetic natural variation in Arabidopsis thaliana. *PLoS Biol* **5**: e174–e174
- Wachsmann G, Modliszewski JL, Valdes M, Benfey PN (2017) A SIMPLE pipeline for mapping point mutations. *Plant Physiol* **174**: 1307–1313
- Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, et al. (2017) ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* **29**: 1806–1821
- Wei Q, Wang J, Wang W, Hu T, Hu H, Bao C (2020) A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Hortic Res* **7**: 153
- Weigel D, Mott R (2009) The 1001 genomes project for Arabidopsis thaliana. *Genome Biol* **10**: 107
- Weintraub H, Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. *Science* **193**: 848–856
- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**: 1431–1443
- West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, Schmitz RJ, Springer NM (2014) Genomic distribution of H3K9me2 and DNA methylation in a maize genome. *PLoS One* **9**: e105267
- Wilson-Sánchez D, Lup SD, Sarmiento-Mañús R, Ponce MR, Micol JL (2019) Next-generation forward genetic screens: using simulated data to improve the design of mapping-by-sequencing experiments in Arabidopsis. *Nucleic Acids Res* **47**: e140
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**: e718
- Woody ST, Austin-Phillips S, Amasino RM, Krysan PJ (2007) The WiscDsLox T-DNA collection: an Arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J Plant Res* **120**: 157–165
- Wu C, Wong YC, Elgin SC (1979) The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* **16**: 807–814
- Xia E, Tong W, Hou Y, An Y, Chen L, Wu Q, Liu Y, Yu J, Li F, Li R, et al. (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol Plant* **13**: 1013–1026
- Yazaki J, Gregory BD, Ecker JR (2007) Mapping the genome landscape using tiling array technology. *Curr Opin Plant Biol* **10**: 534–542
- Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, et al. (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355–364
- Zapata L, Ding J, Willing E-M, Hartwig B, Bezdán D, Jiao W-B, Patel V, Velikkakam James G, Koornneef M, Ossowski S, et al. (2016) Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. *Proc Natl Acad Sci U S A* **113**: E4052–E4060
- Zhang H, Zhang F, Yu Y, Feng L, Jia J, Liu B, Li B, Guo H, Zhai J (2020) A comprehensive online database for exploring ~20,000 public Arabidopsis RNA-seq libraries. *Mol Plant* **13**: 1231–1233
- 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. *Mol Plant* **12**: 648–660
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. *Genome Biol* **10**: R62
- Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE (2007) Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci U S A* **104**: 4536–4541
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126**: 1189–1201
- Zhao L, Xie L, Zhang Q, Ouyang W, Deng L, Guan P, Ma M, Li Y, Zhang Y, Xiao Q, et al. (2020) Integrative analysis of reference epigenomes in 20 rice varieties. *Nat Commun* **11**: 2658–2658
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol* **136**: 2621–2632