# Genome-wide association study reveals influence of cell-specific gene networks on Soybean root system architecture

**Todd Michael**

tmichael@salk.edu

Salk Institute for Biological Studies https://orcid.org/0000-0001-6272-2875

**Ying Sun**

RareFlora

**Charlotte Miller**

Salk Institute for Biological Studies

**Ashish Rajurkar**

Salk Institute for Biological Studies

**Anthony Alyward**

RareFlora

**Ling Zhang**

Salk Institute

**Marieken Shaner**

San Diego State University

**Charles Copeland**

University of California San Diego

**Sean Jarrell-Hurtado**

Salk Institute for Biological Studies https://orcid.org/0009-0008-3488-4824

**Henge Ye**

University of Missouri-Columbia, Columbia

**Henry Nguyen**

University of Missouri https://orcid.org/0000-0002-7597-1800

**Wolfgang Busch**

Salk Institute for Biological Studies https://orcid.org/0000-0003-2042-7290

**Ryan Lynch**

Salk Institute for Biological Studies https://orcid.org/0000-0003-2032-6078

**Article**

**Keywords:** genome-wide association study (GWAS), Soybean (Glycine max), root system architecture 28 (RSA), metaphloem, endodermis, lateral root length

**Posted Date:** June 10th, 2025

**DOI:** https://doi.org/10.21203/rs.3.rs-6719211/v1



**Additional Declarations: Yes** there is potential Competing Interest. T.P.M. and W.B. are founders of the carbon sequestration company Cquesta.

## Title

Genome-wide association study reveals influence of cell-specific gene networks on Soybean root system architecture


## Authors

Ying Sun[1*], Charlotte Miller[1*], Ashish B. Rajurkar[1*], Ryan C. Lynch[1], Anthony Alyward[1], Ling Zhang[1], Marieken Shaner[1,6], Charles D. Copeland[1,7], Sean Jarrell-Hurtado[1], Heng Ye[5], Henry T. Nguyen[5], Wolfgang Busch[1#], Todd P. Michael[1,2,3,4#]

## Affiliations

[1]The Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[2]Department of Cell and Developmental Biology, School of Biological Sciences, University of California, San Diego, La Jolla, CA, 92093, USA
[3]Center for Marine Biotechnology and Biomedicine, University of California, San Diego, La Jolla, CA, 92093, USA
[4]Science and Conservation, San Diego Botanical Garden, Encinitas, CA, 92024, USA
[5]Division of Plant Sciences and National Center for Soybean Biotechnology, University of Missouri-Columbia, Columbia, MO, 65211, USA
[6] Present Address: Department of Biology, San Diego State University, San Diego, CA 92182,
[7] Present Address: School of Global Policy and Strategy, University of California, La Jolla, CA 92093

* Authors contributed equally
# correspondence: wbusch@salk.edu ; tmichael@salk.edu



## Key Words

genome-wide association study (GWAS), Soybean (*Glycine max*), root system architecture (RSA), metaphloem, endodermis, lateral root length



## Abstract

Root system architecture (RSA), the three-dimensional arrangement of roots in soil, is a critical determinant of plant productivity, resource use efficiency, and resilience to environmental stress. Despite its agronomic importance, RSA remains a largely untapped breeding target due to historical technical barriers in root phenotyping. We present *RADICYL* (Root Architecture 3D Cylinder), a scalable, non-invasive, gel-based platform enabling high-throughput, high-resolution quantification of 15 RSA traits in intact root systems. Applying *RADICYL* to a genetically diverse panel of 371 soybean accessions, we combined 3D phenotyping with genome-wide association studies (GWAS), single-nucleus RNA sequencing (snRNA-seq), and gene co-expression network (GCN) analysis to identify *RCE1* and *NPR3* as central regulators of RSA, suggesting auxin and salicylic acid-mediated signaling impacts RSA in specific root tissues. Functional validation in *Arabidopsis* mutants revealed conserved effects on root width and lateral root development. Our findings position the endodermis and metaphloem as key regulatory cell types and demonstrate how multi-omic frameworks can accelerate the discovery of functional genes underlying complex traits. This study establishes a foundation for cell-type-targeted genome editing and climate-smart crop engineering, offering actionable genetic targets to optimize root systems for improved nutrient acquisition, drought resilience, and deep carbon sequestration. By bridging genotype, cellular context, and phenotype, this work redefines RSA as a tractable and transformative trait for the future of crop improvement.


## Introduction

Root system architecture (RSA), the spatial configuration of roots in the soil is a fundamental trait that determines a plant's ability to acquire water and nutrients, directly influencing both fitness and yield. RSA is shaped by root growth, growth direction, and branching patterns (Slovak et al. 2016), all of which are governed by complex interactions between genetic factors and environmental cues (Lynch 2022). Considerable variation in RSA exists between species. For example, dicots typically form a root system with a dominant primary root that gives rise to multiple orders of lateral roots, whereas monocots develop more complex systems composed of primary, seminal, shoot-borne, and lateral roots. In both plant types, lateral root branching is a critical component of RSA and is initiated post-embryonically from the pericycle, a specialized tissue located between the central vascular cylinder and the endodermis (Parizot et al. 2008; Slovak et al. 2016). Lateral root initiation begins when a subset of founder cells in the pericycle is specified through an auxin-dependent process (De Smet et al. 2007). These founder cells subsequently divide to form a lateral root primordium, which develops into a mature lateral root. The spatial positioning of these initiation events is tightly regulated and, in *Arabidopsis thaliana*, has been shown to depend on an oscillatory, clock-like mechanism that determines the periodicity of lateral root formation (Wachsman et al. 2020). Beyond auxin signaling, a variety of regulatory pathways contribute to lateral root development, including those involving receptor-like kinases (RLKs) and additional signaling molecules (Rodriguez-Villalon et al. 2015; Jourquin, Fukaki, and Beeckman 2020; Ou, Kui, and Li 2021). These networks underscore the complexity and developmental plasticity of RSA, making it a highly responsive and adaptable trait across diverse plant species.

While genes and molecular processes involved in RSA have been extensively studied in *Arabidopsis*, they are understudied in crop species. One of the most important crop species is soybean (*Glycine max*), which ranks as the fourth largest crop globally. Soybean seeds contain high protein and edible oil levels and are used for human consumption, animal feed, and oil production (Guo et al. 2022; Chuang Zhao et al. 2017). Several genome-wide association studies (GWAS) have been conducted on RSA in soybean, although most of these studies focused on RSA data obtained by growing roots in environments restricting their growth to two-dimensions (2D). Two of the studies utilized pouch and wick systems based growth on moistened blue paper (Falk et al. 2020, Chandnani et al. 2023). While RSA could be accurately quantified in these studies, the 2D root growth is very far removed from environments found in the field. Other studies used restricted soil-based root systems, soil grown roots either in seedling cone systems, rhizoboxes or PVC pipes (8 cm diameter and 35 cm height). Images of these space restricted root systems were obtained in a 2D way either by a flatbed scanner or by a camera taking an image of the flat surface of the rhizobox (Prince et al. 2019, Seck et al. 2020, Mandozai et al. 2020). Finally, a GWAS was conducted on crown roots of field grown soybeans, providing trait associations of the uppermost root system parts at the end of a field season (Dhanpal et al. 2020). Overall, these studies did not address soybean root systems that grow unconstrained in a three-dimensional (3D) environment or quantify developmental traits such as growth rate or the angles of the developing taproot vs. lateral roots. Despite previous research efforts, the molecular mechanisms underlying RSA in soybean remain poorly understood, particularly regarding the coordinated roles of multiple genes that collectively drive significant phenotypic outcomes.

One promising avenue for prioritizing candidate genes found in GWAS is to employ gene regulatory networks (GRN). GRN have contributed to several phenotypic discoveries including developmental patterns in fruit flies and sea urchins, the circadian rhythm of plants, flowering time regulation, and plant responses to abiotic stress (Tarsis et al. 2022; Imaizumi 2010; Sun et al. 2022). The increasing volume of expression data has popularized gene co-expression network (GCN) approaches as a proxy for GRN. Several strategies have been developed to

construct GCNs that depict mRNA as nodes, co-expression relationships as edges, and co-expressed modules as connected components (Tantardini et al. 2019; Huynh-Thu et al. 2010; Cliff et al. 2019; Moerman et al. 2019; R. Zheng et al. 2019; Sun and Dinneny 2018). Weighted gene co-expression network analysis (WGCNA) has been applied to predict tissue-specific networks, identify networks related to lateral root and nodule formation, and identify the regulatory components of flooding tolerance (Jhan et al. 2023; Smita et al. 2020; Juexin Wang et al. 2019). GCN approaches leveraging RNA-seq data have been successfully applied across a range of crop species, including soybean (Azam et al. 2023; Y. Yao et al. 2023; Gao et al. 2018). These approaches enable the identification of functional modules and sub-networks defined by centrally connected hub genes, offering key insights into the regulatory architecture underlying plant development and other complex traits (Ko and Brandizzi 2020; X. Zhu, Duren, and Wong 2021).

Single-cell and single-nucleus RNA sequencing (scRNA-seq/snRNA-seq) provide powerful complementary approaches to resolve the functional roles of gene network components at cell-type resolution (Jia et al. 2022; Jagadeesh et al. 2022). In plants, these technologies have already uncovered novel developmental phenotypes, defined spatial and temporal patterns during biotic stress, facilitated comparisons of common cell-types in *Arabidopsis* and crops, and identified specialized cell-types and specific gene expression patterns (Dorrity et al. 2021; Shahan et al. 2022; J. Zhu et al. 2023; Apelt et al. 2022; Yılmaz et al. 2023; Song et al. 2020; Guillotin et al. 2023; Shahan, Nolan, and Benfey 2021). In soybeans, scRNA-Seq has also played a crucial role in unraveling the rhizobium-legume symbiosis by classifying major cell-types in both the root and root nodules (Liu et al. 2023).

Here we integrate GWAS, GCN, and snRNA-seq analysis to uncover the genetic regulation that underlies RSA in soybeans. In our efforts to provide an unrestricted root growth environment, we have developed a high-throughput 3D imaging system called **R**oot **A**rchitecture 3**D** **I**maging **Cyl**inder (*RADICYL*), which allows for the growth of seedlings unimpededly in cylinders filled with transparent gel media (Clark et al. 2011; Iyer-Pascuzzi et al. 2010). *RADICYL* provides advancements over previous methods in several aspects including size, hardware, the imaging camera, and system throughput. We developed a deep learning image segmentation approach that is capable of accurately quantifying RSA traits from the resulting images in a non-supervised manner. Employing these tools, we successfully quantified root traits in 371 diverse soybean varieties. Subsequently, we made use of publicly available whole genome resequencing data (Valliyodan et al. 2021) to conduct GWAS using multi-locus models (Jiabo Wang and Zhang 2021). We integrated GWAS results with a GCN constructed from publicly available bulk RNA-seq data, and a newly generated snRNA-seq atlas of the soybean root to refine candidate gene selection. This multi-layered approach enabled the prioritization of high-confidence GWAS candidate (GC) genes with cell-type–specific expression and network connectivity. We functionally validated two GC genes, RCE1 and NPR3, in *Arabidopsis*, demonstrating their conserved roles in regulating RSA traits across species. These genes exemplify promising targets for future genetic improvement of RSA, with potential applications in enhancing stress resilience, resource acquisition, and yield stability, all of which are critical for addressing the pressing challenges of climate change and global food security.

## Results

### *RADICYL*: A high-throughput phenotyping platform for screening Root System Architecture (RSA) traits

Root system architecture (RSA) is a complex trait that is most accurately quantified using imaging platforms capable of capturing roots in their natural three-dimensional (3D) configuration. We developed a high-throughput phenotyping platform, ***R**oot **A**rchitecture 3**D***

***I**maging **Cyl**inder* (*RADICYL*) to achieve the necessary throughput required for screening such traits in the context of natural variation studies. In developing *RADICYL* we used the principles laid out by a previously published imaging method that makes use of gel-filled cylinders within which rice roots could grow unimpededly for approximately two weeks (Clark et al. 2011; Iyer-Pascuzzi et al. 2010). We enhanced throughput by decreasing the quantity of the gel medium required and utilized smaller polystyrene containers, instead of expensive and heavy glass containers (Clark et al. 2011; Iyer-Pascuzzi et al. 2010). Using *RADICYL,* a single person can screen 75 cylinders per hour. Within this time, the user was able to acquire a rotational image series consisting of 72 images at 5º intervals, giving a total of 5,400 captured images (Supplementary Fig. 1).

We developed a deep-learning-based image processing pipeline to process the large number of images collected. In this process, the images were trimmed to a size of 990 x 860 pixels to center the plant within the cylinder and eliminate pixels outside of the cylinder area (Supplementary Fig. 2a). We then utilized a trained model obtained using the convolutional neural network (CNN)-based semantic segmentation architecture UNet++, to segment primary roots and lateral roots in each image. Following this, we skeletonized the roots in each segmented image and measured a total of 15 traits. The majority of the traits were directly computed based on the image (width, depth, convex hull area, biomass, total root length, primary root length, lateral root length, lateral root tip depth, primary root tip depth, all tip depth, vertical angles), while three were compound traits (SDx, SDy, and SDx/SDy) and one physical mass trait (dry root biomass) was measured after imaging on dried roots (Supplementary Fig. 2b, Supplementary Table 1). We also implemented an efficient quality control (QC) step to discard samples with poor germination or other quality issues and to achieve higher fidelity. Overall, using the *RADICYL* imaging setup and the image processing workflow described, we were able to capture high-quality data elucidating a wide range of RSA traits in soybeans.

**Early RSA traits cylinder-grown soybean seedlings are heritable and display notable natural variation**

We screened the 15 RSA parameters described above across a diverse set of 371 publicly available USDA soybean accessions using the *RADICYL* phenotyping platform (**Figure 1**). This collection of germplasm comprises a high level of genetic diversity with genotypes collected from more than 20 countries and across a range of maturity groups. Using the *RADICYL* system to evaluate the 15 traits across these genotypes uncovered a high level of phenotypic variation (Supplementary Fig. 3a, Supplementary Table 2). We calculated broad-sense heritability (BSH) from all the accessions to gain an understanding of how much of the phenotypic variation observed can be accounted for by genetic variation. BSH values ranged from 13% to 32%, with lateral root length showing the lowest BSH and dry biomass showing the highest. The identification of heritable variation for these traits is promising and emphasizes the inherent potential of the population to capture a substantial proportion of variability in RSA traits (Supplementary Table 3).

We performed a correlation analysis to explore the relationship between the early RSA traits screened and identified traits with the highest positive correlation to be convex hull area, biomass area, lateral root length, and total root length, which showed an $R^2$ of 0.9 (Supplementary Fig. 3b). The overall orientation of the root in the X direction (SDx) is strongly positive ($R^2 = 0.9$) correlated with the width of the root system. Similarly, the orientation in the Y direction (SDy) is positively correlated ($R^2 = 0.7$) with the depth of the root system. Early-stage root biomass was predominantly predicted by primary root length and lateral root length, and dry root biomass showed positive correlations ($R^2 = 0.4$-$0.5$) with both of these root length traits. Additionally, the dry root biomass displayed a strong positive correlation ($R^2 = 0.6$) with biomass

area, indicating that the root system captured in pixels is predictive of the actual physical mass of the early seedling stage root structures. The identification of relatively high positive correlations between traits highlights that these traits measure related aspects of the root and despite employing different measurement approaches.

We generated a Principal Component Analysis (PCA) plot based on the values of each trait across all accessions to further explore the relationships among computed traits (Figure 1a). The results demonstrated a clear separation of these traits along two distinct axes: PCA1 and PCA2, which collectively accounted for over 50.7% of the variance. The PCA separated the traits into mostly two clusters: the first cluster comprised traits including lateral root tip depth, primary root tip depth, all tip depth, primary root length, and depth, which we collectively referred to as “depth traits.” The second cluster contained traits such as width, lateral root length, biomass area, total root length, and convex hull, which we collectively referred to as “size and shape traits.” We next identified the accessions with the most significant phenotypic variability across the quantified traits by ranking the accessions in descending order of overall variation across all traits (Supplementary Fig. 4) and summarizing the top four accessions in a radial plot (Figure 1b). The accessions PI548540, PI548667, PI597464, and PI603458A underscored that the most significant differences in root architecture traits are associated with the number, length, and angle of the lateral roots (Figure 1c). Our data suggest that these accessions constitute the most variable subset within our dataset, highlighting the potential for identifying and selecting accessions using our methods.

Finally, to evaluate the reproducibility of quantified soybean root traits, we conducted an independent experiment using a subset of 24 soybean accessions that spanned the range of observed trait variation. Importantly, this follow-up experiment was carried out in October–November 2024, whereas the original experiment had been conducted in May–August 2021 (Supplementary Table 4). Both experiments took place in the same greenhouse but differed in temperature, humidity, daylight conditions, seed batch, materials, and experimenters--factors that introduce realistic environmental variation. We computed pairwise Pearson correlations for our 15 RSA and biomass-related traits to assess trait stability under these differing conditions. Several traits exhibited strong positive correlations across experiments (Supplementary Fig. 5), indicating robust genotypic control. Notably, dry root mass showed a high correlation ($r = 0.69$, $p < 0.001$), suggesting consistent biomass accumulation. Similarly, primary root length ($r = 0.61$, $p < 0.01$), biomass area ($r = 0.49$, $p < 0.05$), and total root length ($r = 0.41$, $p < 0.05$) demonstrated substantial cross-experiment reproducibility. Traits related to root system shape, such as convex hull area ($r = 0.37$), vertical angles ($r = 0.40$), and SDx_SDy, a measure of spatial dispersion ($r = 0.23$), were also moderately correlated. In contrast, traits reflecting absolute root depth, including depth, SDy, lateral root tip depth, and all tip depth, exhibited low, near-zero, or even negative correlations across experiments. Overall, these patterns were mirrored in the broad-sense heritability estimates (Supplementary Fig. 5). Traits that were well explained by genetic variation within a single experiment also tended to exhibit greater correlations across the two independent experiments. These findings suggest that while overall root system size and lateral spread are reproducible across experimental conditions in our system, absolute depth traits are more environmentally responsive or more challenging to quantify consistently.

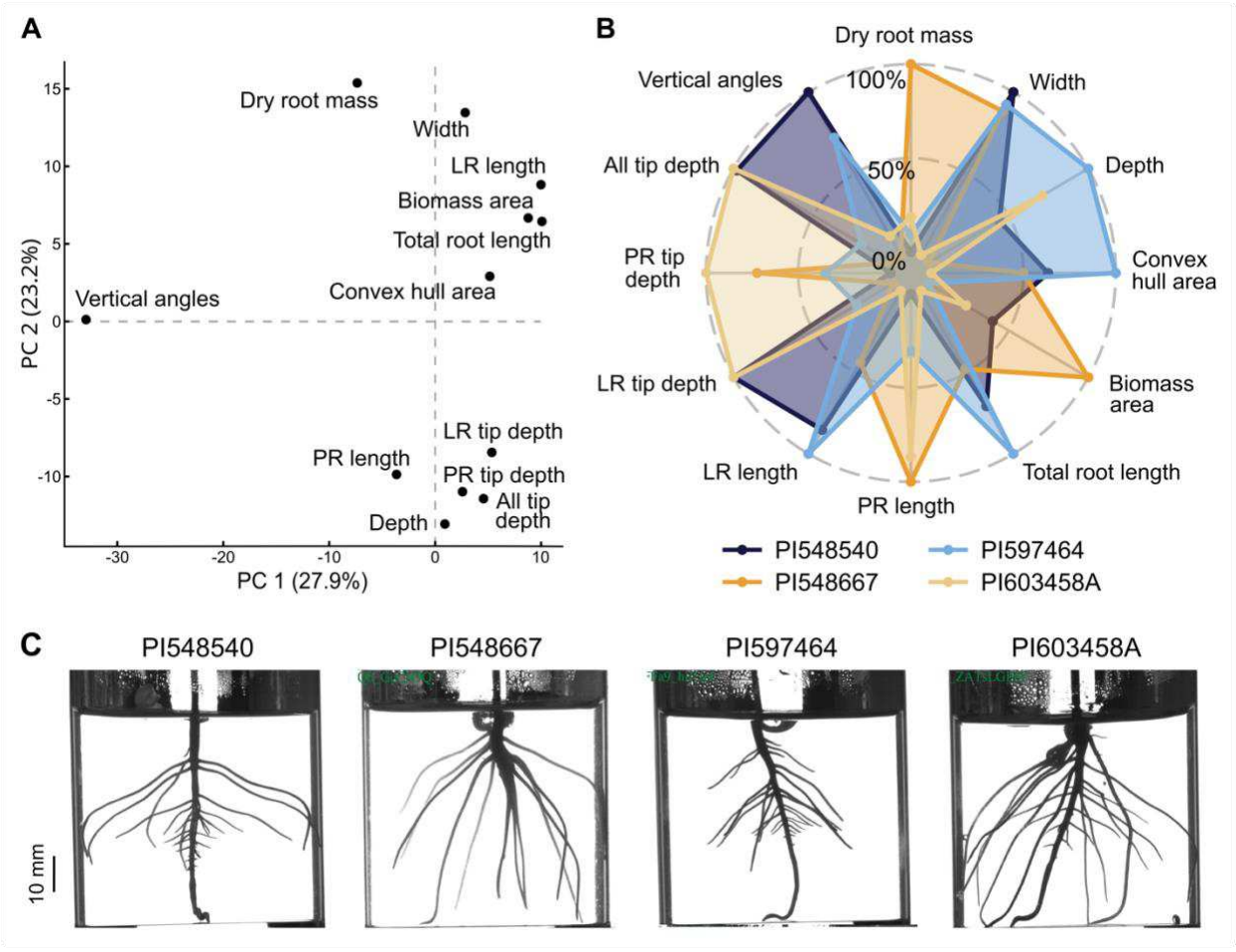



**Figure 1. *RADICYL* captures a substantial range of variation in root system architecture (RSA) through the characterization of twelve directly computed trait measurements.** A) Principal Component Analysis (PCA) plot of directly computed traits using phenotype data for all samples. B) A radial plot showcasing the top four accessions characterized by the highest cumulative variance across traits that exhibit the most significant deviations from the mean. C) Representative image of 6-day-old seedlings corresponding to the cylinder data presented in panel B, illustrating the phenotypic variation among accessions.

**Genome-wide association study (GWAS) pinpoints genetic associations and hotspots for soybean root architecture**

We set out to use a genome-wide association study (GWAS) approach to uncover the genetic component of trait variation observed across the 371 soybean genotypes (Figure 2; Supplementary Table 5). The samples for our study encompassed diverse accessions, including cultivars and landraces from China (222), USA (52), Korea (36), Japan (23) and Russia (11) (Figure 2a). Genotypes across the 371 phenotyped accessions were called from publicly available whole genome Illumina sequencing data, yielding 4,815,704 high-quality single nucleotide polymorphisms (SNPs). The distribution of SNPs across the genome varied from 0 to 14,774 per megabase (Mb) (Supplementary Fig. 6). A PCA of accession genotypes found 34.83% of variance was explained by two PCs, which distinguished a cluster of cultivar type accessions from landrace type accessions (Figure 2b). Notably, the top two PCs did not differentiate accessions by maturity group. Instead, the observed clustering reflected a

correlation between geographic origin and accession type, with most cultivars originating from the USA and most landraces from Asian countries (Supplementary Figure 7).

We leveraged the program fastSTRUCTURE to assess the influence of population structure on root system architecture (RSA) traits (Raj, Stephens, and Pritchard 2014), which revealed subpopulation clusters that broadly correspond to the five countries of origin represented by our soybean accessions (Supplementary Fig. 8a). In parallel, we performed a linkage disequilibrium (LD) analysis to estimate the physical distance over which SNPs could be linked to potential causal genes. Our analysis showed that LD decayed to an $r^2$ value of 0.2 at approximately 300 kb (Supplementary Fig. 8b), indicating long-range LD with SNP-trait associations potentially reflecting causal variants located several hundred kilobases away. This extended LD likely reflects the genetic composition of our panel, which consists primarily of soybean landraces. These traditional, locally adapted varieties have undergone long-term natural and farmer-driven selection, leading to reduced effective recombination and broader haplotype blocks (Figure 2b). Notably, soybean is recognized as an outlier among crop species due to its characteristically long-range LD, previously reported to range between 100 and 500 kb across various *G. max* populations (Hyten et al. 2007; Lam et al. 2010; Potapova et al. 2023; Valliyodan et al. 2021). Our findings are consistent with these earlier estimates, reinforcing the idea that LD structure in soybean imposes significant challenges for mapping causal variants in association studies.

A total of 30 model-RSA trait combinations were tested for statistical associations in our GWAS, which included 15 traits, each tested by Farm-CPU and BLINK models. In some cases, we observed near overlaps or exact matches of associated loci (Figure 2c). In total, our analysis yielded 54 Bonferroni corrected (alpha = 0.05; p-value < 1.03e-08) significant SNPs, of which 49 are unique loci distributed across 18 of the 20 soybean chromosomes (Figure 2d). QQ plots revealed no highly skewed p-values, suggesting that these results are reliable and not systematically biased by population structure (Supplementary Fig. 9). We identified candidate genes for each significant SNP, within a 300 kb window based on our estimated LD decay, limiting the search to 10 genes upstream and downstream. This resulted in a total of 633 GWAS candidates (GC) genes (Supplementary Table 5). We then incorporated relevant gene ontology (GO) terms associated with the GC genes and the description from the identified *Arabidopsis* ortholog. Functional annotation with *Arabidopsis* orthologs of the 633 putative GC gene candidates revealed that 487 of these genes are predicted to play roles in regulating root morphology and function, supporting their potential relevance to RSA traits (Supplementary Table 5).

Of the 49 unique significant SNPs, twelve fell within gene models: six in coding and five in non-coding DNA (Supplementary Table 6). Among these GC genes, seven have predicted *Arabidopsis* orthologs, while the function is unknown for the remaining genes. We examined their gene descriptions and revealed several genes with annotations related to regulating RSA. For example, Glyma.02G149100, which harbors a primary tip depth-associated SNP, Gm02:15770618, in the first coding region, is one of four soybean orthologs of the *Arabidopsis GLUTATHIONE PEROXIDASE 4* (*GPX4;* AT3G63080) (Passaia et al. 2014). Glyma.11G062900, one of eight soybean orthologs of *Arabidopsis CLATHRIN HEAVY CHAIN 1* (*CHC1;* AT3G11130), contains a SNP, Gm11:476025, in the 22nd coding region. The *Arabidopsis* mutant ortholog, *chc1*, demonstrated an increase in primary root length (Ormancey et al. 2020). Glyma.02G191500, one of two soybean orthologs of *Arabidopsis DIHYDROFOLATE SYNTHETASE FOLYLPOLYGLUTAMATE SYNTHETASE (DHFS-FPGS*) homolog C, contains a SNP, Gm02:38012571, within the 13th intron. Earlier research in *Arabidopsis* shows that maintaining folate levels in the leaves is required for maintaining plant metabolic homeostasis. Consequently, the *Arabidopsis* ortholog, *FOLYLPOLYGLUTAMATE*

*SYNTHETASE 2* (*FPGS2*) has been associated with shorter root length (Zhang et al. 2023; Mehrshahi et al. 2010).

We also identified the presence of pleiotropic loci in SNPs that were associated with multiple traits. Specifically, the biomass area and primary root tip depth were associated with the same significant SNP loci (Gm06:17855981 and Gm07:35512631 respectively) in both the FarmCPU and Blink models. Additionally, two SNPs (Gm09:22884486, Gm06:41838269) were significant in multiple model-trait analysis, including FarmCPU-biomass area, FarmCPU-convex hull area, and FarmCPU-total root length. This may suggest a causal relationship behind the high correlations observed among these RSA traits ($R^2$ = 0.9). Additionally, we identified a polymorphic gene, Glyma.06G250300, *ULTRAPETALA 2* (*ULT2*), which is an ortholog of *Arabidopsis ULT1* that maintains the root stem cell niche (Carles and Fletcher 2009; Ornelas-Ayala et al. 2020). It was associated with two SNPs (Gm06:41838269 and Gm06:42199048) and exhibited correlations with FarmCPU-convex hull area and FarmCPU-total root length respectively (Supplementary Table 7).

We identified 37 significant GWAS SNPs located outside of annotated gene models (Supplementary Table 7). Notably, complex traits such as total root length and primary root tip depth were associated with SNPs distributed across multiple chromosomes, suggesting a polygenic basis. Total root length was linked to loci on Chromosomes 2, 6, 7, 9, 10, 13, and 14, while primary tip depth was associated with SNPs on Chromosomes 2, 5, 10, 12, 15, 18, and 20. In contrast, several traits, including dry root mass, primary root length, and width, were each associated with SNPs localized to a single chromosome, suggesting simpler genetic architectures (Figure 2c). Additionally, we identified GWAS “hotspots” where multiple significant SNPs clustered within a 500 kb window, indicating regions of potential pleiotropic control. For example, loci on Chromosome 2 (~37 Mb) and Chromosome 6 (~41 Mb) harbored tightly linked SNPs associated with multiple traits, including depth, biomass area, convex hull area, and total root length (Supplementary Table 8). Overall the patterns of the significant SNPs that we observe are consistent with previous findings that RSA traits are polygenic (LaRue et al. 2022), and that long-distance LD in soybean (Hyten et al. 2007; Lam et al. 2010; Potapova et al. 2023; Valliyodan et al. 2021) may require additional approaches to GC candidate genes.

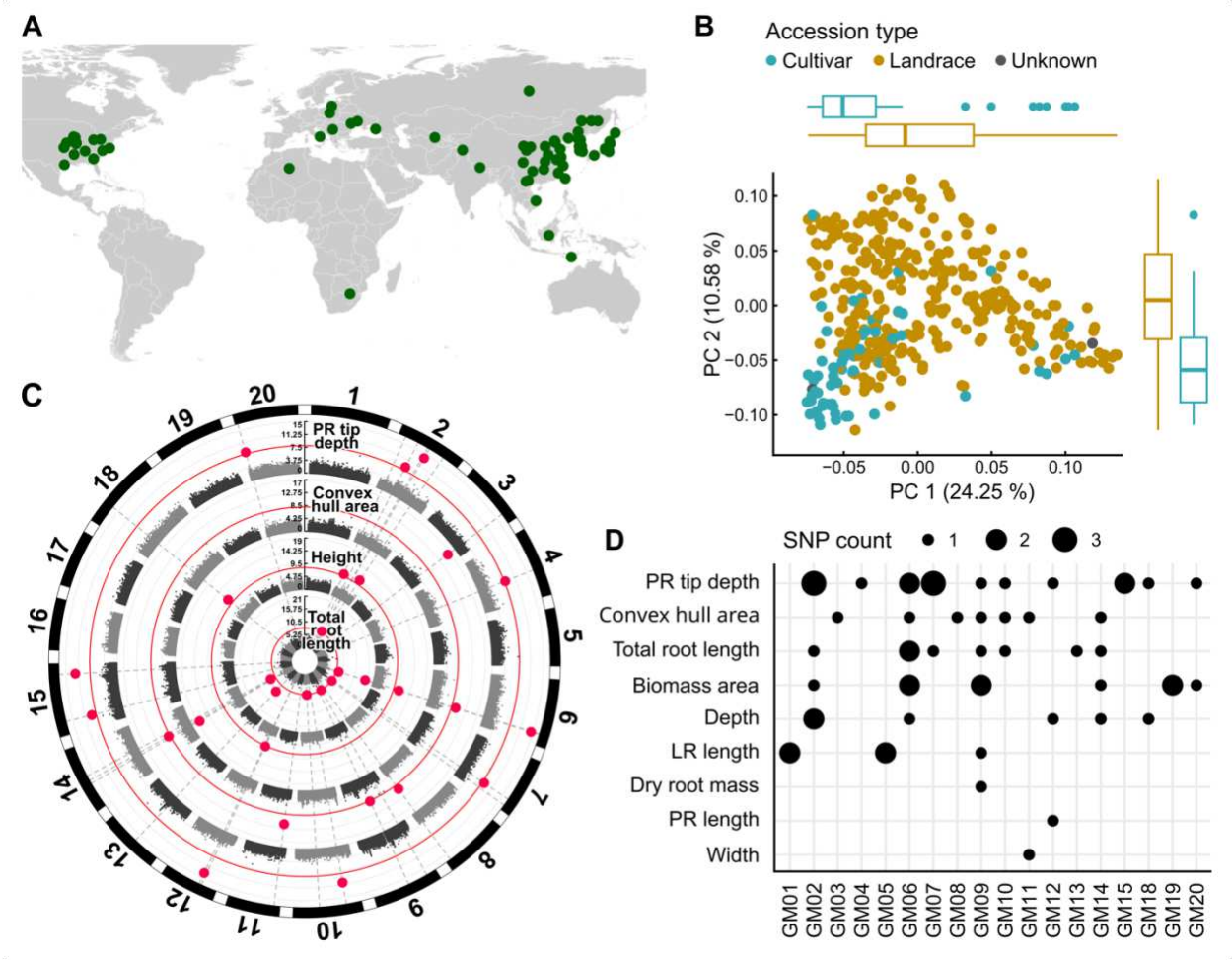



**Figure 2. Genome-Wide Association Study (GWAS) identified specific and significant Single Nucleotide Polymorphisms (SNPs) associated with RSA across a panel of 371 soybean accessions.** A) Soybean accessions included in this study originate from across the globe but with an emphasis on the USA and Asia, where soybean breeding is particularly prevalent . B) Principal Component Analysis (PCA) plot revealed population structure across soybean accessions. C) Manhattan plot displaying significant SNPs as red dots, using the FarmCPU model, associated with four traits from the inner to outer tracts: total root length, height, convex hull area, and primary root tip depth. D) Plot illustrating significant SNPs associated with specific traits, with larger dots indicating chromosomal regions housing a higher density of SNPs “hot spots”.

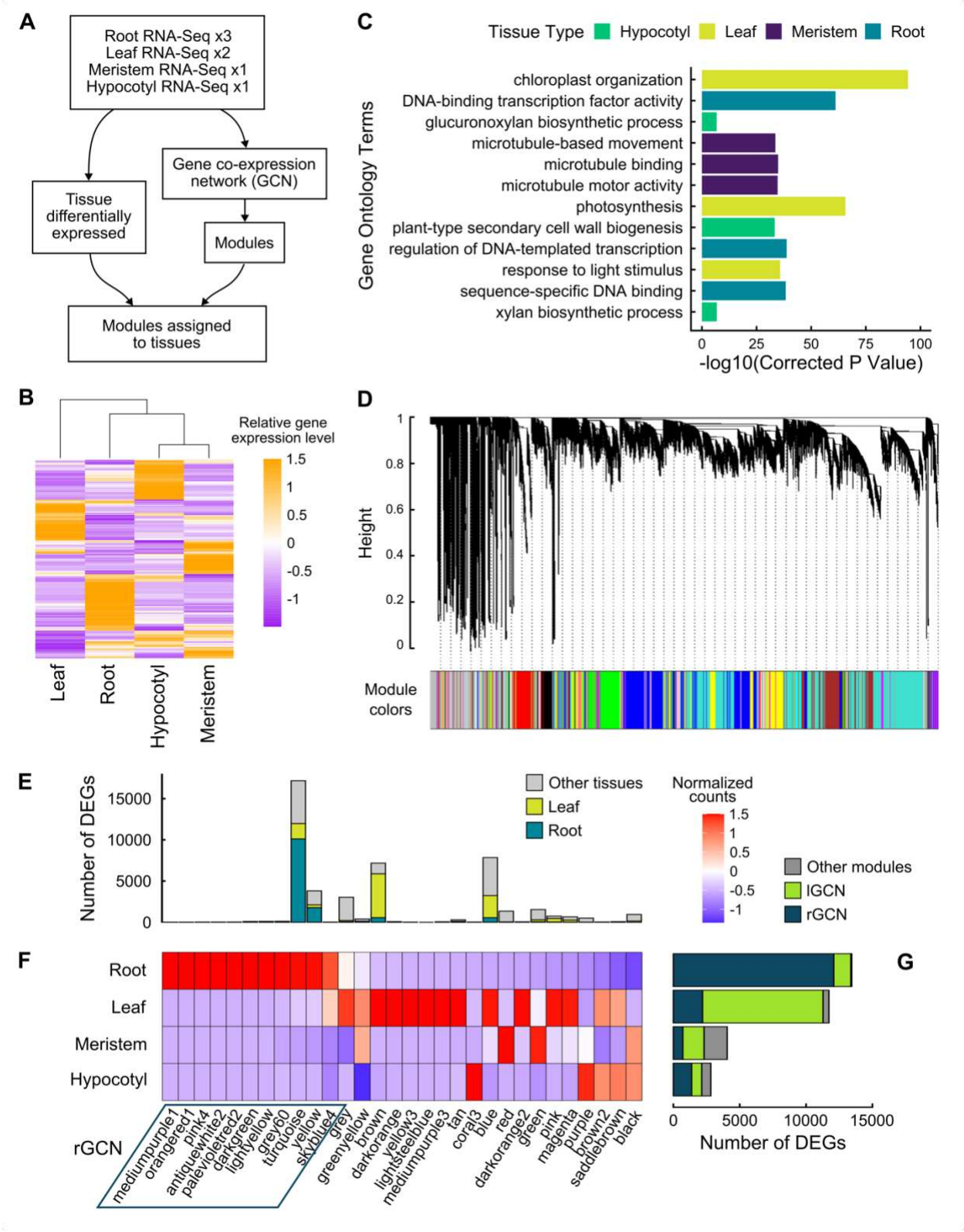
**Differential expression and co-expression network analysis uncovered root-enriched network modules**

Given the complex, polygenic architecture of RSA traits and the presence of long-range LD in soybean, we hypothesized that a gene co-expression network (GCN) approach would help uncover not only RSA-associated GCs, but also additional regulatory genes involved in RSA (Schaefer et al. 2018; M. Yao et al. 2020; Piya et al. 2023). Constructing a GCN would enable us to identify modules of co-expressed genes and assess potential functional interactions among known GCs. Importantly, this approach allows for the discovery of novel genes that, while not directly linked to significant SNPs, are consistently co-expressed with RSA-associated

genes, highlighting them as promising new targets for dissecting and engineering RSA traits in soybean. Therefore, we identified 71 publicly available Williams82 bulk RNA sequencing (RNA-seq) datasets from National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) spanning a wide range of tissues and developmental stages including 27 root, 18 leaf, 3 hypocotyl, and 3 shoot meristem tissues to carry out a weighted gene co-expression network analysis (WGCNA) (Supplementary Table 9).

We first validated our soybean atlas by identifying tissue-enriched differentially expressed genes (DEGs) (Figure 3a–b; Supplementary Figure 10). This validation step allowed us to identify WGCNA modules that were associated or entirely root-specific. We conducted GO enrichment analysis to assess the functional relevance of the DE gene sets, which revealed highly specific enrichment for known tissue-related biological processes. The 10,670 genes enriched in leaf tissue were significantly associated with GO terms including chloroplast organization (corrected $p = 4.82 \times 10^{-95}$), photosynthesis ($1.94 \times 10^{-66}$), response to light stimulus ($1.57 \times 10^{-36}$), thylakoid membrane organization ($4.42 \times 10^{-34}$), and chlorophyll-binding ($1.24 \times 10^{-30}$). The 2,721 shoot meristem-enriched genes showed strong enrichment for processes such as microtubule binding ($1.35 \times 10^{-35}$), microtubule motor activity ($2.16 \times 10^{-35}$), microtubule-based movement ($2.41 \times 10^{-34}$), and cell division ($1.40 \times 10^{-17}$). The 1,856 genes enriched in the hypocotyl were linked to plant-type secondary cell wall biogenesis ($4.89 \times 10^{-34}$), xylan biosynthetic process ($1.52 \times 10^{-7}$), and glucuronoxylan biosynthetic process ($1.79 \times 10^{-7}$) (Figure 3c). These findings confirm that the combination of DE analysis and GO enrichment reliably identifies tissue-specific gene activity. The resulting high-confidence, tissue-enriched gene sets serve as a foundational resource for constructing gene co-expression networks and probing the regulatory mechanisms underlying root system architecture and other complex traits in soybean.

WGCNA identified 99 distinct gene modules, each labeled by a unique color (Figure 3d). A hierarchical clustering analysis was used to separate samples and determine the optimal soft-thresholding power, ensuring the robustness of network construction (Supplementary Figure 11). Analysis of these modules revealed that 30 contained DEGs relevant to specific tissues (Figure 3e–g). The number of DEGs from each tissue was counted within each module to pinpoint modules most relevant to root biology, and modules were ranked based on the predominant tissue represented (Figure 3e–g; Supplementary Figure 12). The analysis identified eleven modules enriched for root-expressed genes, with the “turquoise” and “yellow” modules emerging as the most prominent. The turquoise module contained 10,116 root DEGs out of 17,175 total genes (59%), and the yellow module included 1,779 root DEGs among 3,805 genes (47%). Functional enrichment within the turquoise module revealed strong associations with regulatory processes, including GO terms such as metal ion binding (corrected $p = 3.14 \times 10^{-50}$), DNA-binding transcription factor activity ($1.18 \times 10^{-39}$), and protein serine/threonine kinase activity ($1.40 \times 10^{-27}$). The yellow module was characterized by enrichment in stress and defense-related functions, as indicated by terms such as response to chitin ($1.88 \times 10^{-62}$), cellular response to hypoxia ($2.98 \times 10^{-62}$), and response to wounding ($2.09 \times 10^{-22}$). The PyGNA geneset network analysis tool (Fanfani, Cassano, and Stracquadanio 2020) was adapted to further investigate the structure and connectivity within the GCN (Supplementary Figure 13). A subnetwork composed of genes from the eleven root-enriched modules was extracted and defined as the Root Gene Co-expression Network (rGCN) (Supplementary Figure 14). The rGCN maintained similar structural features to the full GCN, including component size and degree distributions, suggesting that it preserves the underlying complexity and biological relevance required to explore root-specific gene regulation.




425

426 **Figure 3. Differential gene expression and Weighted Gene Co-expression Network**
427 **Analysis (WGCNA) identified root gene co-expression modules.** A. Schematic of RNA-seq

datasets and analysis. B. Heatmap of relative gene expression levels across tissues (10% most variable genes). C. Top Gene Ontology (GO) terms for Differentially Expressed Genes (DEGs) across four distinct tissue types. D. Identification of co-expressed modules using Weighted Correlation Network Analysis (WGCNA). Different colors represent multiple co-expressed gene modules of varying sizes. Identification of gene co-expression modules via hierarchical average linkage clustering. The color row underneath the dendrogram shows the module assignment determined by the dynamic tree cut. E-G. Enrichment of WGCNA modules with DEGs. (E) Total genes and root or leaf DEGs in each module. (F) Tissue enrichment of each module. (G) Total number of DEGs for each tissue.

**Network analysis prioritized GWAS candidates and specific root-enriched network modules for subsequent investigation**

In our analysis, we refined the initial 633 GC genes by focusing on only those differentially expressed in root compared to those differentially expressed in leaf, hypocotyl, shoot meristem or across several tissue types (Figure 4a). This subset of GC DEGs (DE-GC) was distributed across 13 of the 99 WGCNA modules (Figure 4b-d). We identified 142 root DE-GCs, which we termed rDE-GC, across 40 SNPs (Supplementary Table 10). Similar to the overall set of rGCN (Figure 3f), the rDE-GC were prominently concentrated in the turquoise and yellow modules (Figure 4c).

Next, we explored the significance of rDE-GC and other DE-GC enrichment in modules by investigating their connection to network topology. Our goal was to determine whether these genes collaboratively functioned within specific modules, and if any of these modules exhibited tissue specificity. We employed the Topology Internal Degree (TID) test of PyGNA to quantify interconnectedness of gene sets, to assess whether the average internal degree of a gene set was greater than expected by chance. The internal degree statistic represents the average number of edges shared by genes within the set, indicating their connectivity within the network. Upon analyzing all GC genes collectively, we observed that the set was not topologically significant ($P=0.07$), likely due to the distribution of GC genes across multiple modules within the GCN. However, when we focused on the expression of tissue-specific DE-GC in the root, leaf, hypocotyl, and shoot meristem, the results were topologically significant ($P<0.008$). This indicated that the subsetted genes formed cohesive clusters within distinct modules, indicating their connectivity to specific modules within the GCN (Figure 4e). Our findings demonstrated the feasibility of identifying specific tissue-associated modules within the broader context of the GCN by discerning the connections of DE-GC within these modules. Additionally, we delved into the rDE-GC set across nine traits. Notably, the turquoise and yellow modules of the rGCN stood out as the most predominant, containing GC associated with nine and six traits respectively. Our observation suggested a higher level of complexity and interconnectedness within these modules (Figure 4f). Conversely, non-rGCN modules contained a minimal number of rDE-GC.

Building on our findings, we implemented an improved filtering strategy aimed at narrowing down our gene candidates. This approach was designed to minimize potential false positives and negatives from our GWAS analysis. The specific steps of this filtering process are visually depicted in a Venn diagram, illustrating the prioritization of GC genes (Figure 4g). This refined filtering involved narrowing down the pool of rDE-GC to those specifically present in a rGCN (Supplementary Table 11). Through this filtering process, we identified 131 genes meeting these criteria and referred to them as Genes Related to Identified Traits (GRIT).

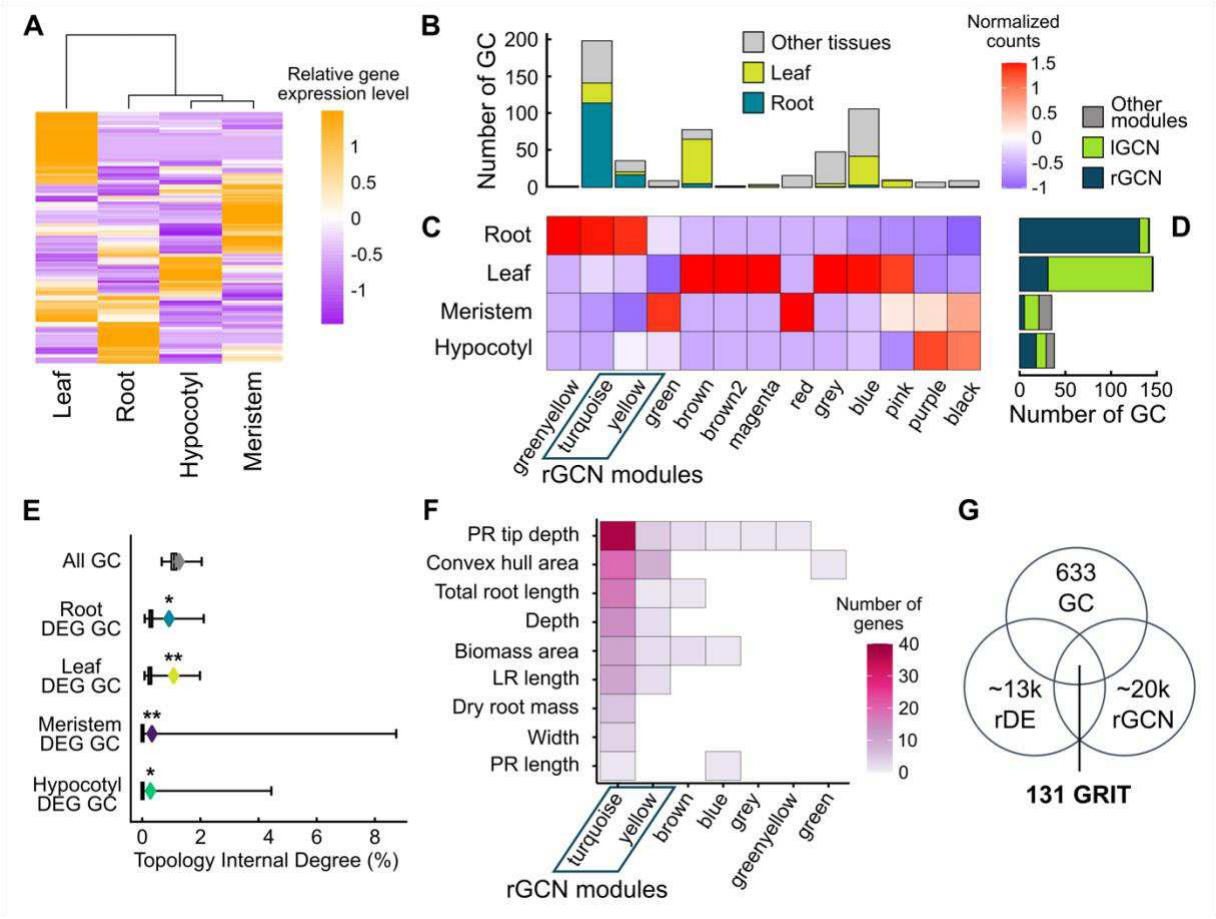



**Figure 4. A high number of differentially expressed GWAS candidates (DE-GC) are concentrated in the turquoise subnetwork.** A. Heatmap of relative gene expression level across tissues (GWAS candidates). B-D. Module enrichment of GWAS candidates. (B) Total GC and root or leaf DE-GC (rDE-GC and lDE-GC, respectively) in each module. (C) Tissue enrichment of each module. D. Total number of DE-GC for each tissue, with membership in rDE-GC and lDE-GC modules are shown. E. Network topology internal degree test (%) tested for potential subnetworks between genesets and co-expression modules identified using WGCNA. "All GC" encompasses all GWAS candidates. DE-GC indicates differentially expressed GWAS candidates from root, shoot meristem, leaf or hypocotyl tissues. A significant association (P< 0.5) is indicated by (*). F. Heatmap of per-trait GC membership across GCN modules. Color bar indicates the number of genes. G. Venn diagram illustrating the process of identifying genes associated with traits (GRIT) genes. The filtering process involves selecting GC genes, root genes that are DEGs for root vs. other tissues and present in a root-enriched gene coexpression network (rGCN).

**Exploring cell specificity of RSA trait regulation through single nuclei sequencing**

Single-nuclei and single cell RNA sequencing (snRNA-seq/scRNA-seq) in plants has become a powerful technique to identify specific cell-types in complex tissues (D. Zheng et al. 2023; Bawa et al. 2022). Gene expression data at this high level of resolution can provide valuable insights into gene function and also be used to guide decisions when aiming to develop future crops with precise alterations in gene activity. We performed snRNA-seq on six-day-old whole roots of

Williams82 soybean seedlings to gain further insight into the candidate genes identified in this study and to deepen our understanding of the soybean root expression landscape. After filtering for low quality nuclei, we obtained a dataset comprising gene expression data across 17,636 high quality nuclei capturing the expression of 47,095 transcripts. Following dimensional reduction and uMAP clustering, we identified 10 main clusters of nuclei with significantly unique transcriptional landscapes (Figure 5a).

Due to the absence of specific markers for the majority of cell types in soybean, we employed a dual-strategy to annotate cell clusters within the Soybean Cell Atlas. Initially, a subset of clusters was classified utilizing established marker genes derived from recent studies (Liu et al. 2023); these clusters were annotated with the most likely cell-type identities compared with *Arabidopsis* (Supplementary Fig. 15; Supplementary Table 12). This annotation delineated several key cellular structures, including the metaphloem (cluster 12), cortex (clusters 0, 3, 6, and 9), epidermis (clusters 7 and 8), vascular bundle (cluster 11), and the xylem pole pericycle (cluster 2). Subsequently, we leveraged orthologs of well-characterized *Arabidopsis* marker genes to annotate the remaining cell clusters, which comprised the pericycle (cluster 1), xylem (cluster 10), and endodermis (cluster 5) (Supplementary Fig. 16). We generated cell-type specific gene sets by ranking genes according to variance in average gene expression across cell-types, selecting the top 10% most variable genes, and assigning each to 1 or 2 cell-type groups where their normalized expression value exceeded a constant threshold (Figure 5b, methods). The gene sets included 465 cortex-specific genes, 768 pericycle-specific genes, 410 vascular bundle and xylem pole pericycle-specific genes, 427 stele-specific genes, 672 endodermis-specific genes, 751 epidermis-specific genes, 886 xylem-specific genes, 840 vascular bundle-specific genes, 759 metaphloem specific genes, and 722 cluster13 specific genes (Supplementary Table 13).

Analysis of the snRNA-seq dataset revealed that 117 out of the 131 GRIT genes were expressed (Supplementary Table 14). All downstream analyses focused on this expressed subset. Within these 117 genes, several candidates were identified with putative roles in root development, based on functional annotations of their *Arabidopsis* orthologs. For example, Glyma.19G076800, orthologous to AT5G40780 in Arabidopsis, encodes *LYSINE HISTIDINE TRANSPORTER 1* (*LHT1*). This gene emerged as a candidate associated with root biomass, was found to be exclusively expressed in vascular bundles and has been shown to be crucial for amino acid uptake at concentrations typically found in soils (Svennerstam et al. 2011). Another gene, Glyma.06G249700, shares homology with AT1G54890 and encodes a *LATE EMBRYOGENESIS ABUNDANT* (*LEA*) gene. This gene was linked to total root length, displayed ubiquitous expression across all root cell types, and may play a role in stress-related root development (Battaglia and Covarrubias 2013). Additionally, Glyma.06G230300, an ortholog of AT1G16890, encodes *UBIQUITIN-CONJUGATING ENZYME* (*UBC36/UBC13B*). This gene was identified as a candidate for primary root tip depth, exhibited broad expression across cell types, and is known to be involved in root developmental responses to iron deficiency in *Arabidopsis* (Li and Schmidt, 2010). These findings highlight specific GRIT genes with strong functional relevance to root traits and provide targets for further investigation into the genetic regulation of root system architecture.

We next proceeded to test for associations between specific cell-types and the networks where GRIT genes might be located. For this, we employed the Random Walk with Restart test from PyGNA to test for topological association of the GRIT gene sets and the cell-type gene sets within the context of the rGCN. Our analysis revealed several notable associations across different traits and cell-types (Supplementary Fig. 17). Of the different associations, we found that lateral root length exhibited the highest number of associations across multiple cell-types.

Furthermore, the root width trait demonstrated a significant association with the endodermis (Figure 5c, Supplementary Fig. 18). This finding suggested that multiple cell-types may govern lateral root development while the endodermis may play a dominant role in governing root width. Variations in lateral root length and root width indicate that root traits might be regulated by multiple or a single cell-specific subnetwork and our analysis methods may potentially be instrumental in elucidating the cell-type specific genes governing these traits.

Additionally, our findings showed an exclusive association between the root metaphloem cell-type and lateral root length (Figure 5d). We determined that the metaphloem cell-type specific GRN was the only cell-type specific GRN associated with lateral root length with no significant association with other traits (Figure 5c). This suggested a potential metaphloem-specific role in influencing RSA through lateral root length (Rodriguez-Villalon 2016). Our findings emphasized that our approach is useful in unraveling the complex regulatory pathways that govern RSA (Supplementary Fig. 19). The finer resolution provided by snRNA-Seq allowed us to identify potential cell-type-specific regulation of RSA in soybean. Further analyzing genes within cell-specific subnetworks could provide valuable insights into the molecular mechanisms governing variations in root width and lateral root length variation in soybean.

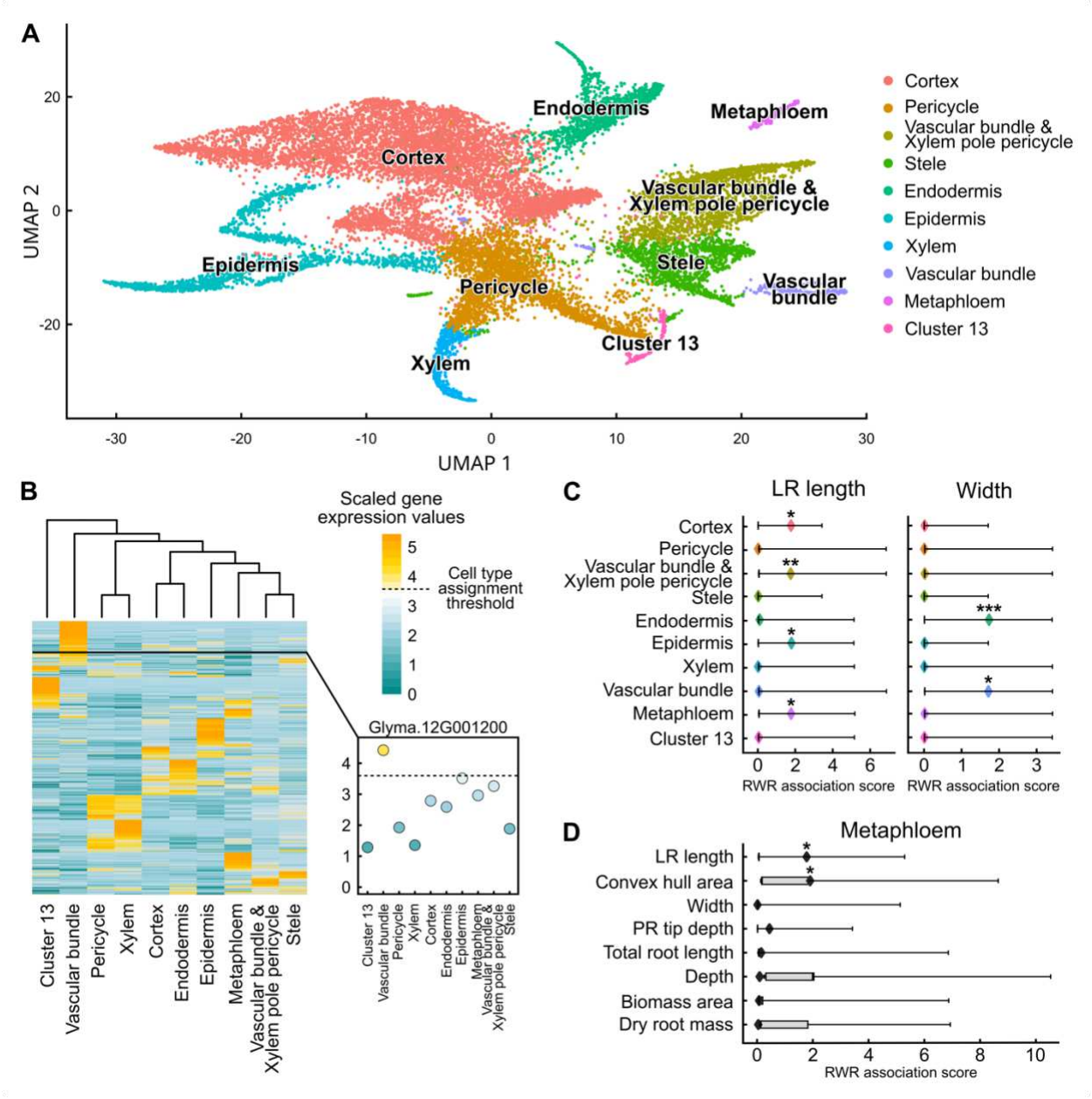



**Figure 5. Single nuclei RNA-sequencing (snRNA-seq) enhances the resolution for spatial expression patterns of GWAS gene candidates, revealing insights into their cellular localization and potential functions.** A. Uniform Manifold Approximation and Projection (UMAP) visualization illustrating distinct single-cell clusters, each representing a different cell-type. B. Heatmap depicting the expression levels of the top 10% most variable genes identified from a single-cell experiment, highlighting genes specific to different cell-types. Scaled expression values are shown. C. Boxplot of Random Walk with Restart (RWR) association test from PyGNA comparing GRIT genes separated by GWAS trait to the genes expressed from snRNA enriched in the metaphloem cell-type. D. Boxplot of RWR association test comparing GRIT genes for Lateral root length and Width traits to cell-type genesets. For both C and D, Boxes show null distributions, with whiskers showing the full range. Diamonds indicate observed values. * = p-value <0.01, ** = p-value <0.005, *** = p-value <0.001.

## Exploring gene regulatory networks at the single cell level highlights a key role for endodermis and metaphloem-expressed genes in regulating RSA in soybean

Building on the identification of subnetworks associated with specific cell types, we investigated whether the genes within these modules exhibited expression patterns consistent with functional roles in root development. We began with the endodermis–root width-associated subnetwork, one of the most significant modules within the broader turquoise co-expression network. The endodermis, which encases the vascular tissues, acts as a selective barrier that regulates the transport of water-soluble ions and shields the root from external environmental stress. Within this subnetwork, we identified two GRIT genes, Glyma.11G176302 and Glyma.11G176301, both located near the GWAS-significant SNP on Chromosome 11:31034808 and associated with variation in root width. Glyma.11G176302 is the predicted ortholog of *Arabidopsis UMAMIT41* (AT3G28050), a member of the *USUALLY MULTIPLE ACIDS MOVE IN AND OUT TRANSPORTERS 41* family (Chengsong Zhao et al. 2021), while Glyma.11G176301 is homologous to a leucine-rich repeat protein kinase (AT2G26730). Both genes are implicated in intercellular transport and signaling, processes that may underlie their role in shaping root system architecture.

We evaluated the potential roles of candidate genes within the endodermis-associated subnetwork by analyzing their connectivity in the root-specific gene co-expression network (rGCN) (Figure 6a, Supplementary Table 15). This analysis measured the number of co-expression links (edges) each gene shared with others in the network. *GmUMAMIT41* emerged as the most highly connected gene, with a degree of 4, followed by Glyma.11G176301, which had a degree of 3. These high connectivity scores suggest central regulatory roles within the subnetwork. Three additional GRIT genes also demonstrated notable co-expression patterns: Glyma.09G099500 (degree = 2), the predicted ortholog of *Arabidopsis METAL-TOLERANCE PROTEIN 10* (*MTP10*; AT1G16310), a cation efflux transporter (Ge et al. 2022); Glyma.20G058500 (degree = 2), corresponding to *WSS1/SPRTN-TYPE REPAIR PROTEASE B* (*WSS1B*; AT5G35690), a DNA repair-related protein (Enderle et al. 2019); Glyma.15G127200 (degree = 3), the ortholog of *NONEXPRESSER OF PATHOGEN RELATED GENES 3* (*NPR3*; AT5G45110), that functions as a salicylic acid (SA)-binding adaptor protein that promotes degradation of the immune coactivator *NPR1*, thereby acting as a negative regulator of SA-mediated immune responses in *Arabidopsis* (Fu et al. 2012; Kong et al. 2020; Ding et al. 2018). Gene expression patterns from our snRNA-seq data revealed that *GmUMAMIT41* and *GmMTP10* show peak expression in the endodermis, reinforcing the significance of this cell type in the regulation of root width (Figure 6b). These findings indicate that the endodermis subnetwork may influence root system architecture through the coordinated activity of transporters, signaling molecules, and stress-responsive genes, identifying a promising set of candidates for future functional validation.

The metaphloem-associated subnetwork emerged as another key module, revealing a strong link between the metaphloem cell type and GRIT genes associated with lateral root length. As a critical tissue responsible for transporting sugars and solutes to the root meristem the metaphloem plays an essential role in supporting root development (Graeff and Hardtke 2021; Hardtke 2023; Rodriguez-Villalon 2016). This subnetwork includes several genes co-expressed with metaphloem-enriched genes, encompassing seven genes associated with lateral root length, two with convex hull area, and two with primary tip depth (Figure 6c, Supplementary Table 16). Among these, Glyma.11G064000, the predicted ortholog of *Arabidopsis RUB1 CONJUGATING ENZYME 1* (*RCE1*; AT4G36800), showed the highest connectivity within the subnetwork (*degree = 5*). *RCE1* functions in auxin signaling, and its disruption in *Arabidopsis* leads to morphological defects resembling those of auxin-resistant mutants (Dharmasiri et al.

2003; Pacurar et al. 2014; del Pozo and Estelle 1999). *GmRCE1* is co-expressed with four additional genes related to lateral root length: Glyma.09G070700 (*degree = 2*), the predicted ortholog of *5-METHYLTHIORIBOSE-1-PHOSPHATE ISOMERASE* (*MTI1*; AT2G05830) (Méteignier et al. 2022); Glyma.01G074600 (*degree = 3*), corresponding to *UDP-GLUCOSE-DEPENDENT GLYCOSYLTRANSFERASE 72B1* (*UGT72B1*; AT4G01070) (Speeckaert et al. 2022); Glyma.08G322600 (*degree = 1*), the predicted ortholog of *SYNTAXIN OF PLANTS 51* (*SYP51*; AT1G16240) (C. Luo, Shi, and Xiang 2022); and Glyma.09G069700 (*degree = 3*), the ortholog of *FH INTERACTING PROTEIN 1* (*FIP1*; AT2G06005) (Geraldo et al. 2009). These connections suggest that *GmRCE1* and its associated genes may collectively regulate lateral root development through processes linked to auxin signaling, vesicle trafficking, and metabolite processing within the metaphloem, highlighting this subnetwork as a strong candidate for further functional validation.

The genes with the highest connectivity in the metaphloem-associated subnetwork were not exclusively linked to lateral root length or metaphloem-specific expression. Once again, the SA receptor *GmNPR3* (Glyma.15G127200) exhibited the highest degree of connectivity (degree = 6) in this network. Notably, *GmNPR3* was identified in the endodermis-associated subnetwork, where it was also co-expressed with genes influencing root width, suggesting a broader regulatory role across cell types. In the metaphloem network, *GmNPR3* connects to three endodermis-associated genes and five genes related to lateral root length, highlighting its potential as a key integrator of developmental signals. Gene expression analyses revealed that many genes in the metaphloem subnetwork were expressed across multiple root cell types, rather than being restricted to the metaphloem. Several co-expressed genes did not display peak expression in the metaphloem (Figure 6d), suggesting possible cross-cell-type functional coordination. Given the metaphloem's role in long-distance transport of signaling molecules and metabolites (Hardtke 2023), it is plausible that genes expressed in non-metaphloem tissues may influence lateral root development by acting through the metaphloem as a conduit. These findings support the idea that RSA traits, such as lateral root length, may be shaped by a complex interplay of signals between cell types.

We expanded our network analysis to gain deeper insight into the strong associations between lateral root length and multiple root cell types identified in earlier results. This broader analysis incorporated all significantly associated cell types, including the metaphloem, vascular bundle, xylem pole pericycle, cortex, and epidermis (Supplementary Fig. 20). Through this expanded framework, we prioritized 26 GRIT genes that were robustly associated not only with lateral root length but also with other root architectural traits such as convex hull area, primary tip depth, biomass area, total depth, and total root length (Supplementary Table 17). Among the top candidates, *GmNPR3* once again stood out, exhibiting the highest degree of network connectivity (degree = 16). This reinforces its central role in disease resistance, salicylic acid (SA) signaling, and auxin-related pathways that contribute to RSA regulation. Additionally, *GmRCE1*, previously identified in the metaphloem-associated subnetwork, ranked sixth in this analysis with a degree of 9, further supporting its relevance across multiple cell types and traits. These highly connected genes were embedded in subnetworks enriched for trait-associated links, offering a more integrated view of the regulatory landscape underlying root development. The cell-type-specific context provided by this analysis enhances our understanding of how these genes influence RSA. Overall, the results underscore the power of multi-cell-type network integration in dissecting complex traits like lateral root length and emphasize the need to assess candidate genes within their full developmental and regulatory framework.

**_NPR3_ and _RCE1_ are regulators of RSA in _Arabidopsis_**

Given the identification of *NPR3* and *RCE1* as potential regulators of RSA, we examined whether loss-of-function mutations in these genes impact RSA traits using the *RADICYL* assay. The role of *RCE1* in root development is well characterized. Previous studies have demonstrated that *rce1-1* and *rce1-2* mutants exhibit shortened primary roots and fewer lateral roots, phenotypes attributed to impaired auxin signaling and reduced auxin sensitivity (Dharmasiri et al. 2003; Pacurar et al. 2014; del Pozo and Estelle 1999). In contrast, *NPR3* has primarily been studied in the context of SA-mediated immune responses (Fu et al. 2012; Ding et al. 2018). However, emerging evidence suggests a broader role in root development. Notably, *npr3-1/npr4-3* double mutants display a short-root phenotype, and *npr1-1/npr3-1/npr4-3* triple mutants exhibit reduced lateral root emergence, despite normal lateral root development in the double mutant (Kong et al. 2020). *NPR3* may influence RSA by modulating crosstalk between SA and auxin signaling. Notably, *NPR3* may act antagonistically to *AUXIN RESPONSE FACTOR 7* (*ARF7*), a key regulator of lateral root development, by suppressing *ARF7*-mediated auxin responses during SA-induced immune signaling (Kong et al. 2020). Supporting this model, we also identified *GmARF7* as a GC gene for total root length (Supplementary Table 5), in line with the well-documented role of auxin signaling in shaping RSA (Perianez-Rodriguez et al. 2021).

In the *RADICYL* assay, the *rce1-1* mutant exhibited significant RSA alterations ($P < 0.05$) across eleven traits, with the most pronounced effect being a marked reduction in convex hull area, reflecting a narrower and more compact root system compared to wild-type (Col-0) plants (Figure 6E–F; Supplementary Fig. 21). These phenotypes are consistent with our soybean GWAS results, which identified *RCE1* as a top candidate gene associated with one of the most significant SNPs on chromosome 11 (Gm11_4760258) for convex hull area. Similarly, the *npr3-1* mutant showed significant RSA alterations ($P < 0.05$) across six traits, most notably characterized by shortened lateral roots compared to wild-type (Figure 6G–H; Supplementary Fig. 22). These observations are in line with our GWAS findings that linked *GmNPR3* to SNP Gm15_10024740, which was significantly associated with root tip depth. Taken together, these results provide functional evidence that both *RCE1* and *NPR3* play important roles in shaping RSA. Their involvement across multiple traits and cell types, and indirect impact on auxin signaling, highlights them as high-priority targets for future crop improvement efforts, particularly for enhancing RSA in soybean and potentially other agriculturally important species.

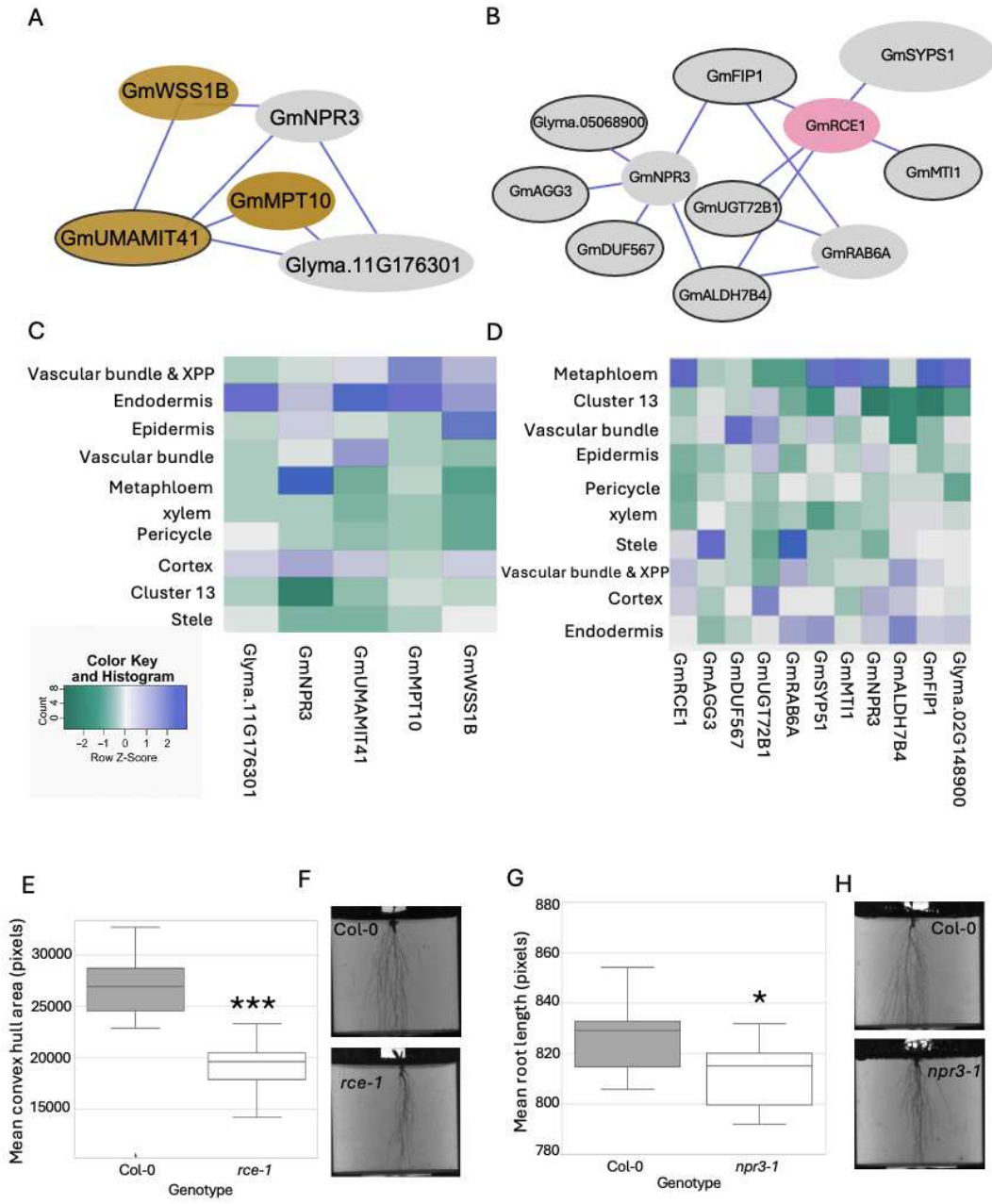



**Figure 6. Identification of key gene co-expression network hubs for root width and lateral root development.** A. Subnetwork highlighting GRIT genes (grey) associated with endodermis-specific GRIT genes in soybean (green) B. Subnetwork highlighting GRIT genes (grey) associated with metaphloem-specific GRIT genes in soybean (magenta) C-D. Heatmap depicting the average expression levels per cell for single-cell expression candidates. E-F. Root phenotypes for the Arabidopsis rce-1 mutant compared to its wild type (wt) control. G-H. Root phenotypes for the Arabidopsis npr-3 mutant compared to its wild type (wt) control.

## Discussion

We developed Root Architecture 3D Cylinder (*RADICYL*), a novel high-throughput phenotyping platform that enables non-destructive, high-resolution assessment of root system architecture (RSA) across a diverse panel of 371 soybean accessions. *RADICYL* represents a significant advancement over existing methods by streamlining trait extraction and capturing complex spatial features of root systems at a larger scale. Leveraging this platform, we conducted a genome-wide association study (GWAS) of 15 RSA traits, including 12 directly measured and 3 derived traits, resulting in the identification of 54 significant SNPs supported by at least one GWAS model across nine distinct traits. We integrated gene co-expression network (GCN) analysis and filtered 633 genes based on their differential expression in root tissues (rDE) and network connectivity (rGCN) to refine candidate genes from these loci. This integrative framework prioritized 131 Genes Related to Identified Traits (GRIT), 117 of which were expressed in our single-nucleus RNA sequencing (snRNA-seq) dataset. Among these, *GmRCE1* (*RUB1 CONJUGATING ENZYME 1*) and *GmNPR3* (*NONEXPRESSER OF PATHOGEN RELATED GENES 3*) emerged as key candidates associated with root width and lateral root length, respectively, and showed cell-type-specific expression in the endodermis and metaphloem. Functional validation in *Arabidopsis* revealed that *rce1-1* and *npr3-1* mutants exhibited RSA phenotypes consistent with those observed in our *RADICYL* assay, supporting the evolutionarily conserved role of these genes in root development. Collectively, our findings highlight the power of combining 3D phenotyping with multi-omic integration to identify the genetic and cellular basis of complex root traits and to advance trait dissection in soybean and other crop species.

Previous phenotyping methods have struggled to fully capture the complexity of root growth due to limitations in spatial and temporal resolution. In field settings, variable environmental conditions and reliance on destructive sampling further complicate accurate root phenotyping. Two-dimensional systems also fall short, as they fail to represent the natural three-dimensional (3D) growth patterns of RSA. In contrast, *RADICYL* enables non-destructive, continuous observation of root development without constraining natural growth, an approach previously applied only to rice roots, and even then at limited scale and resolution (Iyer-Pascuzzi 2010, Clark 2011). The reproducibility of key traits such as dry root mass, primary root length, and biomass area across two independent experiments highlights *RADICYL's* ability to capture stable, genetically regulated aspects of RSA. However, the relatively low consistency observed in absolute depth-related traits points to a limitation: while the method offers detailed spatial quantification, depth measurements may be more sensitive to environmental variation or inconsistencies in image processing and segmentation. These findings underscore both the potential and current limitations of *RADICYL* and suggest that certain traits may benefit from further refinement in methodology.

Soybean stands out among crop species due to its exceptionally long-range linkage disequilibrium (LD). Our analysis revealed LD extending to approximately 300 kb, aligning with previous estimates ranging from 100 to 500 kb in *G. max* populations (Hyten et al. 2007; Lam et al. 2010; Potapova et al. 2023; Valliyodan et al. 2021). This pattern reflects the effects of self-pollination, domestication bottlenecks, and breeding practices, which reduce genetic recombination and elevate LD across the genome. Our study introduced an integrated multi-omic framework combining snRNA-seq, GWAS, and GCN analysis to improve candidate gene prioritization in the context of extended LD. Traditional proximity-based approaches often misidentify causal genes in soybean due to broad LD blocks. Using our framework, we instead leveraged cell-type-specific expression and network connectivity to functionally refine GWAS signals with higher biological accuracy. This approach revealed strong associations between root width and the endodermis, as well as lateral root length and the metaphloem. Our framework enabled functional fine-mapping, improving SNP-to-gene resolution by integrating

genetic associations with gene expression and interaction data. GWAS identified trait-linked loci, GCN analysis mapped co-regulatory networks, and snRNA-seq contextualized gene function within specific cell types. These results underscore the importance of cell-type-aware analysis for dissecting complex traits in species like soybean, where long-range LD can obscure causal variation. The combined insights offer a path forward for more precise gene discovery, trait dissection, and crop improvement in soybean and other plant species.

In comparison to earlier GWAS studies, our analysis aligns with past sample sets, typically involving several hundred samples ranging from 137 to 397 varieties (Salim et al. 2021; Dhanapal et al. 2020; Seck, Torkamaneh, and Belzile 2020; Prince et al. 2019; Kim et al. 2023; Chandnani et al. 2023; Falk et al. 2020). These studies measured between seven and thirteen traits and identified two to 70 candidate genes, spanning a wide range of annotations (Salim et al. 2021; Dhanapal et al. 2020; Seck, Torkamaneh, and Belzile 2020; Prince et al. 2019; Kim et al. 2023; Chandnani et al. 2023; Falk et al. 2020). Our analysis identified overlaps with two specific loci adding to the credibility of the discovered associations. We found that Gm18:51895181 from our study and associated with primary tip depth, is located 22,650 bp from one of the candidate SNPs: Gm18:51917831, which was associated with total root volume (Chandnani et al. 2023). Additionally, we found that a total root length SNP in our study (Gm09:7681193) is near (~11kb) two candidates for the trait “number of forks” (NF) in (Kim et al. 2023) (Gm9:7691924, Gm9:7699360), suggesting a potential quantitative trait loci (QTL) region for NF and number of tips (NT). It’s plausible that the variations in our dataset are from differences due to environment and developmental stages considered during our measurements.

Our analysis uncovered key insights into the primary cell types and gene networks underlying RSA in soybean. Among the most striking findings was a strong association between root width and the endodermis, a trait often overlooked in favor of more commonly studied features like root length and branching patterns. Enhancing root width, however, may present new opportunities for improving soybean planting density and nutrient use efficiency by increasing the soil volume explored between neighboring plants. Within the endodermis, we identified a subnetwork of GWAS-prioritized genes, many of which encode amino acid and ion transporters. This expression profile aligns well with the endodermis's established role as a regulatory barrier, mediating the flow of water and nutrients into the vascular system. These findings suggest that natural variation in endodermal function could directly influence root width, highlighting a promising avenue for crop improvement. Beyond the endodermis, we detected associations between lateral root length and multiple cell types, including the cortex, epidermis, xylem pole pericycle, and notably, the metaphloem. Within the metaphloem, lateral root length emerged as the sole trait for which GWAS-associated genes demonstrated strong topological connectivity within the co-expression network. While auxin signaling is well known to mediate lateral root development, the regulatory landscape governing long-distance signaling and systemic coordination remains complex and less well understood (Geng et al. 2023). Given the transport-specialized function of the phloem, gene expression in the metaphloem may reflect systemic regulatory processes rather than locally restricted control. In this context, we identified a metaphloem-expressed gene involved in auxin signaling, which may offer new insights into how this vascular tissue contributes to lateral root development. These results underscore the importance of resolving cell-type-specific gene expression and functional networks when studying RSA, especially for traits influenced by intercellular communication and systemic signaling pathways.

*RCE1* and *NPR3* emerged as central regulators within the metaphloem-associated gene network, with both genes predicted to influence plant hormone signaling based on the functional

annotations of their *Arabidopsis* orthologs. In *Arabidopsis*, *RCE1* encodes a RUB1-conjugating enzyme that facilitates auxin signaling by promoting the degradation of AUX/IAA transcriptional repressors. It also plays a role in ethylene biosynthesis, further contributing to hormonal control of root development (Dharmasiri et al. 2003; del Pozo and Estelle 1999). In contrast, *NPR3* functions primarily as a salicylic acid (SA) receptor and was identified in our study as a highly connected hub across multiple gene subnetworks. *NPR3* regulates lateral root formation through crosstalk between SA and auxin pathways, balancing immune responses and growth. Together with its paralog *NPR4*, *NPR3* serves as a negative regulator of SA-mediated defense, acting both as a transcriptional repressor of defense genes and as an adaptor protein for *NPR1* degradation (Kong et al. 2020). The *npr3-1/npr4-3* double mutant exhibits a short-root phenotype, while the *npr1-1/npr3-1/npr4-3* triple mutant shows reduced lateral root emergence, highlighting the contribution of this signaling module to RSA. Since *NPR1* overactivation can suppress growth and compromise pathogen resilience, *NPR3* likely plays a key role in maintaining the growth-defense trade-off (Huot et al. 2014).

Given their network centrality and established roles in hormone-related pathways, we further investigated the functional relevance of *RCE1* and *NPR3* using *Arabidopsis* loss-of-function mutants. Both *rce1-1* and *npr3-1* mutants exhibited distinct RSA phenotypes relative to wild-type (Col-0), consistent with their identification as GC genes in soybean. The *rce1-1* mutant showed a significant reduction in convex hull area, suggesting a more compact and spatially restricted root system. In contrast, the *npr3-1* single mutant displayed a pronounced decrease in lateral root length, aligning with its predicted role in auxin and SA crosstalk during lateral root development (Ding et al., 2018; Fu et al., 2012; Kong et al., 2020). Importantly, *RCE1* and *NPR3* represent highly connected, cell-type-specific regulators that likely act downstream or in parallel to core hormone signaling networks to fine-tune growth responses to internal and external cues. This localized, context-specific control makes them attractive targets for engineering RSA traits, particularly when compared to modifying core components of hormone pathways, such as *ARF7*, which we also identified as a candidate in this study. Direct manipulation of central phytohormone regulators like *ARF7* often leads to pleiotropic effects, disrupting multiple developmental processes beyond the intended target trait (W.-G. Luo et al. 2023). By contrast, targeting *RCE1* and *NPR3* offers a strategic advantage by potentially enabling trait-specific modulation of RSA without compromising broader aspects of plant development or stress responses.

Precision genome editing of genes like *RCE1* and *NPR3* in targeted root cell types offers a compelling strategy for developing next-generation crop varieties with optimized root traits. Designing deeper and steeper root systems could substantially enhance water and nutrient uptake, particularly under drought or nutrient-limited conditions, thus contributing to climate resilience. Moreover extended and deeper root systems promise to facilitate carbon dioxide removal from the atmosphere via increased soil carbon sequestration, particularly in deeper soil layers (Faggiani Dias et al. 2025). These combined benefits position *RCE1* and *NPR3* as high-priority targets for RSA optimization and climate-smart agricultural innovation. The findings reported here represent the first example of integrating GWAS, GCN, and snRNA-seq to dissect a complex and environmentally relevant trait in a species like soybean that has long-range LD. This multi-omic approach has enabled the identification of functionally validated candidate genes and their regulatory contexts, providing a powerful framework for trait dissection in other crops. The genes and naturally occurring variation highlighted in this study offer valuable genetic tools for breeders aiming to meet critical challenges at the intersection of food security and climate adaptation.

## Materials and Methods

### *Germplasm collection*

A diverse set of 371 soybean plant introductions (PIs) was selected from the USDA Soybean Germplasm Collection (Valliyodan et al. 2021). The collection represents a wide genetic diversity with genotypes collected from over 20 different countries and includes maturity groups (MG) ranging from II to V. The population is mainly/majority have individuals/genotypes from China (222), USA (52), Korea (36), Japan (23) and Russia (11) (Supplementary Table 1).

### *Genotyping*

Sequencing data was generated using previously described methods (Valliyodan et al. 2021). In short, trimmed paired-end Illumina reads from each sample were aligned to the Williams 82 reference (Glycine max Wm82.a4.v1) using BWA (Li and Durbin 2009) and Picard tools (Wysoker, Tibbetts, and Fennell, n.d.). GATK (https://github.com/broadinstitute/gatk/) was then used to call variants. Next, VCF tools (Petr et al., n.d.) was then used to select biallelic SNPs that were present in at least 50% of samples with a minor allele frequency of no less than 5%, a minimum depth per sample (DP) of 10 reads, and minimum genotype quality score (GQ) of 10. This yielded 4,815,704 SNPs for downstream GWAS analyses.

### *High throughput gel-based phenotyping for root system architecture*

Soybean seeds were surface sterilized with chlorine gas (4mL HCL in 250 mL bleach solution) in a fume hood for 12 hours. One seed was sowed for each plastic cylinder filled with 170 ml of half strength Murashige and Skoog (MS) medium (pH 5.7) solidified using 0.8% phytagel. The dimensions of the plastic cylinders (Greiner Bio-One Polystyrene Container, Product No.: 968161) were 110 mm height and 68 mm diameter with volumetric capacity of 330 mL. Eight individual replicates were generated for each accession and placed in a randomized block design in a greenhouse with natural and artificial lighting (28 ± 8 °C, 14-h photoperiod). The images of the root systems were acquired 5- or 6-days post germination using our imaging system called Root Architecture 3D Imaging Cylinder (*RADICYL*). Similarly, an independent validation experiment was conducted using a subset of 24 soybean accessions from a diverse panel, with a new batch of seeds (Supplementary Table 4). This validation was performed in a separate bay of the same greenhouse during October–November 2024 and imaged on day 5, following the protocol described above.

The imaging system consisted of a telecentric lens mounted in a Basler acA2000-50gm GIgE camera, an aquarium, and a light source mounted on an optical breadboard. Gel-filled cylinders containing the soybean seedlings were placed in the water-filled aquarium that was placed on a turntable utilized to position the cylinders. This setup was back-illuminated by a near-infrared light source (Supplementary Fig. 1) and a rotational series of images of the cylinder were acquired (72 images in 5º steps). After image acquisition, seedlings were pulled out from the gel. Roots were separated from shoots and pinned on a cardboard with respective labels and then dried for 72 hours at 50ºC. The dried roots were then weighed using a fine scale.

### *Image analysis, skeletonizing and trait extraction*

We employed UNet++, a deep learning model, to segment primary/lateral roots from images collected using *RADICYL*. The UNet++ architecture was optimized using the Adam optimizer with an initial learning rate of 0.0001 for 60 epochs. We used ResNet101 as the encoder of the UNet++ model and softmax2d as the activation function of the last layer. The best validated model with the highest IoU score was saved as the trained model. During the segmentation phase, 72 images for each cylinder were cropped and used for primary/lateral roots segmentation based on the trained UNet++ framework. Roots on the images were then manually labeled using LabelMe (https://github.com/wkentaro/labelme) with three classes of labeled pixels: primary roots, lateral roots and background. The training-set contained 90

*RADICYL* captured images with a size of 2048 x 1080 pixels that were randomly selected and not part of our analysis. We cropped the images to a size of 990 x 860 pixels centered on the cylinder to remove pixels outside of the cylinder area.

The pcv.morphology.skeletonize function in PlantCv was used to skeletonize the roots (primary roots and lateral roots) from segmented images. The python library, OpenCV, was used to measure convex hull; the Hough Line Transform function from OpenCV was used to detect line segments originating from a given base of a lateral root on the primary root and ending on the tips of a lateral root, the average vertical angle for these short straight lines for all lateral roots was computed and output as vertical angle. During the prediction phase, 72 images for each cylinder were cropped and used for primary/lateral roots segmentation based on the trained UNet++ framework. The resulting black and white ground truth images were generated and cross checked with the manually annotated images. In total, nine different RSA-related traits were measured from each 2D image. A quality control step was performed by producing collages with segmented images of each plant belonging to a single genotype. We then identified poorly germinated or abnormally grown plants and excluded them from the final analysis. The trait mean and median data of the final set of images were used in the subsequent analysis. Each trait data set was tested for normality utilizing the Kolmogorov–Smirnov test and applied transformations to data that failed the normality test using the BestNormalize package in R.

***GWAS analysis pipeline***

GAPIT3 (Jiabo Wang and Zhang 2021) was utilized to test for associations between the SNPs from 371 samples and the 15 root traits (Supplementary Table 1). Using the median values for each accession and trait, plus the two most statistically powerful models (BLINK, FarmCPU) lead to 22 model-trait combinations being tested (gitlab.com/salk-tm/snake_gapit_gwas). BLINK and FarmCPU are both multilocus models, which incorporate the top three principal components derived from all the markers as covariates to reduce false positives. Additionally, BLINK iteratively incorporates associated markers as covariates to control for relationships among individuals. The associated markers are first selected using linkage disequilibrium (LD), then optimized for Bayesian information content, and finally reexamined across multiple tests to reduce false negatives. Results from each model-trait combination were then pooled and treated as independent tests. P-values of less than 5% after Bonferonni correction for multiple testing (p = 0.05 / 4.1 million SNPs) were considered as significant. We considered genes within 300 Kb of the significant SNP, which was based on our LD estimate, or up to ten genes on each side of the significant SNP.

***Construction and analysis of root gene networks***

Gene expression data representing diverse tissue types of soybean was obtained from Short Read Archive (SRA) (Supplementary Table 9). Samples were filtered for only samples related to the Wm82 accession, which resulted in a total of 71 gene expression samples selected for further analysis. The raw gene expression data was preprocessed and normalized using Salmon and transcript counts were quantified (Patro et al. 2017). DESeq2 was used to identify differentially expressed genes for all four tissue types (Love, Huber, and Anders 2014). The WGCNA package in R was applied to construct a gene co-expression network (Langfelder and Horvath 2008; Chang, n.d.). Network construction was performed using the blockwiseModules function for the entire dataset to calculate a pair-wise correlation matrix and adjacency matrix for each set of genes. We calculated the topological overlap matrix (TOM) for all pairs of genes in the co-expression network to quantify the strength of interconnection between two genes before

defining the modules within a network. Using the modules, ten was set as the minimum number of genes in the module and the threshold of cutting height. The gene modules were identified using the dynamic cut tree method and the modules with high similarity were combined to obtain 99 modules. Root-enriched modules were defined as the modules that had the highest number of root DEGs within the module. Initially, WGCNA generated a network comprising 219 million edges. After filtering only the root-enriched modules, this number was reduced to 154 million edges. Then by applying a threshold for the correlation coefficient (≥ 0.1), the network was further refined to include 10 million edges. A summary of the network was generated using PyGNA to investigate properties of the root-enriched subnetwork, including number of nodes (genes), edges (co-expression relationships), and degree (number of edges) of each node. We identified the root-enriched network to consist of 18,191 nodes, and 10,326,288 edges, with individual nodes having a minimum degree of 1 and a max degree of 7,516.

***Nuclei extraction and snRNAseq library construction***

Wm82 soybean seeds were sterilized in the same manner as the cylinder experiment above and grown on filter paper soaked in ½ MS for germination in a Percival growth chamber at 28ºC/18ºC and 12 hr/12 hr day night conditions. Seven days post germination, root material was harvested in liquid nitrogen and stored at -80 until further processing. For nuclei extraction (Lee et al. 2023), 75 roots were ground to a powder using a cooled pestle and mortar and homogenized in 20 ml Nuclei Extraction buffer (NEB) (20 mM MOPS (pH 7), 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, Supplemented with 0.5 % SUPER RNase inhibitor, 0.5 mM Spermidine, 0.2 mM Spermine, 1:100 dilution Roche Complete Protease Inhibitors). Samples were then sequentially filtered through a 70 µm and then 40 µm cell strainer and centrifuged for 5 minutes at 700 rcf (all centrifugation steps were performed at 4°C). The liquid phase was removed using an aspirator and the pellet was resuspended in NEB + 0.1% triton. Samples were incubated on ice for 15 minutes before being centrifuged at 700 rcf for 5 mins. This step was repeated for a total of three washes. Following the third wash, the pellet was resuspended in 4 ml NEB. A density gradient was used to separate the nuclei. For this 1 volume of diluent (120 mM Tris-Cl pH 8, 150 mM KCl, 30 mM MgCl2) was added to 5 volumes 60% Optiprep to make a 50% density buffer. This 50% stock was then used to make 45% and 15% solutions using a second dilution buffer (400 mM Sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-Cl pH 8). The gradients were made in 15ml tubes with 1 ml of 15% solution and 2 ml of 45% solution. Two ml of each sample was added to the gradient and centrifuged at 1500 rcf without breaks. Following centrifugation nuclei can be seen as a layer at the 45% mark in the gradient. These were collected into a 15ml falcon tube and centrifuged at 1000 rcf for 5 mins. The nuclei pellet was resuspended in 1 ml NEB and nuclei were sorted using Hoechst stain. The sorted nuclei were centrifuged at 700 rcf for 5 mins and resuspended in 50 ul 1xPBS to ensure compatibility with 10X Genomics library preparation. Libraries were made using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 according to manufacturer's instructions. cDNA and final library quality were assessed with a Bioanalyzer D1000 DNA Chip (Agilent), and libraries were sequenced using the NovaSeq SP 100 cycle kit (Illumina).

***Raw snRNA-Seq data pre-processing***

Initial processing of the raw single-nucleus RNA sequencing (snRNA-seq) dataset was performed using Cell Ranger v6.1.2 (10x Genomics), beginning with the mkfastq module for

demultiplexing. This workflow includes read alignment and the generation of gene-by-cell expression matrices. The reference genome and GTF annotation files for *G. max* were obtained from the Wm82 accession, and a custom reference was built using the cellranger mkref command with the --genome, --fasta, and --genes parameters. Single-cell gene expression quantification was conducted using the cellranger count command, specifying --id, --transcriptome, --fastqs, --sample, and --force-cells options.

***Cell clustering reconstruction via nonlinear dimensionality reduction***

We next aimed to determine if candidate genes associated with root-specific network modules were expressed in the same cell-type. Therefore, we leveraged the *G. max* snRNA-seq described above. This approach was instrumental in investigating the co-expression of gene networks within specific cell-types, offering a more refined understanding of the spatial and functional organization of genes in the root. In our analysis of Soybean snRNA-Seq datasets, we initiated the process by normalizing the UMI counts for each gene. This was achieved by dividing the UMI counts by the total UMIs in each cell, scaling the result by 10,000, and then applying a logarithmic transformation. We rigorously filtered the cells, setting stringent criteria based on the percentage of mitochondrial transcripts (percent.mt), the number of detected genes (nFeature_RNA), and the total mRNA molecules in each cell (nCount_RNA). Our thresholds were tailored to soybean datasets: nFeature_RNA per cell had to be more than 300 but less than 2500, percent.mt was capped at 1%, and nCount_RNA had to be between 300 and 4000. Following quality control, we retained 17,636 high-quality nuclei capturing the expression of 47,095 genes.

Next, we identified the 4,000 highly variable genes using the FindVariableFeatures function in capturing the broader variability in transcriptomes. Dimensionality reduction was then performed via PCA using the 'RnPCA' function. To address batch effects, we employed Harmony, ensuring accurate integration and analysis of our snRNA-Seq data. For unsupervised clustering, we utilized the 'FindNeighbors' function with the top 20 PCs and the 'FindClusters' function at a resolution setting of 0.3. These steps are crucial in revealing inherent grouping patterns within the data. The resulting clusters were visualized using the UMAP method, facilitating an intuitive understanding of the data's underlying structure. We successfully identified cortex, pericycle, vascular bundle/xylem pole pericycle, stele, endodermis, epidermis, xylem, vascular bundle, metaphloem with known soybean marker genes, orthologs of marker genes in *Arabidopsis*, (Supplementary Table 12).

***snRNA-Seq data post-processing***

10 cell-types were annotated based on known marker genes and gene expression profiles. Cell clusters "inner cortex", "outer cortex", and "potential cortex" were merged into the cortex cluster to simplify the analysis. 47,094 genes were sorted by their variance across cell-types, and the top 10% (4,709) most variable genes were selected for downstream analysis. For each variable gene, we scaled expression values to a standard deviation of one across cell-types, and a common mean across genes. We divided the range of scaled expression values into three equal bins, and for each cell-type chose genes with scaled expression values in the top bin as the cell-type specific gene set.

***PyGNA 2 analysis***

We developed a Python package based on the API of PyGNA, which we called PyGNA2, to facilitate our hypothesis tests and visualizations (Fanfani, Cassano, and Stracquadanio 2020). PyGNA2 was employed to summarize the gene networks. We used the 'pygna2 summary' function to obtain key network statistics, providing an overview of network characteristics. We employed the 'pygna2 test' function to assess the relationship between the root-enriched gene

network and the single-cell gene sets. This function tests the associations between two networks using the Topology Random Walk with Restart test and the Topology Internal Degree test (TID). We performed all our final analysis with 8000 permutations. We visualized the results using 'pygna2 cytoscape' with the –minimal option to visualize subnetworks including specific cell-type gene sets, GWAS candidate genes, and genes on shortest paths between them.

**Network filtering and Cytoscape analysis**

We identified the largest connected component within the significant cell-type vs. trait network and designated this as "cell-type and trait". We retained only the nodes corresponding to genes associated with cell-type and trait within the larger network component to simplify the network further. Node selection was guided by the 117 GRIT genes identified from filtering by root DEGs within root modules in single nucleus data. Additionally, we determined the degrees of connectivity for each subnetwork and determined that these central connecting nodes to be the most significant.

***Arabidopsis* mutant analyses**

*Arabidopsis* GABI-Kat mutant lines were obtained from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University. Homozygous individuals were identified by resistance screening on 1% agarose [Phytotech] plates supplemented with 2.22 g/L Murashige & Skoog Basal Salts (pH 5.8) and 7.5 mg/L sulfadiazine [Sigma-Aldrich]. Prior to phenotyping, mutant seeds and wild-type (Col-0) controls were sterilized using chlorine gas (generated from 3.5 mL HCl and 250 mL of 6% bleach) for one hour, then stratified in the dark at 4°C for three days. Stratified seeds were sown on 1% Phytagel [PlantMedia] plates containing 1% sucrose and 2.22 g/L Murashige & Skoog Basal Salts (pH 5.8) [Phytotech], in 120 × 120 × 15 mm square Petri dishes [Greiner]. Plates were oriented vertically and incubated for four days at 21°C under a 16-hour light/8-hour dark photoperiod. Seedlings were then transferred to 110 × 68 mm (height × diameter) vertical cylinders [Greiner] filled with 170 mL of the same medium and maintained under identical environmental conditions Root system architecture of mutant and wild-type (Col-0) control plants was phenotyped at multiple timepoints between 14 and 28 days post-transfer. At each timepoint, individual plants were imaged on a rotating platform with standardized captures taken every 5°, resulting in 72 images per plant per timepoint for subsequent 3D root reconstruction. Phenotypic traits were quantified using the same automated pipeline described above for soybean.

**Data and code availability**

The code to analyze the WGCNA network and single-cell data can be found here: https://gitlab.com/salk-tm/soybean-root-gwas/. RADYCL Segmentation pipeline for image analysis can be found here: https://github.com/Salk-Harnessing-Plants-Initiative/SSRAPC-Soy-Segmentation-Root-Architecture-Phenotyping-for-Cylinder.git. PyGNA2 is available on PyPI (https://pypi.org/project/pygna2/) and GitLab (https://gitlab.com/salk-tm/pygna2).

**Author Contributions**

T.P.M., and W.B. conceived the study. C.D.C. and M.S. conducted phenotyping measurements. A.R. developed and implemented the image analysis pipeline and conducted image quality control and phenotypic validation experiments. C.M. performed single-nuclei RNA sequencing. R.C.L. and A.A. performed statistical analyses and contributed to trait data transformation and interpretation. A.A. and Y.S. developed the network analysis. L.Z. supported the development of the phenotyping workflow, contributed to image analysis and single cell. D.L. assisted with data processing and management. C.M. and S,J-H. conducted *Arabidopsis* RSA experiments. H.Y. and H.T. N. provided reagents (seeds), WGS, genotyping, and curated VCF files. T.P.M., H.T.N.

and W.B. supervised the project and contributed to experimental design, data interpretation, and manuscript writing. All authors reviewed and approved the final manuscript.

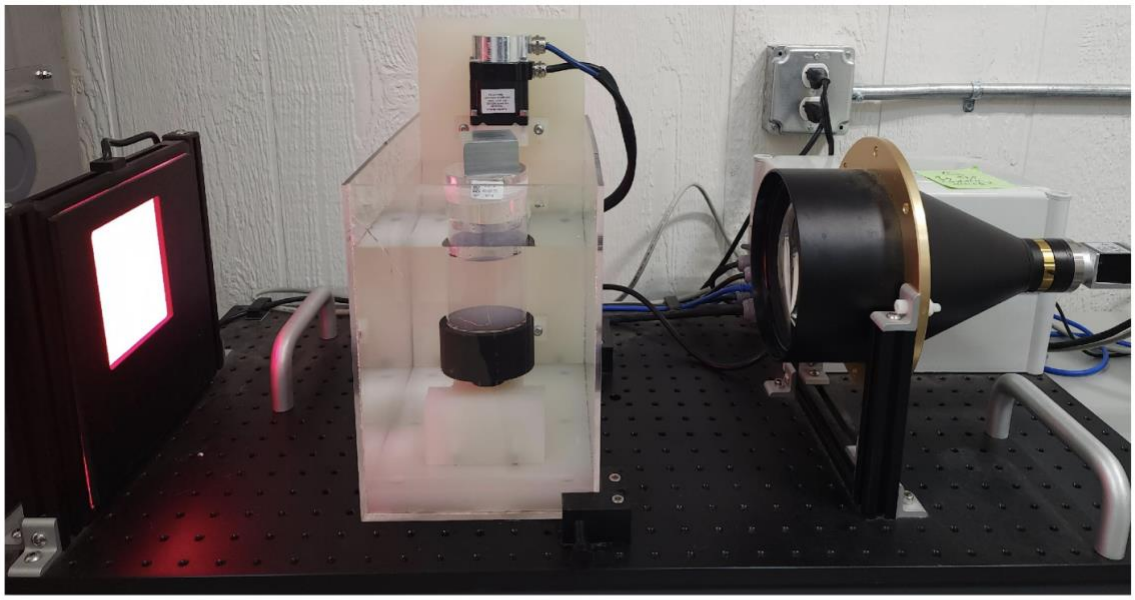
**Acknowledgments**

We thank Emily Shane and Charidimos Georgousakis for phenotyping. This work was supported through the Salk Harnessing Plants Initiative (HPI) with funding from the TED Audacious, Bezos Earth Fund, and Hess Corporation.

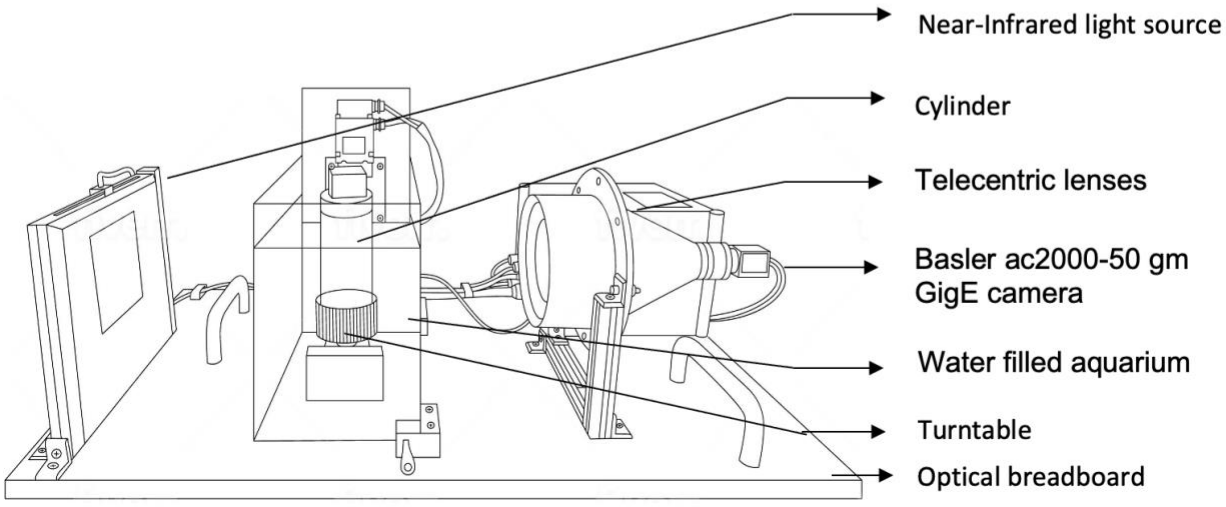
***Declaration of interest***

W.B. and T.P.M. are co-founders of Cquesta, a company that works on crop root growth and carbon sequestration.

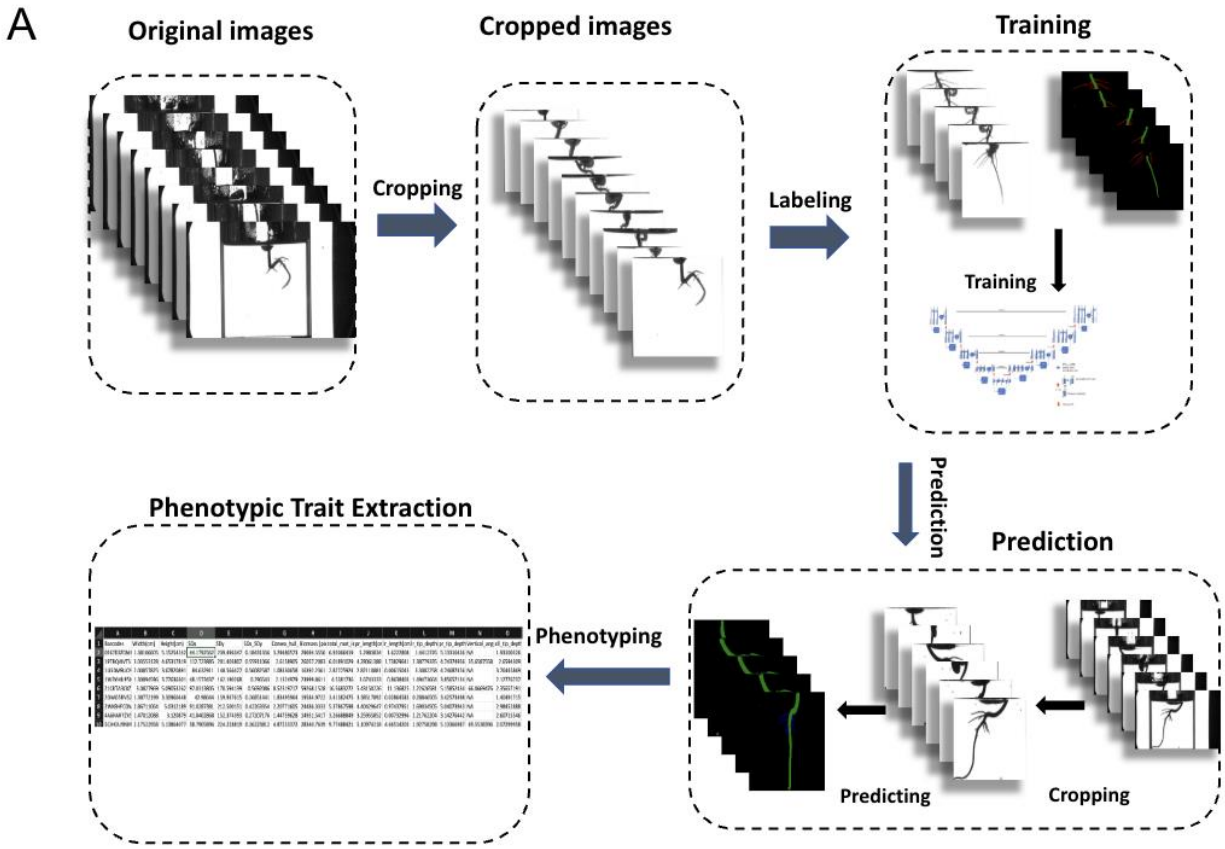
**Supplemental Figures**

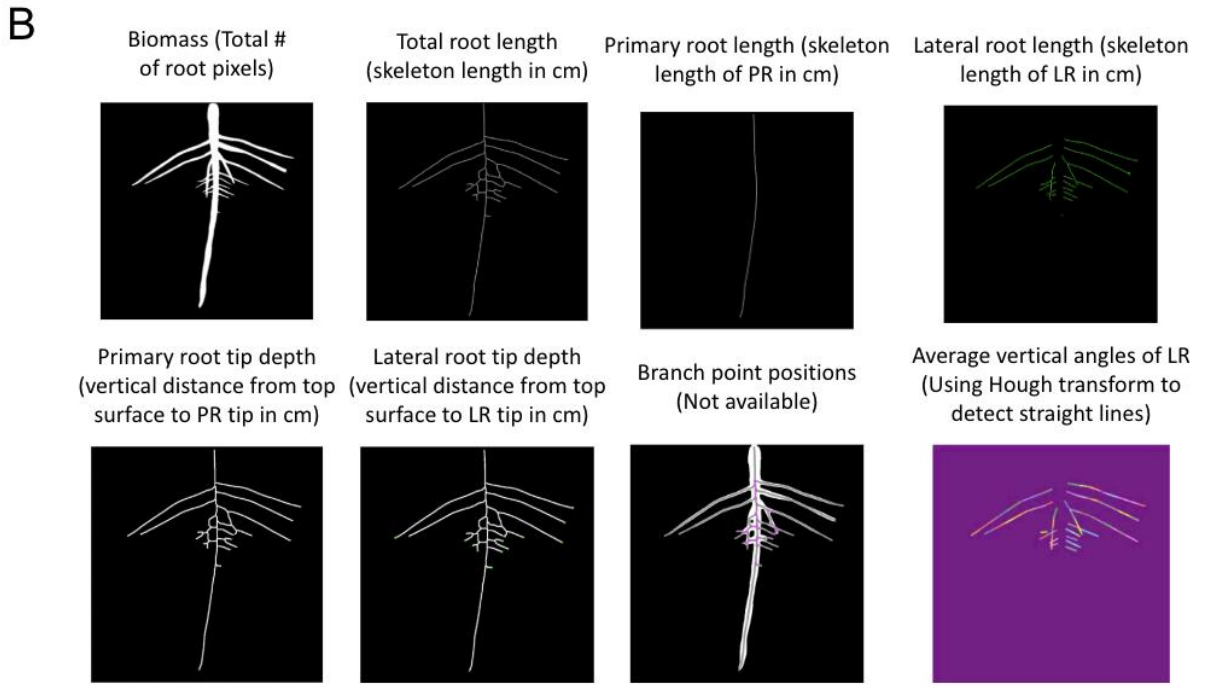


*RADICYL* (**R**oot **A**rchitecture 3**D** **Cyl**inder)



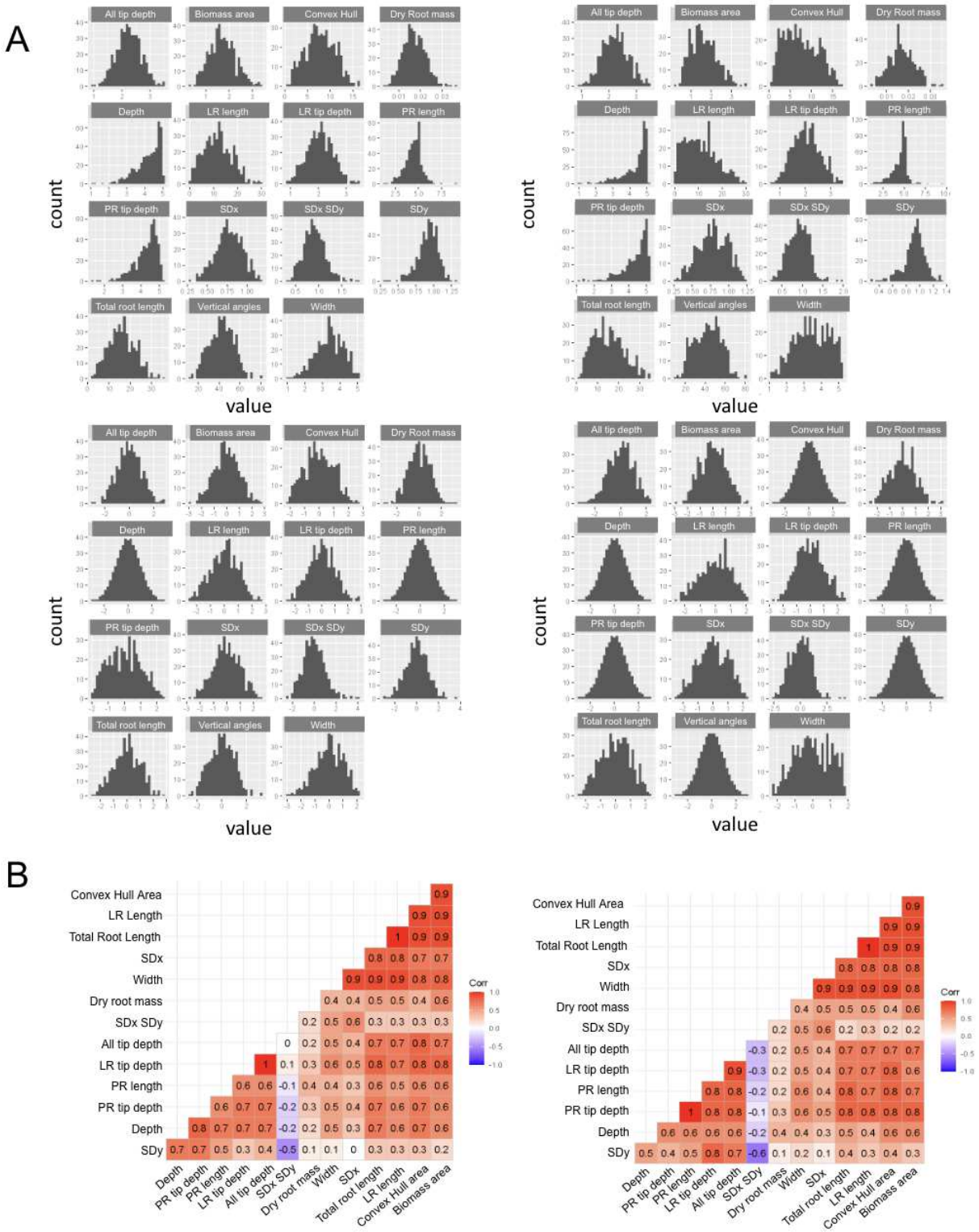
**Supplementary Figure 1. Root Architecture 3D Cylinder (*RADICYL*), a high-throughput phenotyping system.** The *RADICYL* system comprises key components, including an automated turntable equipped with a Basler acA2000-50gm GigE camera and a telecentric lens, positioned above an aquarium. Additionally, an optical breadboard-mounted light source is used to illuminate the subject. In the experimental setup, cylindrical samples are submerged in a water-filled aquarium, which is then placed on the turntable. A rotational series of images is captured, consisting of 72 images at 5-degree intervals. Subsequently, seedlings are carefully extracted from the hydrogel medium following image acquisition for dry biomass measurements.



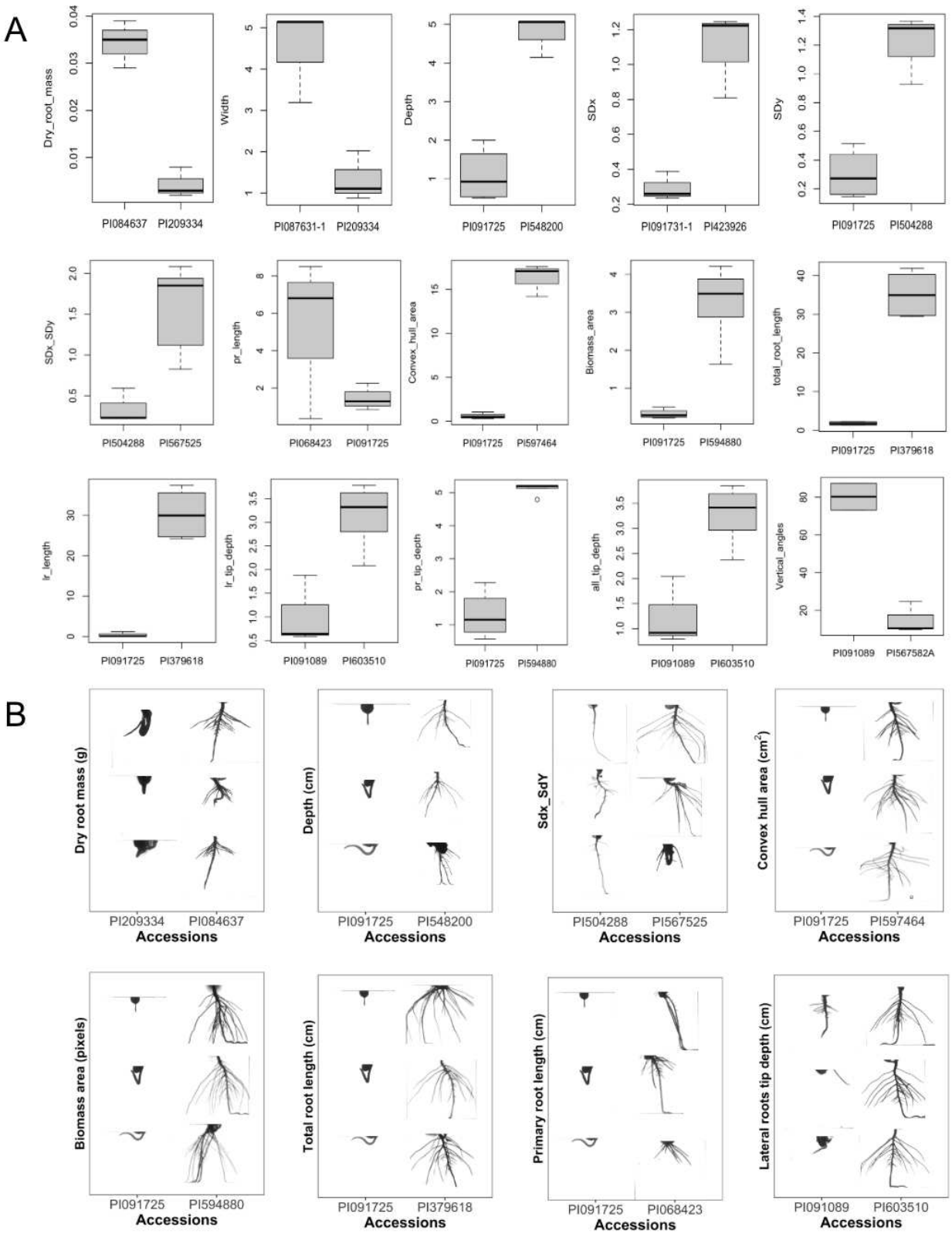




**Supplementary Figure 2. Analysis of Root Architecture 3D Cylinder (RADICYL) images.** A) An illustrative example of the processing steps undertaken to extract the trait quantifications after capturing the initial images. B) Segmentation of primary roots and lateral roots was accomplished using the UNet++ architecture, a convolutional neural network (CNN) designed

for semantic segmentation tasks. During the segmentation phase, 72 images from each cylinder were cropped and used for primary and lateral root segmentation based on the trained UNet++ framework.



**Supplementary Figure 3. Validation of the Soybean RSA phenotyping using *RADICYL*.** A) The histograms display the mean (on the left) and median value (on the right) for 15 RSA traits. To ensure the data met normal distribution criteria, adjustments were made to transform the data (displayed in the bottom left and right). For the GWAS analysis, the median values were used to avoid potential biases from extreme values that could skew the mean. B) In correlation plots, the left side illustrates correlations using the mean, while the right side does so with the median. Both plots emphasize strong connections between the various quantified traits, revealing a high level of correlation among them.




**Supplementary Figure 4. Extreme soybean RSA phenotypes.** A) Box plots depict the two highest and lowest accessions associated with the median of each trait. Each box plot summarizes the replicates for the respective trait. Notably, several accessions exhibit both high and low associations across multiple traits, revealing complex relationships within the dataset.

B) Phenotypic differences in two extremes, the highest and lowest accessions associated with the median of few traits.

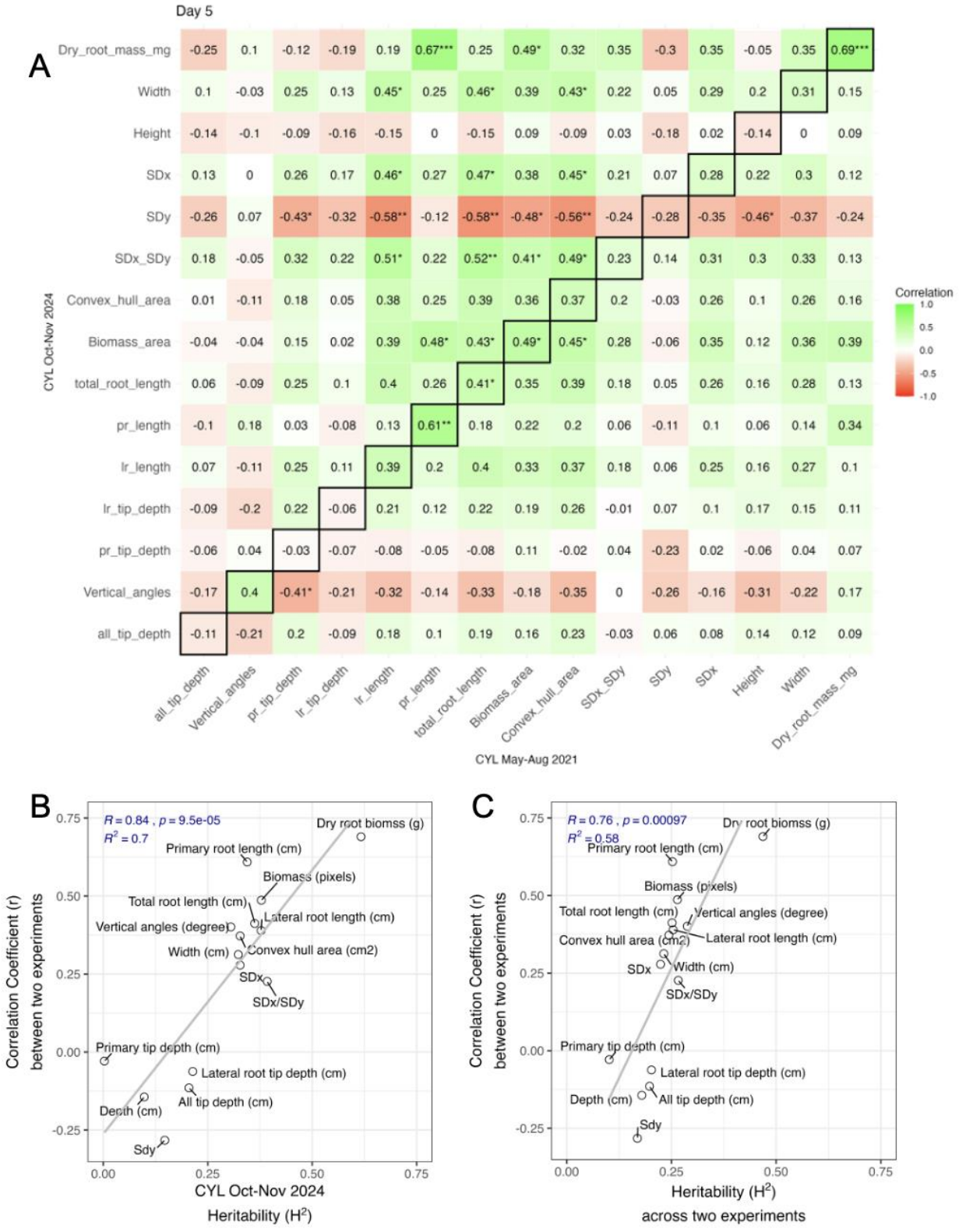
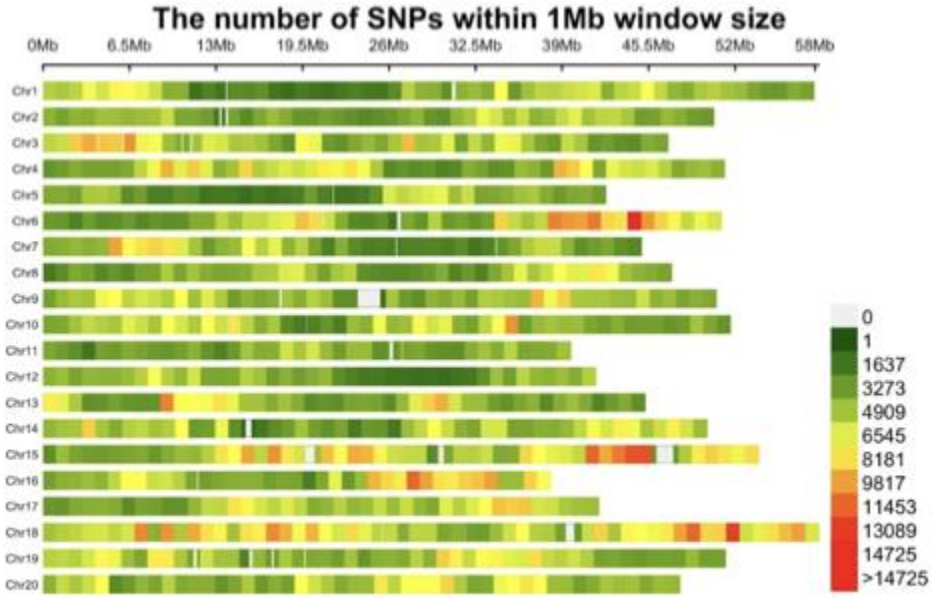
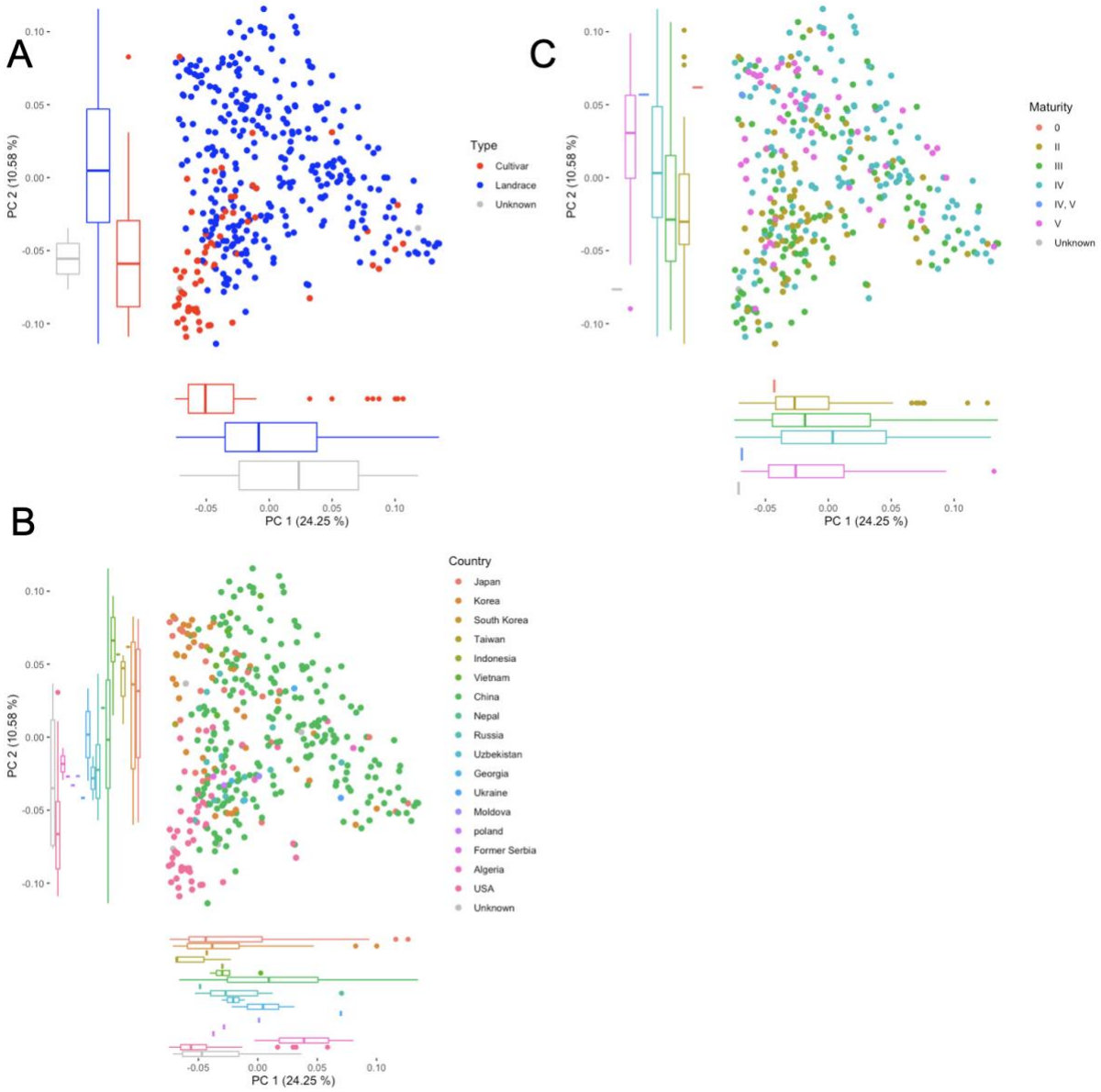


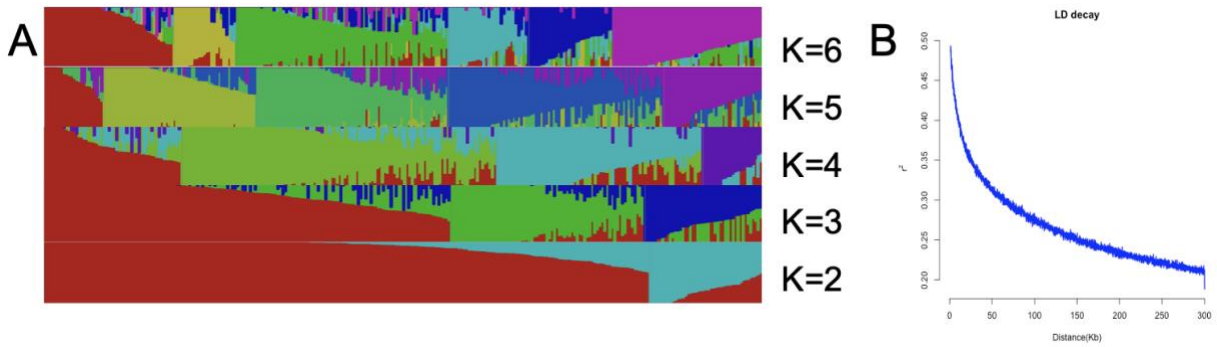
**Figure 5. Reproducibility and heritability of root system architecture (RSA) traits in soybean.** A) Pairwise correlation matrix of RSA traits measured on day 5 across two independent experiments conducted in the same controlled environment. Traits from the October–November 2024 experiment (y-axis) were correlated with the same traits measured in the May–August 2021 experiment (x-axis). Correlation coefficients (Pearson's *r*) are represented by the color scale from -1 (red) to 1 (green). Significant correlations ($p < 0.001$) are outlined in black. B) Relationship between trait heritability ($H^2$) in the October–November 2024 experiment and cross-experiment correlation coefficients (trait reproducibility). Each point represents one trait; traits with higher $H^2$ tend to show stronger reproducibility. C) Same as (B), but $H^2$ values are averaged across both experiments. Solid lines indicate linear regression fits with corresponding *R*, $R^2$, and *p*-values shown above.



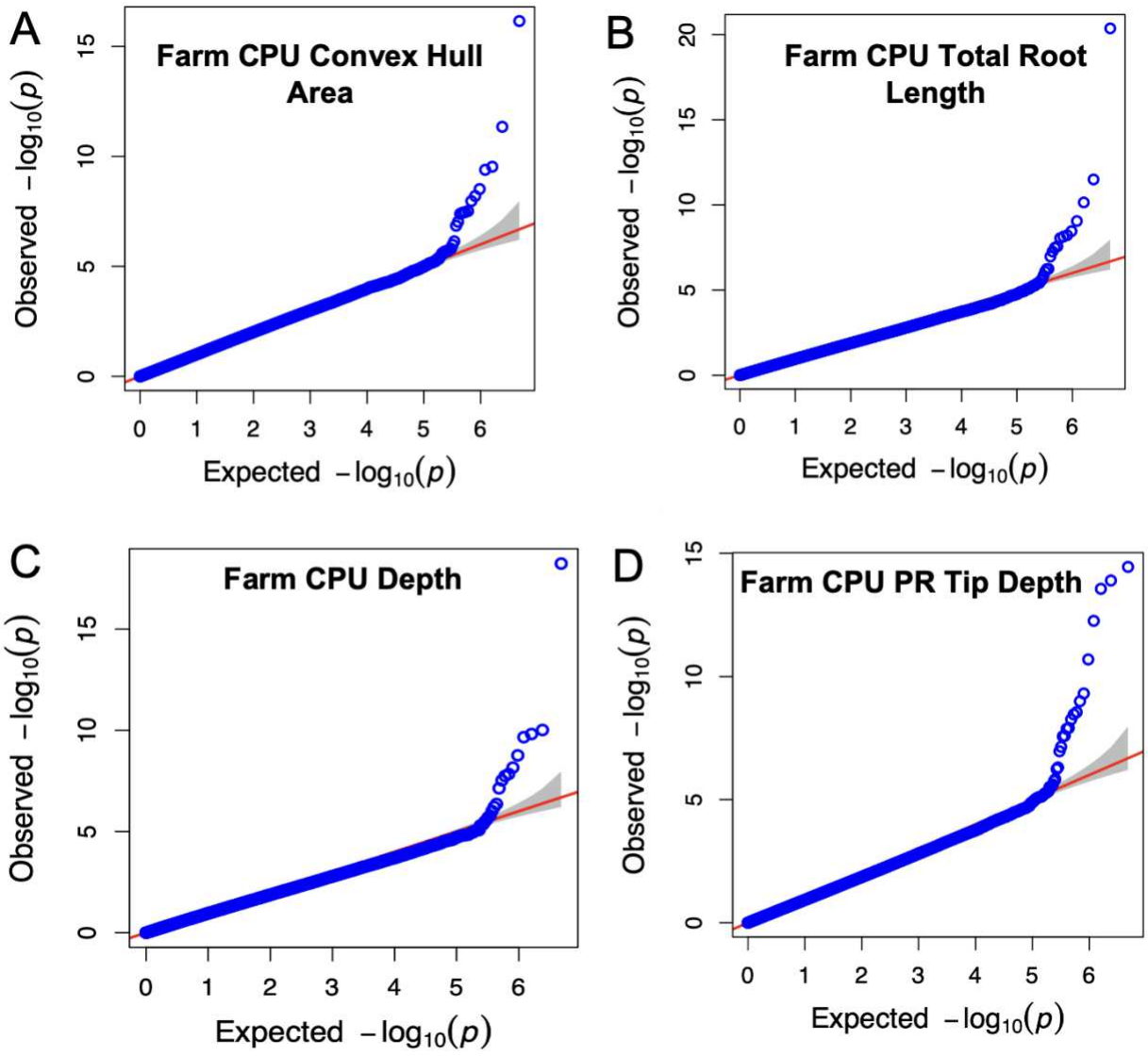

**Supplementary Fig. 6. A Heatmap showcasing the distribution of Single Nucleotide Polymorphisms (SNPs) across chromosomes 1 to 20.** The color gradation, from green indicating fewer SNPs to red indicating higher counts, represents the density of SNPs within 1 Mb window sizes. These SNPs are sourced from the VCF file produced through whole genome sequencing.



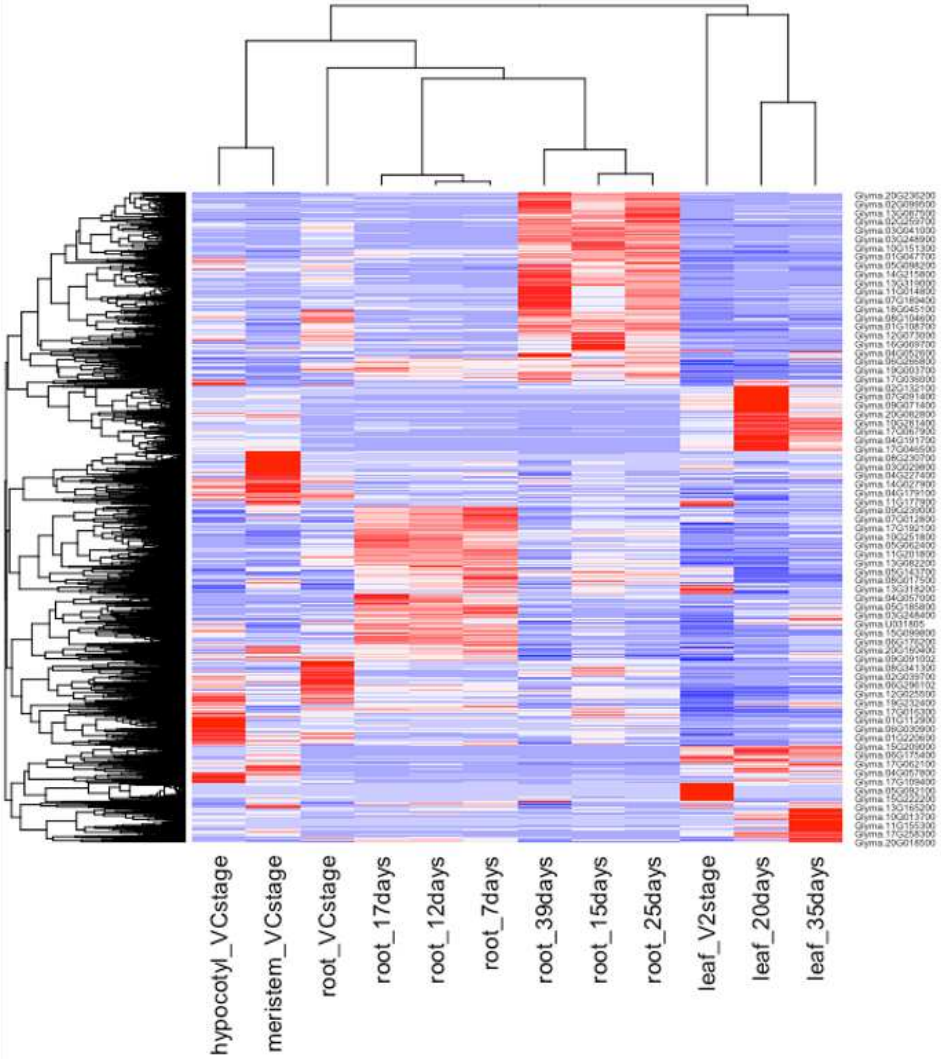

**Supplementary Fig. 7. Principal component analysis (PCA) of soybean accessions based on genomic variation.** PCA was conducted using genome-wide SNP data to assess population structure among soybean accessions. The first two principal components are shown, explaining 24.25% (PC1) and 10.53% (PC2) of the total variance, respectively. **A)** Accessions colored by type: cultivars (red), landraces (blue), and unknown type (gray). Boxplots display the distribution of PC1 and PC2 scores for each group. **B)** Accessions colored by country of origin. Distinct geographic clustering patterns are observed, particularly among accessions from Japan, Korea, China, and Vietnam. **C)** Accessions colored by maturity group (0–V and unknown). Variation in PC1 is partially associated with maturity group, with later-maturing accessions tending toward positive PC1 values. Boxplots in each panel illustrate group-wise dispersion along PC1 and PC2.



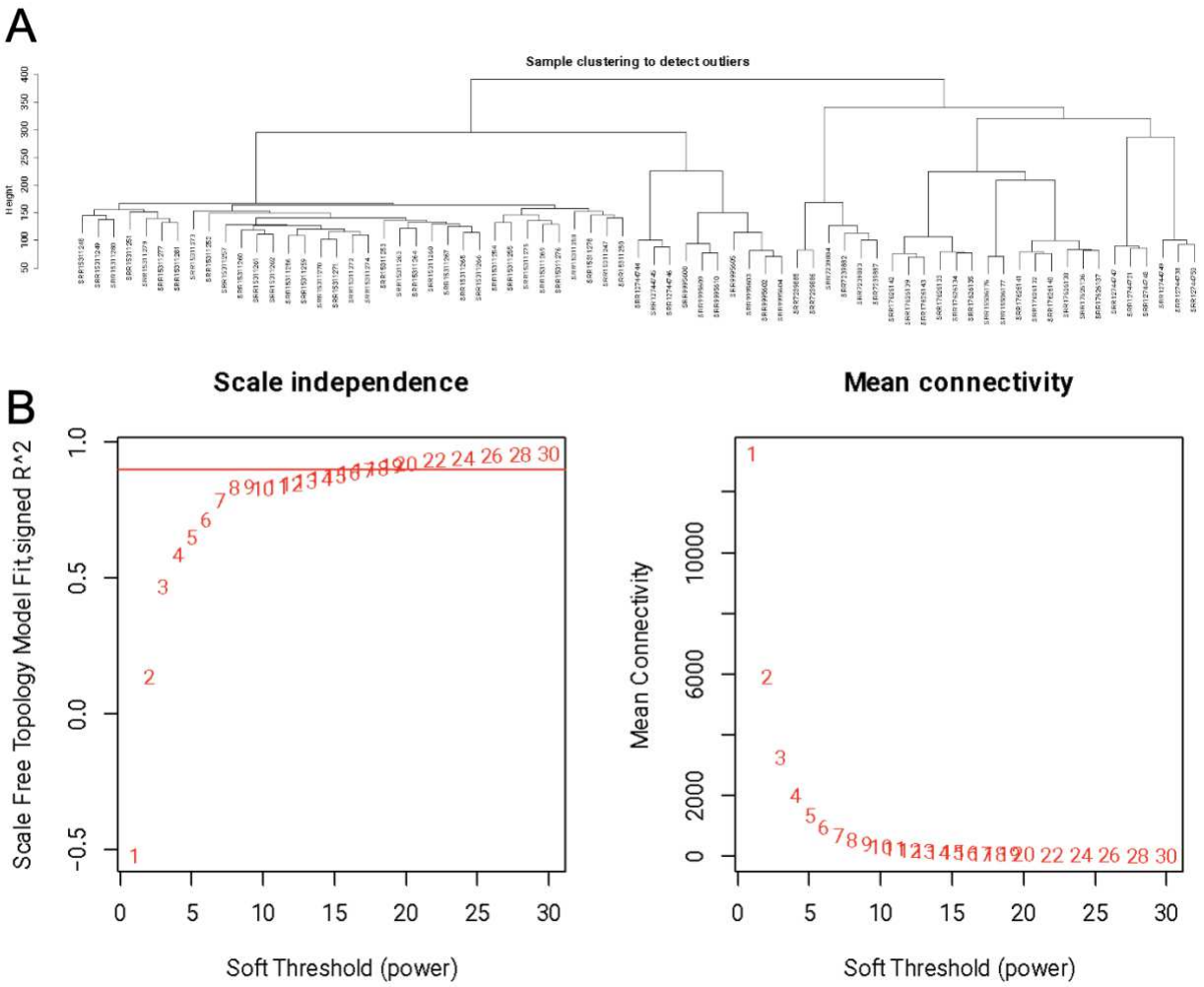

**Supplementary Fig. 8. Population structure and linkage disequilibrium decay in soybean accessions.** A) Population structure analysis of soybean accessions based on genome-wide SNP data using STRUCTURE. Bar plots show individual membership coefficients for K = 2 to K = 6, where each vertical bar represents an individual and colors denote inferred genetic clusters. As K increases, greater subpopulation complexity is revealed, indicating hierarchical structure within the population. B) Linkage disequilibrium (LD) decay plot showing the decline of pairwise correlation coefficient ($r^2$) as a function of physical distance (Kb) between SNPs. LD decays rapidly within the first 100 Kb and plateaus beyond 250 Kb, indicating the resolution limits of association mapping in this population.



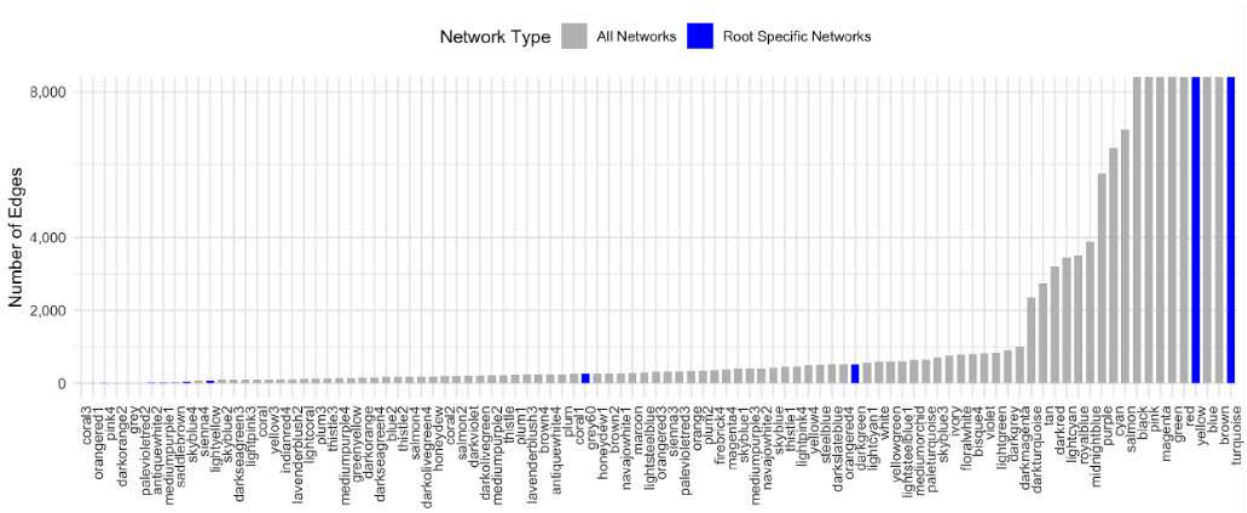

**Supplementary Fig. 9. Quantile–quantile (Q–Q) plots of GWAS for root system architecture traits in soybean using the FarmCPU model.** Q–Q plots show the distribution of observed versus expected $-\log_{10}(p)$ values for association tests conducted on four root traits: A) Convex hull area; B) Total root length; C) Root system depth; and D) Primary root (PR) tip depth. Blue circles represent observed GWAS p-values, the red diagonal line indicates the null expectation under no association, and the gray area shows the 95% confidence interval. Deviations from the null line at higher values suggest significant marker–trait associations.



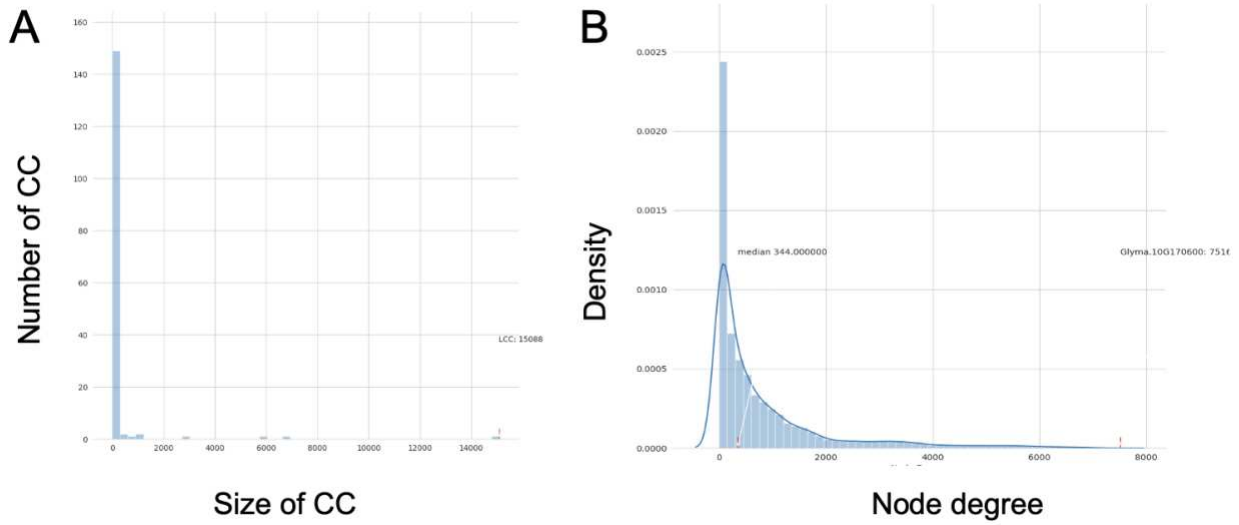

**Supplementary Fig. 10. Heatmap of gene expression across soybean tissues and developmental stages.** Hierarchical clustering heatmap of differentially expressed genes across soybean samples representing various tissues and developmental timepoints. Samples include hypocotyl, meristem, and root at VC stage, and root at 7, 12, 15, 17, 25, and 30 days after germination, as well as leaf at V2 stage, 20 days, and 35 days. Rows represent individual genes (labeled by Glyma IDs), and columns represent tissue-stage combinations. Expression values were normalized and scaled, with red indicating higher expression and blue indicating lower expression. Both genes and samples were clustered using hierarchical clustering, revealing groups of co-expressed genes and their temporal or spatial expression patterns.



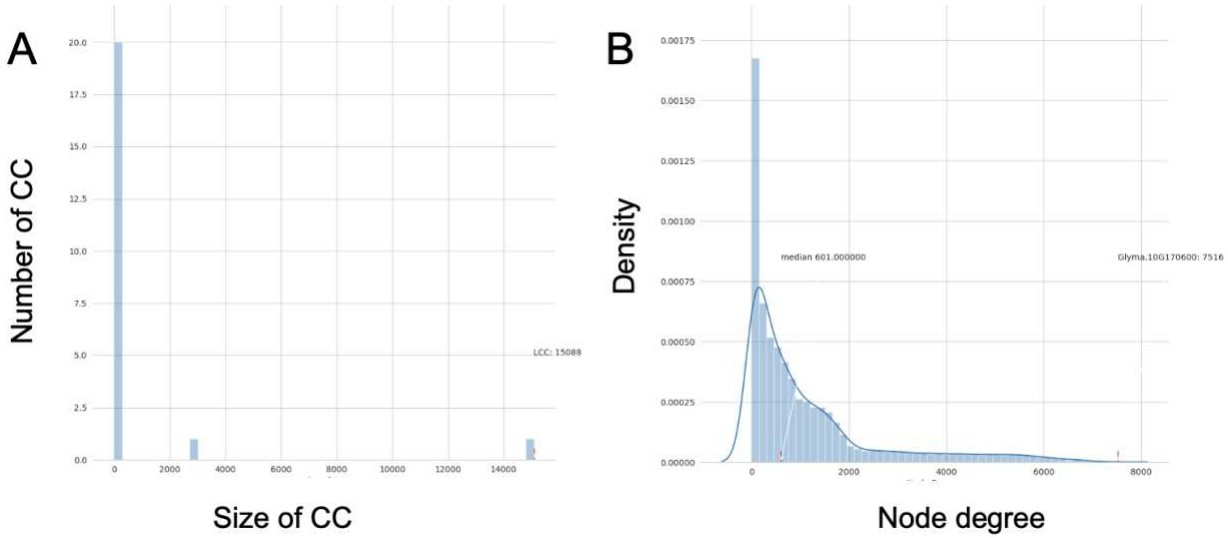

**Supplementary Figure 11. Sample clustering and soft-threshold selection for weighted gene co-expression network analysis (WGCNA). A)** Hierarchical clustering dendrogram of all samples used in the analysis, based on expression similarity. This clustering was used to assess sample quality and identify potential outliers. **B)** Analysis of network topology to determine the appropriate soft-thresholding power. The left panel shows the scale-free topology fit index ($R^2$) as a function of the soft-threshold power, with the optimal threshold selected where $R^2$ exceeds 0.9. The right panel shows the mean connectivity for each power, illustrating how increasing power reduces average connectivity across the network.



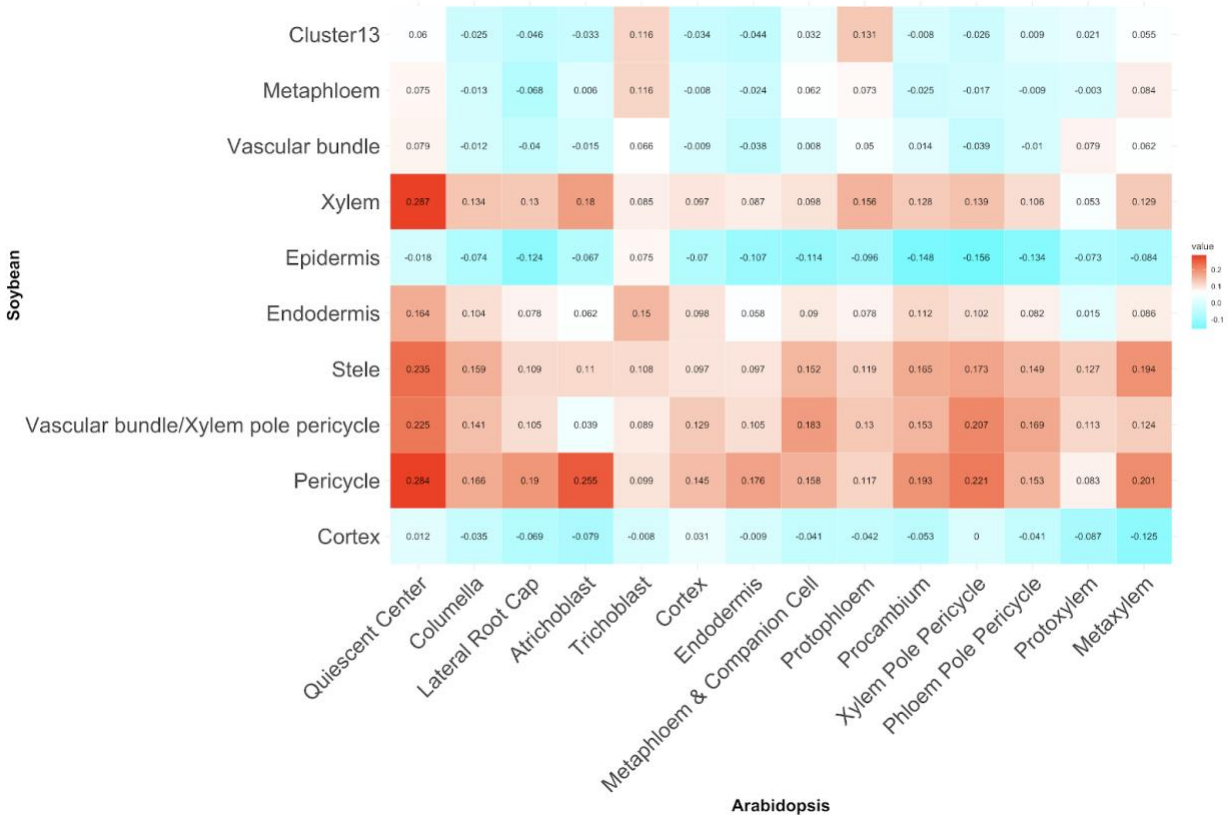

**Supplementary Figure 12. Distribution of root-specific gene co-expression networks.** Bar chart showing the number of edges (gene–gene connections) within each gene co-expression module identified through WGCNA. Modules are color-coded along the x-axis, and bars are colored by network type: all modules (gray) and root-specific modules (blue). The turquoise module had the largest number of root-specific edges, indicating a strong enrichment for differentially expressed genes (DEGs) in root tissue. Several smaller modules also showed root-specific network associations, suggesting modular specialization across the expression dataset.



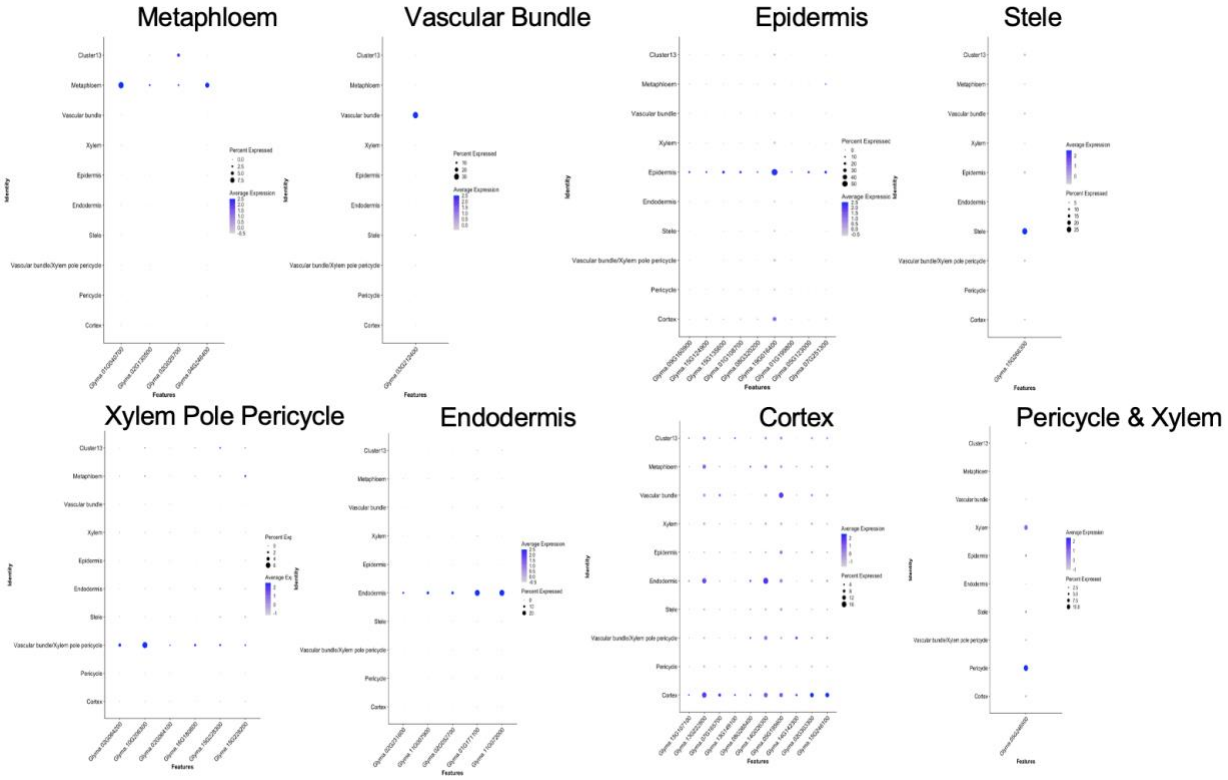

**Supplementary Figure 13. Overview of network connectivity and component structure.** A) Distribution of connected components (CCs) across the full gene co-expression network. The x-axis represents the size of each component (number of genes), while the y-axis indicates the number of components of a given size. The largest connected component (LCC) includes 15,088 genes, highlighting the extensive connectivity of the main network. B) Node degree distribution within the network. The x-axis shows the number of connections (degree) per gene, and the y-axis represents the density of genes with that degree. The median node degree is 344, while one highly connected gene (Glyma.10G170600) exhibits the highest degree of 7,518, indicating it may serve as a major regulatory hub.



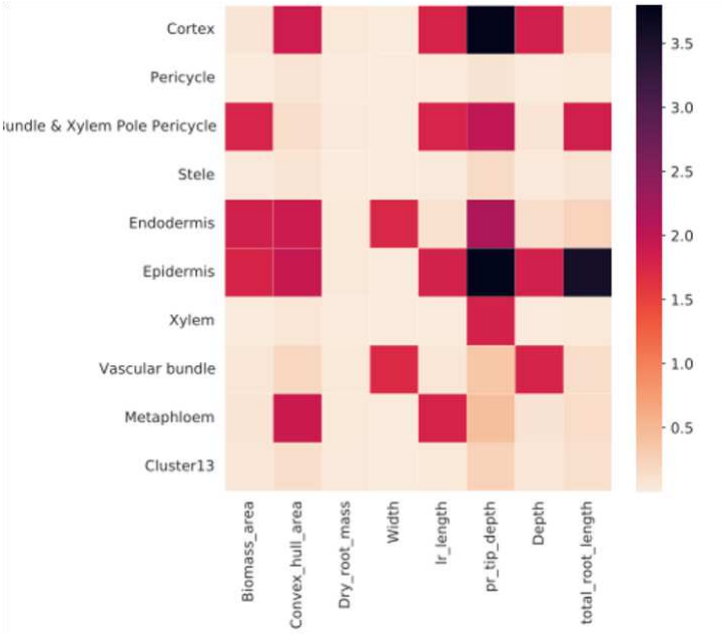

**Supplementary Figure 14. Network connectivity and component distribution in the root-enriched gene co-expression network.** A) Histogram showing the distribution of connected components (CCs) by size within the root-enriched network. The x-axis indicates the number of genes in each component, while the y-axis represents the number of components of a given size. The largest connected component (LCC) includes 15,088 genes, reflecting a highly integrated core network. B) Node degree distribution within the root-enriched network. The x-axis shows the number of connections (degree) per gene, and the y-axis represents the density of genes with a given degree. The median node degree is 601, while the most highly connected gene, *Glyma.10G170600*, has 7,516 connections, suggesting a central regulatory role.



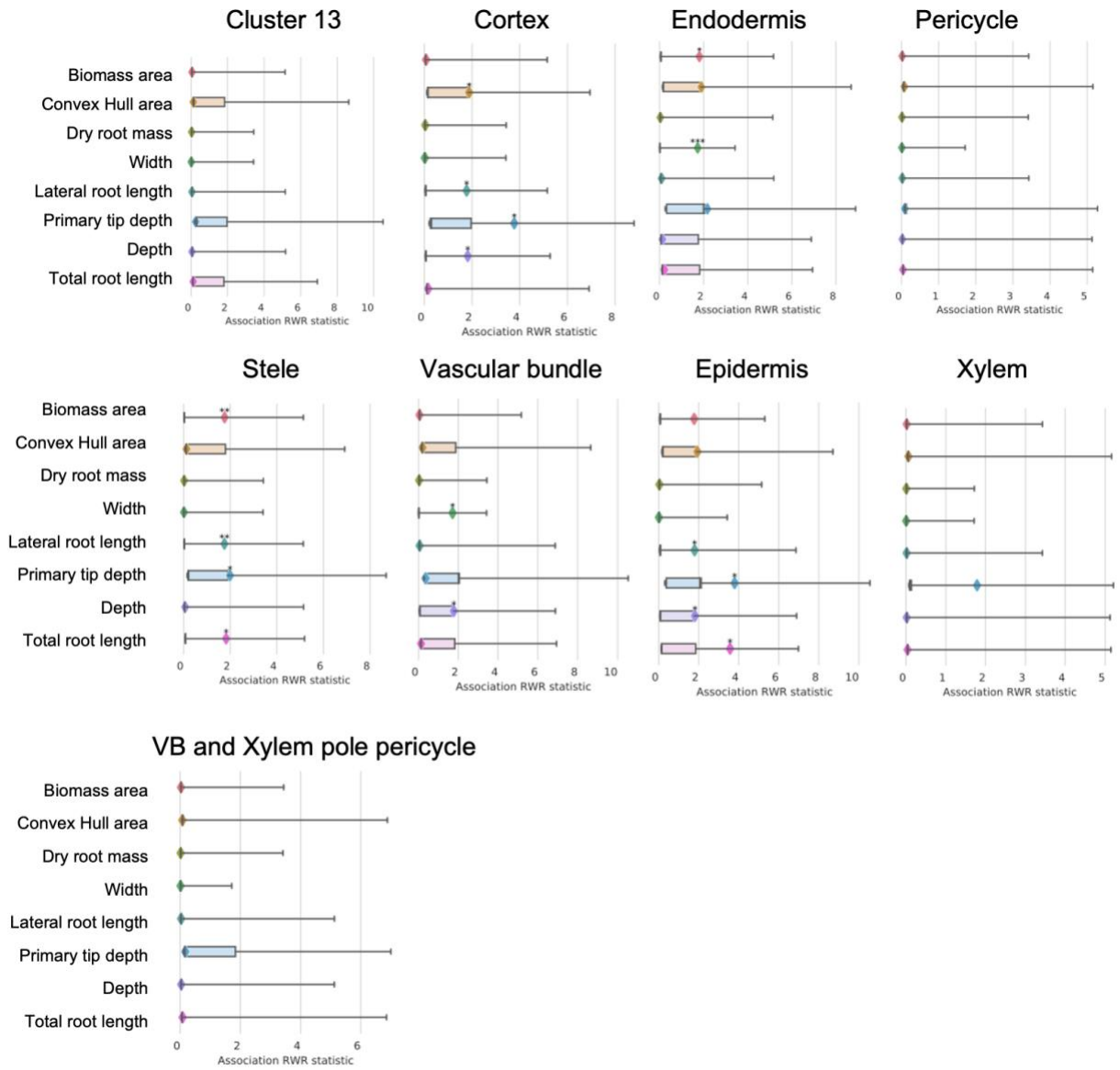
**Supplementary Figure 15. Cross-species correlation of root cell types between soybean and Arabidopsis.** Heatmap showing Spearman rank correlation coefficients between annotated cell types or clusters in soybean (rows) and Arabidopsis (columns). Each value represents the degree of transcriptional similarity between corresponding cell-type gene expression profiles across species. Positive correlations are shown in red, while negative correlations are shown in blue, as indicated by the color scale. Notably, strong positive correlations were observed between Arabidopsis xylem and pericycle cell types and their corresponding soybean counterparts, highlighting conserved root cell-type transcriptional signatures.



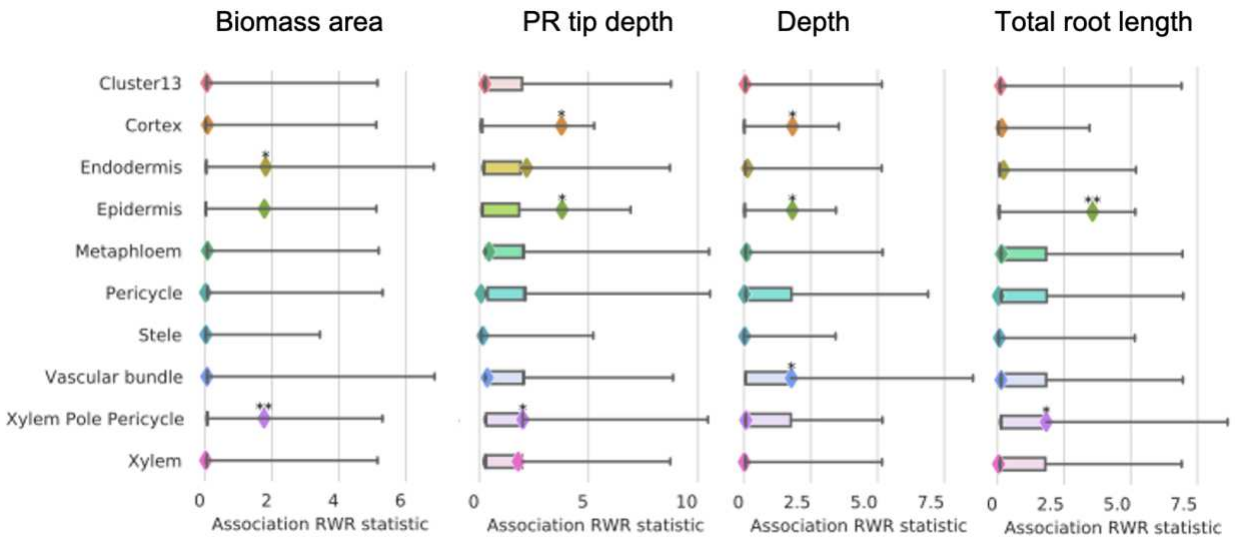

**Supplementary Figure 16. Expression patterns of cell type–specific marker genes in soybean root clusters.** Dot plots display the expression profiles of selected marker genes across annotated soybean root cell types, including metaphloem, vascular bundle, epidermis, stele, xylem pole pericycle, endodermis, cortex, and combined pericycle & xylem. Each dot represents a gene's expression in a given cell type. Dot size corresponds to the percentage of cells expressing the gene, and color intensity indicates the average expression level within those cells. These profiles illustrate the specificity and enrichment of marker genes within distinct root cell identities.



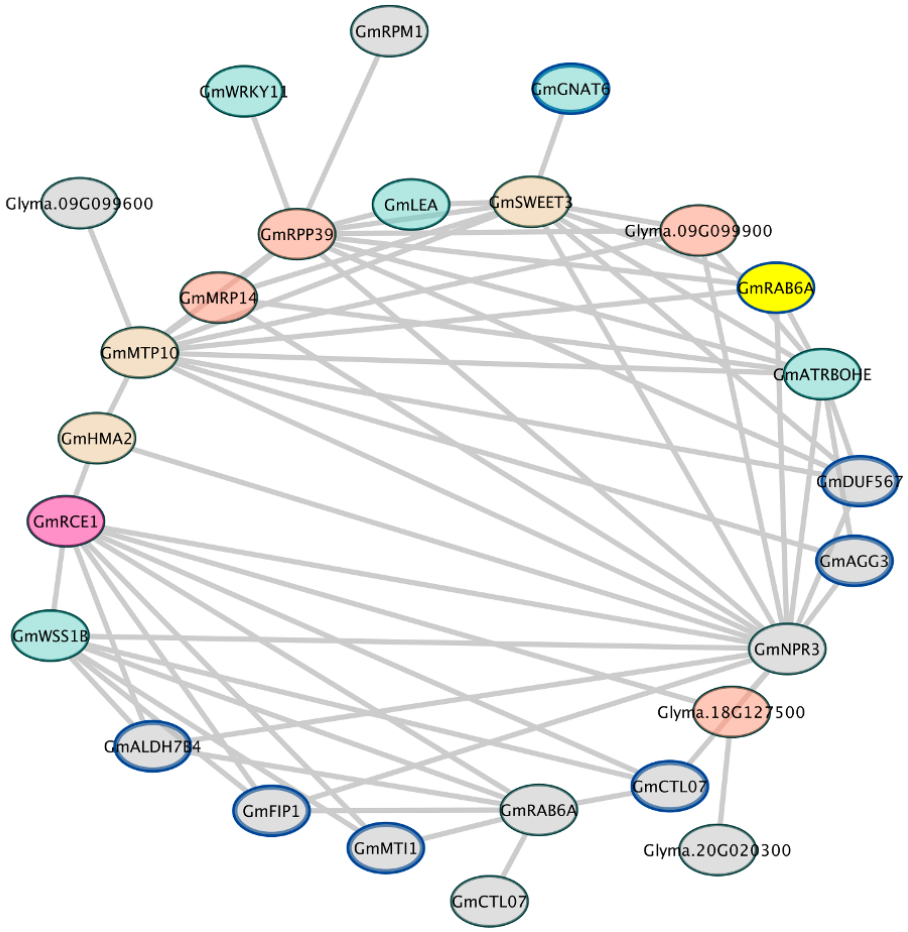

**Supplementary Figure 17. Heatmap of significant trait–cell type associations based on Random Walk with Restart analysis using PYGNA.** The heatmap displays $-\log_{10}(p)$ values representing the statistical significance of associations between specific root traits (columns) and root cell types or clusters (rows) in soybean. Warmer colors (red to black) indicate higher significance, with darker shades corresponding to stronger deviations from the null model. Statistical significance thresholds are annotated as: * $p < 0.01$, ** $p < 0.005$, *** $p < 0.001$.



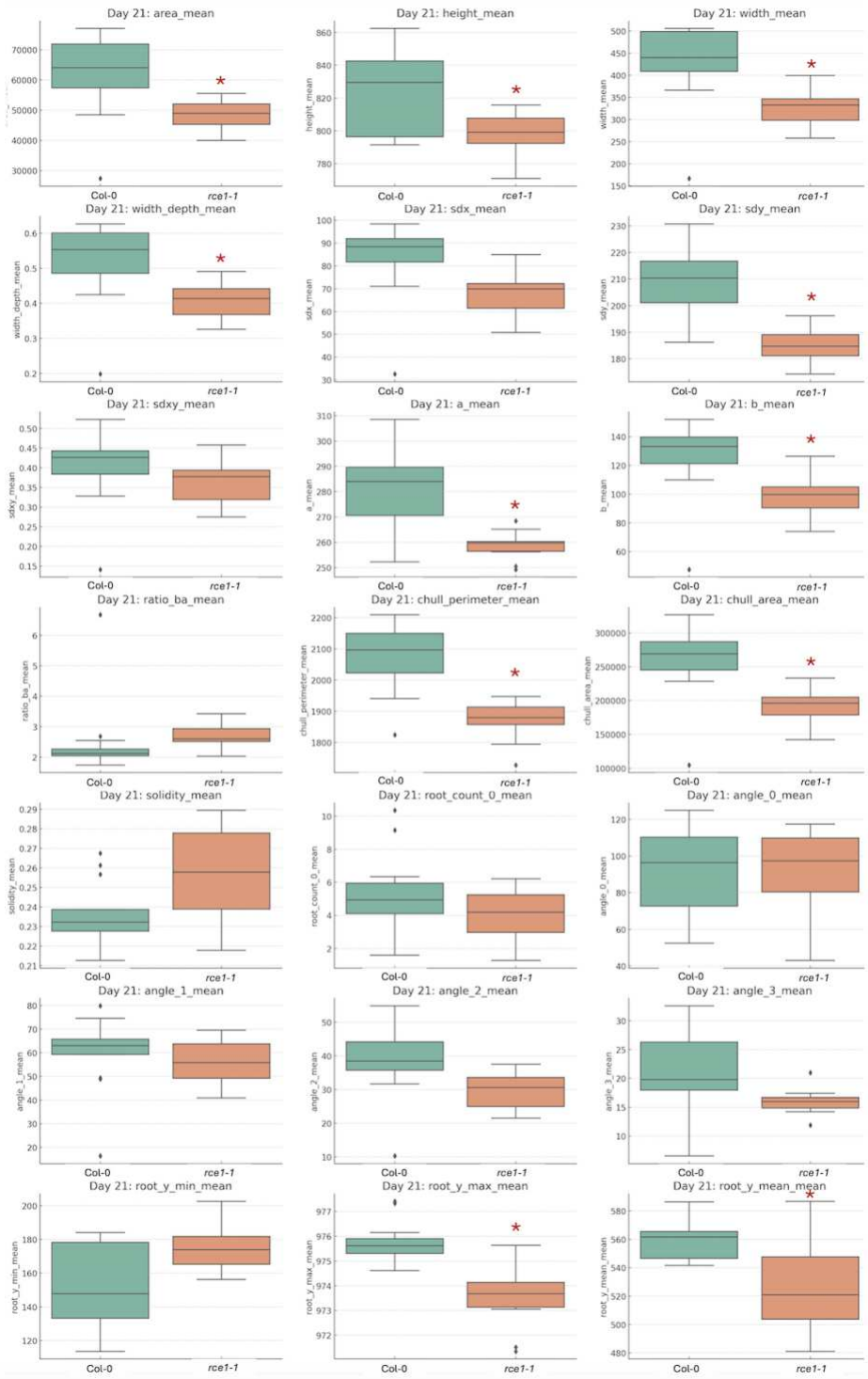

**Supplementary Figure 18. Trait-specific gene set enrichment across root cell types using Random Walk with Restart (RWR) analysis.** Boxplots depict the results of RWR-based gene set enrichment analysis for trait-associated GRIT genes across individual root cell-type subnetworks within the soybean gene co-expression network (GCN). Each panel corresponds to a specific cell type. Traits analyzed include biomass area, convex hull area, dry root mass, width, lateral root length, primary root tip depth, depth, and total root length. Boxes represent the null distribution of association statistics generated by random sampling. Whiskers indicate the full range of null values, and colored diamonds represent the observed RWR association statistic for each trait–cell type pair. Elevated observed values relative to the null distribution suggest trait-specific enrichment in the corresponding cell-type network.




**Supplementary Figure 19. Trait-specific enrichment of GRIT genes across root cell types using Random Walk with Restart (RWR) analysis.** Boxplots illustrate the association strength between GRIT gene sets and individual root cell-type subnetworks in the soybean gene co-expression network (GCN), for four root traits: biomass area, primary root (PR) tip depth, depth, and total root length. Each panel shows enrichment results across multiple root cell types. Boxes represent the null distribution of association RWR statistics generated by randomized sampling, and whiskers indicate the full range of these distributions. Colored diamonds denote observed RWR association scores for each trait–cell type pair.
Statistical significance is annotated as follows: * $p < 0.01$, ** $p < 0.005$.

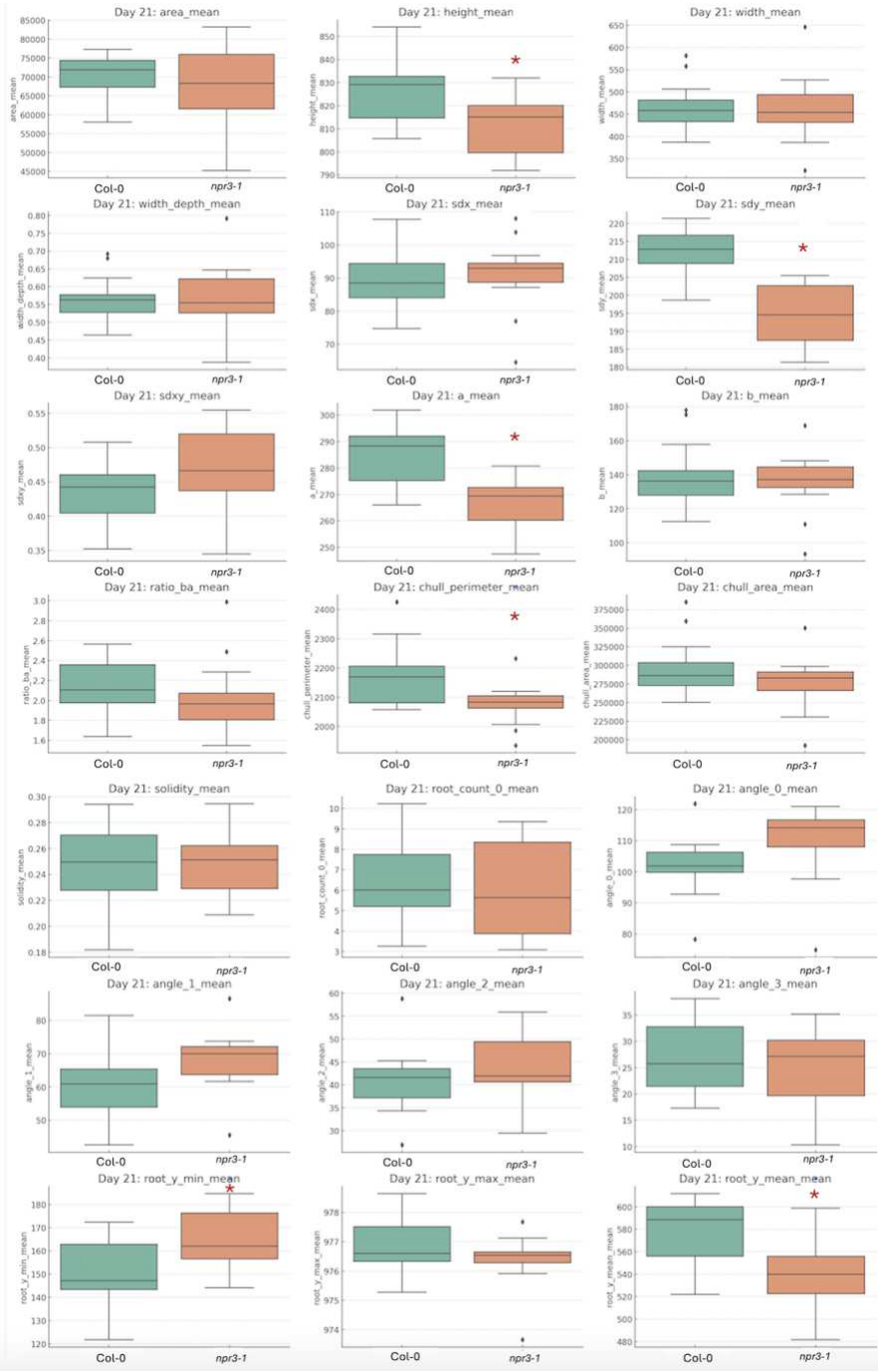



**Supplementary Figure 20. Gene co-expression network highlighting lateral root length-associated genes and cell-type-specific expression.** This network visualization shows the relationships between genes associated with lateral root length (outlined with a blue border) and additional genes identified through genome-wide association studies (GWAS) in soybean. Edges represent co-expression or inferred functional associations between genes. Node fill colors indicate cell-type or tissue-specific expression: Red: Metaphloem; Tan: Xylem pole pericycle (XPP); Blue: Epidermis; Orange: Cortex; and Gray: GWAS-identified genes without specific cell-type enrichment.




**Supplementary Figure 21. Quantification of root system architecture (RSA) traits in the *rce1-1* Arabidopsis mutant.** Box plots summarize 24 RSA traits measured in *rce1-1* mutant and Col-0 wild-type Arabidopsis roots at 21 days after germination. Green boxes represent Col-0, and orange boxes represent *rce1-1*. Each plot shows the distribution of a specific trait, including root system dimensions (e.g., area, width, height), shape descriptors (e.g., convex hull area, solidity, SDX/SDY), and spatial metrics (e.g., angles, root_y coordinates). Statistical comparisons were performed using two-sided Student's *t*-tests followed by Benjamini–Hochberg

false discovery rate (FDR) correction. Red asterisks indicate traits with statistically significant differences between genotypes (*adjusted p* < 0.05).




**Supplementary Figure 22. Quantification of root system architecture (RSA) traits in the *npr3-1* Arabidopsis mutant.** Box plots display 24 RSA traits measured in *npr3-1* mutant and Col-0 wild-type Arabidopsis roots at 21 days post-germination. Each plot compares trait distributions between genotypes, with Col-0 shown in green and *npr3-1* in orange. Traits include root dimensions (e.g., area, width, depth), shape metrics (e.g., solidity, convex hull area/perimeter), and angular or spatial descriptors of root growth (e.g., angles, root_y

coordinates). Statistical significance was assessed using a two-sided Student's *t*-test followed by Benjamini–Hochberg correction for false discovery rate (FDR). Asterisks indicate traits with adjusted *p*-values < 0.05, highlighting significant differences between genotypes.

## References


Apelt, Federico, Eleni Mavrothalassiti, Saurabh Gupta, Frank Machin, Justyna Jadwiga Olas, Maria Grazia Annunziata, Dana Schindelasch, and Friedrich Kragler. 2022. "Shoot and Root Single Cell Sequencing Reveals Tissue- and Daytime-Specific Transcriptome Profiles." *Plant Physiology* 188 (2): 861–78.

Azam, Muhammad, Shengrui Zhang, Jing Li, Muhammad Ahsan, Kwadwo Gyapong Agyenim-Boateng, Jie Qi, Yue Feng, et al. 2023. "Identification of Hub Genes Regulating Isoflavone Accumulation in Soybean Seeds via GWAS and WGCNA Approaches." *Frontiers in Plant Science* 14 (February):1120498.

Battaglia, Marina, and Alejandra A. Covarrubias. 2013. "Late Embryogenesis Abundant (LEA) Proteins in Legumes." *Frontiers in Plant Science* 4 (June):190.

Bawa, George, Zhixin Liu, Xiaole Yu, Aizhi Qin, and Xuwu Sun. 2022. "Single-Cell RNA Sequencing for Plant Research: Insights and Possible Benefits." *International Journal of Molecular Sciences* 23 (9). https://doi.org/10.3390/ijms23094497.

Carles, Cristel C., and Jennifer C. Fletcher. 2009. "The SAND Domain Protein ULTRAPETALA1 Acts as a Trithorax Group Factor to Regulate Cell Fate in Plants." *Genes & Development* 23 (23): 2723–28.

Chandnani, Rahul, Tongfei Qin, Heng Ye, Haifei Hu, Karim Panjvani, Mutsutomo Tokizawa, Javier Mora Macias, et al. 2023. "Application of an Improved 2-Dimensional High-Throughput Soybean Root Phenotyping Platform to Identify Novel Genetic Variants Regulating Root Architecture Traits." *Plant Phenomics (Washington, D.C.)* 5 (September):0097.

Chang, Jennifer. n.d. "WGCNA Gene Correlation Network Analysis." Bioinformatics Workbook. Accessed September 13, 2023. https://bioinformaticsworkbook.org/tutorials/wgcna.html.

Clark, Randy T., Robert B. MacCurdy, Janelle K. Jung, Jon E. Shaff, Susan R. McCouch, Daniel J. Aneshansley, and Leon V. Kochian. 2011. "Three-Dimensional Root Phenotyping with a Novel Imaging and Software Platform." *Plant Physiology* 156 (2): 455–65.

Cliff, Ashley, Jonathon Romero, David Kainer, Angelica Walker, Anna Furches, and Daniel Jacobson. 2019. "A High-Performance Computing Implementation of Iterative Random Forest for the Creation of Predictive Expression Networks." *Genes* 10 (12). https://doi.org/10.3390/genes10120996.

De Smet, Ive, Takuya Tetsumura, Bert De Rybel, Nicolas Frei dit Frey, Laurent Laplaze, Ilda Casimiro, Ranjan Swarup, et al. 2007. "Auxin-Dependent Regulation of Lateral Root Positioning in the Basal Meristem of Arabidopsis." *Development* 134 (4): 681–90.

Dhanapal, Arun Prabhu, Larry M. York, Kasey A. Hames, and Felix B. Fritschi. 2020. "Genome-Wide Association Study of Topsoil Root System Architecture in Field-Grown Soybean [Glycine Max (L.) Merr.]." *Frontiers in Plant Science* 11:590179.

Dharmasiri, Sunethra, Nihal Dharmasiri, Hanjo Hellmann, and Mark Estelle. 2003. "The RUB/Nedd8 Conjugation Pathway Is Required for Early Development in Arabidopsis." *The EMBO Journal* 22 (8): 1762–70.

Ding, Yuli, Tongjun Sun, Kevin Ao, Yujun Peng, Yaxi Zhang, Xin Li, and Yuelin Zhang. 2018. "Opposite Roles of Salicylic Acid Receptors NPR1 and NPR3/NPR4 in Transcriptional Regulation of Plant Immunity." *Cell* 173 (6): 1454–67.e15.

Dorrity, Michael W., Cristina M. Alexandre, Morgan O. Hamm, Anna-Lena Vigil, Stanley Fields, Christine Queitsch, and Josh T. Cuperus. 2021. "The Regulatory Landscape of Arabidopsis

Thaliana Roots at Single-Cell Resolution." *Nature Communications* 12 (1): 3334.
Enderle, Janina, Annika Dorn, Natalja Beying, Oliver Trapp, and Holger Puchta. 2019. "The Protease WSS1A, the Endonuclease MUS81, and the Phosphodiesterase TDP1 Are Involved in Independent Pathways of DNA-Protein Crosslink Repair in Plants." *The Plant Cell* 31 (4): 775–90.
Faggiani Dias, Daniela, Ryan Hanna, Jeffrey Sachnik, Yangyang Xu, Jack Gilbert, Wolfgang Busch, and David G. Victor. 2025. "Removing Atmospheric $CO_2$ through Mass Scaleup of Crops with Enhanced Root Systems." *Environmental Research Letters*, March. https://doi.org/10.1088/1748-9326/adc31b.
Falk, Kevin G., Talukder Zaki Jubery, Jamie A. O'Rourke, Arti Singh, Soumik Sarkar, Baskar Ganapathysubramanian, and Asheesh K. Singh. 2020. "Soybean Root System Architecture Trait Study through Genotypic, Phenotypic, and Shape-Based Clusters." *Plant Phenomics (Washington, D.C.)* 2020 (June):1925495.
Fanfani, Viola, Fabio Cassano, and Giovanni Stracquadanio. 2020. "PyGNA: A Unified Framework for Geneset Network Analysis." *BMC Bioinformatics* 21 (1): 476.
Fu, Zheng Qing, Shunping Yan, Abdelaty Saleh, Wei Wang, James Ruble, Nodoka Oka, Rajinikanth Mohan, et al. 2012. "NPR3 and NPR4 Are Receptors for the Immune Signal Salicylic Acid in Plants." *Nature* 486 (7402): 228–32.
Gao, Huihui, Yan Wang, Wei Li, Yongzhe Gu, Yongcai Lai, Yingdong Bi, and Chaoying He. 2018. "Transcriptomic Comparison Reveals Genetic Variation Potentially Underlying Seed Developmental Evolution of Soybeans." *Journal of Experimental Botany* 69 (21): 5089–5104.
Ge, Haiman, Qiaolin Shao, Jinlin Chen, Jiahong Chen, Xueqin Li, Yu Tan, Wenzhi Lan, Lei Yang, and Yuan Wang. 2022. "A Metal Tolerance Protein, MTP10, Is Required for the Calcium and Magnesium Homeostasis in Arabidopsis." *Plant Signaling & Behavior* 17 (1): 2025322.
Geng, Zhi, Jun Chen, Bo Lu, Fuyuan Zhang, Ziping Chen, Yujun Liu, Chao Xia, et al. 2023. "A Review: Systemic Signaling in the Regulation of Plant Responses to Low N, P and Fe." *Plants* 12 (15). https://doi.org/10.3390/plants12152765.
Geraldo, Nuno, Isabel Bäurle, Shin-Ichiro Kidou, Xiangyang Hu, and Caroline Dean. 2009. "FRIGIDA Delays Flowering in Arabidopsis via a Cotranscriptional Mechanism Involving Direct Interaction with the Nuclear Cap-Binding Complex." *Plant Physiology* 150 (3): 1611–18.
Graeff, Moritz, and Christian S. Hardtke. 2021. "Metaphloem Development in the Arabidopsis Root Tip." *Development* 148 (18). https://doi.org/10.1242/dev.199766.
Guillotin, Bruno, Ramin Rahni, Michael Passalacqua, Mohammed Ateequr Mohammed, Xiaosa Xu, Sunil Kenchanmane Raju, Carlos Ortiz Ramírez, et al. 2023. "A Pan-Grass Transcriptome Reveals Patterns of Cellular Divergence in Crops." *Nature* 617 (7962): 785–91.
Guo, Bingfu, Liping Sun, Siqi Jiang, Honglei Ren, Rujian Sun, Zhongyan Wei, Huilong Hong, et al. 2022. "Soybean Genetic Resources Contributing to Sustainable Protein Production." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 135 (11): 4095–4121.
Hardtke, Christian S. 2023. "Phloem Development." *The New Phytologist* 239 (3): 852–67.
Huot, Bethany, Jian Yao, Beronda L. Montgomery, and Sheng Yang He. 2014. "Growth-Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness." *Molecular Plant* 7 (8): 1267–87.
Huynh-Thu, Vân Anh, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. 2010. "Inferring Regulatory Networks from Expression Data Using Tree-Based Methods." *PloS One* 5 (9). https://doi.org/10.1371/journal.pone.0012776.
Hyten, David L., Ik-Young Choi, Qijian Song, Randy C. Shoemaker, Randall L. Nelson, Jose M.

Costa, James E. Specht, and Perry B. Cregan. 2007. “Highly Variable Patterns of Linkage Disequilibrium in Multiple Soybean Populations.” *Genetics* 175 (4): 1937–44.
Imaizumi, Takato. 2010. “Arabidopsis Circadian Clock and Photoperiodism: Time to Think about Location.” *Current Opinion in Plant Biology* 13 (1): 83–89.
Iyer-Pascuzzi, Anjali S., Olga Symonova, Yuriy Mileyko, Yueling Hao, Heather Belcher, John Harer, Joshua S. Weitz, and Philip N. Benfey. 2010. “Imaging and Analysis Platform for Automatic Phenotyping and Trait Ranking of Plant Root Systems.” *Plant Physiology* 152 (3): 1148–57.
Jagadeesh, Karthik A., Kushal K. Dey, Daniel T. Montoro, Rahul Mohan, Steven Gazal, Jesse M. Engreitz, Ramnik J. Xavier, Alkes L. Price, and Aviv Regev. 2022. “Identifying Disease-Critical Cell Types and Cellular Processes by Integrating Single-Cell RNA-Sequencing and Human Genetics.” *Nature Genetics* 54 (10): 1479–92.
Jhan, Li-Hsin, Chin-Ying Yang, Chih-Min Huang, Mu-Chien Lai, Yen-Hsiang Huang, Supaporn Baiya, and Chung-Feng Kao. 2023. “Integrative Pathway and Network Analysis Provide Insights on Flooding-Tolerance Genes in Soybean.” *Scientific Reports* 13 (1): 1980.
Jia, Peilin, Ruifeng Hu, Fangfang Yan, Yulin Dai, and Zhongming Zhao. 2022. “scGWAS: Landscape of Trait-Cell Type Associations by Integrating Single-Cell Transcriptomics-Wide and Genome-Wide Association Studies.” *Genome Biology* 23 (1): 220.
Jourquin, Joris, Hidehiro Fukaki, and Tom Beeckman. 2020. “Peptide-Receptor Signaling Controls Lateral Root Development.” *Plant Physiology* 182 (4): 1645–56.
Kim, Seong-Hoon, Rupesh Tayade, Byeong-Hee Kang, Bum-Soo Hahn, Bo-Keun Ha, and Yoon-Ha Kim. 2023. “Genome-Wide Association Studies of Seven Root Traits in Soybean (Glycine Max L.) Landraces.” *International Journal of Molecular Sciences* 24 (1). https://doi.org/10.3390/ijms24010873.
Ko, Dae Kwan, and Federica Brandizzi. 2020. “Network-Based Approaches for Understanding Gene Regulation and Function in Plants.” *The Plant Journal: For Cell and Molecular Biology* 104 (2): 302–17.
Kong, Xiangpei, Chunlei Zhang, Huihui Zheng, Min Sun, Feng Zhang, Mengyue Zhang, Fuhao Cui, et al. 2020. “Antagonistic Interaction between Auxin and SA Signaling Pathways Regulates Bacterial Infection through Lateral Root in Arabidopsis.” *Cell Reports* 32 (8): 108060.
Lam, Hon-Ming, Xun Xu, Xin Liu, Wenbin Chen, Guohua Yang, Fuk-Ling Wong, Man-Wah Li, et al. 2010. “Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection.” *Nature Genetics* 42 (12): 1053–59.
Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” *BMC Bioinformatics* 9 (December):559.
LaRue, Therese, Heike Lindner, Ankit Srinivas, Moises Exposito-Alonso, Guillaume Lobet, and José R. Dinneny. 2022. “Uncovering Natural Variation in Root System Architecture and Growth Dynamics Using a Robotics-Assisted Phenomics Platform.” *eLife* 11 (September). https://doi.org/10.7554/eLife.76968.
Lee, Travis A., Tatsuya Nobori, Natanella Illouz-Eliaz, Jiaying Xu, Bruce Jow, Joseph R. Nery, and Joseph R. Ecker. 2023. “A Single-Nucleus Atlas of Seed-to-Seed Development in Arabidopsis.” *bioRxiv*. https://doi.org/10.1101/2023.03.23.533992.
Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.
Liu, Zhijian, Xiangying Kong, Yanping Long, Sirui Liu, Hong Zhang, Jinbu Jia, Wenhui Cui, et al. 2023. “Integrated Single-Nucleus and Spatial Transcriptomics Captures Transitional States in Soybean Nodule Maturation.” *Nature Plants* 9 (4): 515–24.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550.
Luo, Changxin, Yumei Shi, and Yun Xiang. 2022. “SNAREs Regulate Vesicle Trafficking during

Root Growth and Development." *Frontiers in Plant Science* 13 (March):853251.
Luo, Wei-Gui, Qi-Wen Liang, Yi Su, Chao Huang, Bei-Xin Mo, Yu Yu, and Lang-Tao Xiao. 2023. "Auxin Inhibits Chlorophyll Accumulation through ARF7-IAA14-Mediated Repression of Chlorophyll Biosynthesis Genes in Arabidopsis." *Frontiers in Plant Science* 14 (April):1172059.
Lynch, Jonathan P. 2022. "Harnessing Root Architecture to Address Global Challenges." *The Plant Journal: For Cell and Molecular Biology* 109 (2): 415–31.
Mehrshahi, Payam, Sabrina Gonzalez-Jorge, Tariq A. Akhtar, Jane L. Ward, Anahi Santoyo-Castelazo, Susan E. Marcus, Aurora Lara-Núñez, et al. 2010. "Functional Analysis of Folate Polyglutamylation and Its Essential Role in Plant Metabolism and Development." *The Plant Journal: For Cell and Molecular Biology* 64 (2): 267–79.
Méteignier, Louis-Valentin, Cécile Lecampion, Florent Velay, Cécile Vriet, Laura Dimnet, Martin Rougée, Christian Breuer, et al. 2022. "Topoisomerase VI Participates in an Insulator-like Function That Prevents H3K9me2 Spreading." *Proceedings of the National Academy of Sciences of the United States of America* 119 (27): e2001290119.
Moerman, Thomas, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. 2019. "GRNBoost2 and Arboreto: Efficient and Scalable Inference of Gene Regulatory Networks." *Bioinformatics* 35 (12): 2159–61.
Ormancey, Mélanie, Aurélie Le Ru, Carine Duboé, Hailing Jin, Patrice Thuleau, Serge Plaza, and Jean-Philippe Combier. 2020. "Internalization of miPEP165a into Arabidopsis Roots Depends on Both Passive Diffusion and Endocytosis-Associated Processes." *International Journal of Molecular Sciences* 21 (7). https://doi.org/10.3390/ijms21072266.
Ornelas-Ayala, Diego, Rosario Vega-León, Emilio Petrone-Mendoza, Adriana Garay-Arroyo, Berenice García-Ponce, Elena R. Álvarez-Buylla, and María de la Paz Sanchez. 2020. "ULTRAPETALA1 Maintains Arabidopsis Root Stem Cell Niche Independently of ARABIDOPSIS TRITHORAX1." *The New Phytologist* 225 (3): 1261–72.
Ou, Yang, Hong Kui, and Jia Li. 2021. "Receptor-like Kinases in Root Development: Current Progress and Future Directions." *Molecular Plant* 14 (1): 166–85.
Pacurar, Daniel Ioan, Monica Lacramioara Pacurar, John Desmond Bussell, Joseli Schwambach, Tiberia Ioana Pop, Mariusz Kowalczyk, Laurent Gutierrez, et al. 2014. "Identification of New Adventitious Rooting Mutants amongst Suppressors of the Arabidopsis Thaliana superroot2 Mutation." *Journal of Experimental Botany* 65 (6): 1605–18.
Parizot, Boris, Laurent Laplaze, Lilian Ricaud, Elodie Boucheron-Dubuisson, Vincent Bayle, Martin Bonke, Ive De Smet, et al. 2008. "Diarch Symmetry of the Vascular Bundle in Arabidopsis Root Encompasses the Pericycle and Is Reflected in Distich Lateral Root Initiation." *Plant Physiology* 146 (1): 140–48.
Passaia, Gisele, Guillaume Queval, Juan Bai, Marcia Margis-Pinheiro, and Christine H. Foyer. 2014. "The Effects of Redox Controls Mediated by Glutathione Peroxidases on Root Architecture in Arabidopsis Thaliana." *Journal of Experimental Botany* 65 (5): 1403–13.
Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4): 417–19.
Perianez-Rodriguez, Juan, Marcos Rodriguez, Marco Marconi, Estefano Bustillo-Avendaño, Guy Wachsman, Alvaro Sanchez-Corrionero, Hugues De Gernier, et al. 2021. "An Auxin-Regulable Oscillatory Circuit Drives the Root Clock in Arabidopsis." *Science Advances* 7 (1): eabd4722.
Petr, D., A. Auton, A. Goncalo, A. Cornelis, and E. Banks. n.d. "1000 Genomes Project Analysis Group. The Variant Call Format and VCF Tools." *Bioinformatics* .
Piya, Sarbottam, Vince Pantalone, Sobhan Bahrami Zadegan, Sarah Shipp, Naoufal Lakhssassi, Dounya Knizia, Hari B. Krishnan, Khalid Meksem, and Tarek Hewezi. 2023.

"Soybean Gene Co-Expression Network Analysis Identifies Two Co-Regulated Gene Modules Associated with Nodule Formation and Development." *Molecular Plant Pathology* 24 (6): 628–36.
Potapova, Nadezhda A., Alexander S. Zlobin, Roman N. Perfil'ev, Gennady V. Vasiliev, Elena A. Salina, and Yakov A. Tsepilov. 2023. "Population Structure and Genetic Diversity of the 175 Soybean Breeding Lines and Varieties Cultivated in West Siberia and Other Regions of Russia." *Plants* 12 (19): 3490.
Pozo, J. C. del, and M. Estelle. 1999. "The Arabidopsis Cullin AtCUL1 Is Modified by the Ubiquitin-Related Protein RUB1." *Proceedings of the National Academy of Sciences of the United States of America* 96 (26): 15342–47.
Prince, Silvas J., Babu Valliyodan, Heng Ye, Ming Yang, Shuaishuai Tai, Wushu Hu, Mackensie Murphy, et al. 2019. "Understanding Genetic Control of Root System Architecture in Soybean: Insights into the Genetic Basis of Lateral Root Number." *Plant, Cell & Environment* 42 (1): 212–29.
Raj, Anil, Matthew Stephens, and Jonathan K. Pritchard. 2014. "fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets." *Genetics* 197 (2): 573–89.
Rodriguez-Villalon, Antia. 2016. "Wiring a Plant: Genetic Networks for Phloem Formation in Arabidopsis Thaliana Roots." *The New Phytologist* 210 (1): 45–50.
Rodriguez-Villalon, Antia, Bojan Gujas, Ringo van Wijk, Teun Munnik, and Christian S. Hardtke. 2015. "Primary Root Protophloem Differentiation Requires Balanced Phosphatidylinositol-4,5-Biphosphate Levels and Systemically Affects Root Branching." *Development* 142 (8): 1437–46.
Salim, Mohammad, Yinglong Chen, Heng Ye, Henry T. Nguyen, Zakaria M. Solaiman, and Kadambot H. M. Siddique. 2021. "Screening of Soybean Genotypes Based on Root Morphology and Shoot Traits Using the Semi-Hydroponic Phenotyping Platform and Rhizobox Technique." *Agronomy* 12 (1): 56.
Schaefer, Robert J., Jean-Michel Michno, Joseph Jeffers, Owen Hoekenga, Brian Dilkes, Ivan Baxter, and Chad L. Myers. 2018. "Integrating Coexpression Networks with GWAS to Prioritize Causal Genes in Maize." *The Plant Cell* 30 (12): 2922–42.
Seck, Waldiodio, Davoud Torkamaneh, and François Belzile. 2020. "Comprehensive Genome-Wide Association Analysis Reveals the Genetic Basis of Root System Architecture in Soybean." *Frontiers in Plant Science* 11 (December):590740.
Shahan, Rachel, Che-Wei Hsu, Trevor M. Nolan, Benjamin J. Cole, Isaiah W. Taylor, Laura Greenstreet, Stephen Zhang, et al. 2022. "A Single-Cell Arabidopsis Root Atlas Reveals Developmental Trajectories in Wild-Type and Cell Identity Mutants." *Developmental Cell* 57 (4): 543–60.e9.
Shahan, Rachel, Trevor M. Nolan, and Philip N. Benfey. 2021. "Single-Cell Analysis of Cell Identity in the Arabidopsis Root Apical Meristem: Insights and Opportunities." *Journal of Experimental Botany* 72 (19): 6679–86.
Slovak, Radka, Takehiko Ogura, Santosh B. Satbhai, Daniela Ristova, and Wolfgang Busch. 2016. "Genetic Control of Root Growth: From Genes to Networks." *Annals of Botany* 117 (1): 9–24.
Smita, Shuchi, Jason Kiehne, Sajag Adhikari, Erliang Zeng, Qin Ma, and Senthil Subramanian. 2020. "Gene Regulatory Networks Associated with Lateral Root and Nodule Development in Soybean." *In Silico Plants* 2 (1): diaa002.
Song, Qingxin, Atsumi Ando, Ning Jiang, Yoko Ikeda, and Z. Jeffrey Chen. 2020. "Single-Cell RNA-Seq Analysis Reveals Ploidy-Dependent and Cell-Specific Transcriptome Changes in Arabidopsis Female Gametophytes." *Genome Biology* 21 (1): 178.
Speeckaert, Nathanaël, Mondher El Jaziri, Marie Baucher, and Marc Behr. 2022. "UGT72, a Major Glycosyltransferase Family for Flavonoid and Monolignol Homeostasis in Plants." *Biology* 11 (3): 441.

Sun, Ying, and José R. Dinneny. 2018. “Q&A: How Do Gene Regulatory Networks Control Environmental Responses in Plants?” *BMC Biology* 16 (1): 38.
Sun, Ying, Dong-Ha Oh, Lina Duan, Prashanth Ramachandran, Andrea Ramirez, Anna Bartlett, Kieu-Nga Tran, Guannan Wang, Maheshi Dassanayake, and José R. Dinneny. 2022. “Divergence in the ABA Gene Regulatory Network Underlies Differential Growth Control.” *Nature Plants* 8 (5): 549–60.
Svennerstam, Henrik, Sandra Jämtgård, Iftikhar Ahmad, Kerstin Huss-Danell, Torgny Näsholm, and Ulrika Ganeteg. 2011. “Transporters in Arabidopsis Roots Mediating Uptake of Amino Acids at Naturally Occurring Concentrations.” *The New Phytologist* 191 (2): 459–67.
Tantardini, Mattia, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. 2019. “Comparing Methods for Comparing Networks.” *Scientific Reports* 9 (1): 17557.
Tarsis, Kristina, Tsvia Gildor, Miri Morgulis, and Smadar Ben-Tabou de-Leon. 2022. “Distinct Regulatory States Control the Elongation of Individual Skeletal Rods in the Sea Urchin Embryo.” *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 251 (8): 1322–39.
Valliyodan, Babu, Anne V. Brown, Juexin Wang, Gunvant Patil, Yang Liu, Paul I. Otyama, Rex T. Nelson, et al. 2021. “Genetic Variation among 481 Diverse Soybean Accessions, Inferred from Genomic Re-Sequencing.” *Scientific Data* 8 (1): 50.
Wachsman, Guy, Jingyuan Zhang, Miguel A. Moreno-Risueno, Charles T. Anderson, and Philip N. Benfey. 2020. “Cell Wall Remodeling and Vesicle Trafficking Mediate the Root Clock in *Arabidopsis*.” *Science* 370 (6518): 819–23.
Wang, Jiabo, and Zhiwu Zhang. 2021. “GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction.” *Genomics, Proteomics & Bioinformatics* 19 (4): 629–40.
Wang, Juexin, Md Shakhawat Hossain, Zhen Lyu, Jeremy Schmutz, Gary Stacey, Dong Xu, and Trupti Joshi. 2019. “SoyCSN: Soybean Context-Specific Network Analysis and Prediction Based on Tissue-Specific Transcriptome Data.” *Plant Direct* 3 (9): e00167.
Wysoker, A., K. Tibbetts, and T. Fennell. n.d. “Picard Tools Version 1.90.” *Jhpsn*.
Yao, Min, Mei Guan, Zhenqian Zhang, Qiuping Zhang, Yixin Cui, Hao Chen, Wei Liu, et al. 2020. “GWAS and Co-Expression Network Combination Uncovers Multigenes with Close Linkage Effects on the Oleic Acid Content Accumulation in Brassica Napus.” *BMC Genomics* 21 (1): 320.
Yao, Yanjie, Erhui Xiong, Xuelian Qu, Junfeng Li, Hongli Liu, Leipo Quan, Wenyan Lu, et al. 2023. “WGCNA and Transcriptome Profiling Reveal Hub Genes for Key Development Stage Seed Size/oil Content between Wild and Cultivated Soybean.” *BMC Genomics* 24 (1): 494.
Yılmaz, Hikmet, Ceyhun Kayıhan, Halis Batuhan Ünal, Oğuzhan Yaprak, and Emre Aksoy. 2023. “Single-Cell Transcriptional Profiling in Arabidopsis Root Exposed to B Toxicity at Seedling Stages.” *bioRxiv*. https://doi.org/10.1101/2023.03.09.531923.
Zhang, Yijuan, Chongmei Zhang, Xiaxia Man, Yihan Men, Xuemei Ren, Xueyin Li, Lida Han, et al. 2023. “Functional Characterization of the SiFPGS2 Gene of Foxtail Millet in Folate Accumulation and Root Development.” *Plant Growth Regulation* 99 (1): 137–47.
Zhao, Chengsong, Réjane Pratelli, Shi Yu, Brett Shelley, Eva Collakova, and Guillaume Pilot. 2021. “Detailed Characterization of the UMAMIT Proteins Provides Insight into Their Evolution, Amino Acid Transport Properties, and Role in the Plant.” *Journal of Experimental Botany* 72 (18): 6400–6417.
Zhao, Chuang, Bing Liu, Shilong Piao, Xuhui Wang, David B. Lobell, Yao Huang, Mengtian Huang, et al. 2017. “Temperature Increase Reduces Global Yields of Major Crops in Four Independent Estimates.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (35): 9326–31.
Zheng, Dihuai, Jiwei Xu, Yaqian Lu, Hongyu Chen, Qinjie Chu, and Longjiang Fan. 2023.

"Recent Progresses in Plant Single-Cell Transcriptomics." *Crop Design* 2 (2): 100041.
Zheng, Ruiqing, Min Li, Xiang Chen, Fang-Xiang Wu, Yi Pan, and Jianxin Wang. 2019. "BiXGBoost: A Scalable, Flexible Boosting-Based Method for Reconstructing Gene Regulatory Networks." *Bioinformatics* 35 (11): 1893–1900.
Zhu, Jie, Signe Lolle, Andrea Tang, Bella Guel, Brian Kvitko, Benjamin Cole, and Gitta Coaker. 2023. "Single-Cell Profiling of Arabidopsis Leaves to Pseudomonas Syringae Infection." *Cell Reports* 42 (7): 112676.
Zhu, Xiang, Zhana Duren, and Wing Hung Wong. 2021. "Modeling Regulatory Network Topology Improves Genome-Wide Analyses of Complex Human Traits." *Nature Communications* 12 (1): 2851.

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SoyGWASSupplementaryTables051625.xlsx