New Phytologist

Research

Genome-wide association study and network analysis of *in vitro* transformation in *Populus trichocarpa* support key roles of diverse phytohormone pathways and cross talk

Michael F. Nagle[1], Jialin Yuan[2], Damanpreet Kaur[2], Cathleen Ma[1], Ekaterina Peremyslova[1], Yuan Jiang[3], Greg S. Goralogia[1], Anna Magnuson[1], Jia Yi Li[2], Wellington Muchero[4,5,6], Li Fuxin[2] and Steven H. Strauss[1]

[1]Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR 97331, USA; [2]School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331, USA; [3]Statistics Department, Oregon State University, Corvallis, OR 97331, USA; [4]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA; [5]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA; [6]Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville, TN 37996, USA

Author for correspondence:
Steven H. Strauss
Email: steve.strauss@oregonstate.edu

Received: *17 November 2023*
Accepted: *6 March 2024*

New Phytologist (2024)
doi: 10.1111/nph.19737

Key words: genome-wide association study (GWAS), network analysis, *Populus*, regeneration, transformation.

Summary

- Wide variation in amenability to transformation and regeneration (TR) among many plant species and genotypes presents a challenge to the use of genetic engineering in research and breeding. To help understand the causes of this variation, we performed association mapping and network analysis using a population of 1204 wild trees of *Populus trichocarpa* (black cottonwood).
- To enable precise and high-throughput phenotyping of callus and shoot TR, we developed a computer vision system that cross-referenced complementary red, green, and blue (RGB) and fluorescent-hyperspectral images. We performed association mapping using single-marker and combined variant methods, followed by statistical tests for epistasis and integration of published multi-omic datasets to identify likely regulatory hubs.
- We report 409 candidate genes implicated by associations within 5 kb of coding sequences, and epistasis tests implicated 81 of these candidate genes as regulators of one another. Gene ontology terms related to protein–protein interactions and transcriptional regulation are over-represented, among others.
- In addition to auxin and cytokinin pathways long established as critical to TR, our results highlight the importance of stress and wounding pathways. Potential regulatory hubs of signaling within and across these pathways include *GROWTH REGULATORY FACTOR 1* (*GRF1*), *PHOSPHATIDYLINOSITOL 4-KINASE β1* (*PI-4Kβ1*), and *OBF-BINDING PROTEIN 1* (*OBP1*).

Introduction

By enabling the introduction, transfer and manipulation of genetic material within and across genomes, plant transformation has become an important tool for scientific research and plant improvement. *Agrobacterium* is commonly used as a vector to introduce foreign DNA to plant genomes, followed by *in vitro* tissue culture methods to promote regeneration of whole plants. These treatments often employ a high-auxin, low-cytokinin media to promote the formation of dedifferentiated and pluripotent 'callus,' followed by high-cytokinin, low-auxin media to promote redifferentiation of pluripotent tissue into shoot (Thorpe, 2007). However, effective treatments often vary widely across species and genotypes, a challenge that remains a limiting factor in the application of plant transformation and regeneration (TR; Altpeter *et al.*, 2016).

Similarly to other developmental processes in plants, regeneration depends on phytohormones and complex polygenic networks involving transcriptional regulation, protein–protein interaction, post-translational modifications, and other means of gene–gene interactions functioning both upstream and downstream of the phytohormones (Ikeuchi *et al.*, 2018). Auxin is a canonical regulator of cell expansion, long believed to function largely via downstream proton pumps that influence pH-dependent action of enzymes catalyzing loosening of cell walls (Hager, 2003; Majda & Robert, 2018). In contrast to auxin's patterns of localization (Taiz *et al.*, 2015), cytokinin is a canonical regulator of cell division via mechanisms including regulation of cyclin expression (Argueso & Kieber, 2024). Cross talk between auxin and cytokinin signaling includes both antagonistic and cooperative relationships through numerous points of interaction spanning hormone synthesis, transport, and

downstream regulatory cascades, encouraging a view of the complete auxin-cytokinin axis as a regulator of both cell division and expansion (Kotov & Kotova, 2023).

When plant tissue is wounded, either in nature or *in vitro* tissue culture, the auxin/cytokinin axis can be activated by upstream wound response signaling directed by jasmonates (JA; Zhang *et al.*, 2023) in concert with ethylene (Bouchez *et al.*, 2007), while salicylic acid (SA) signaling antagonizes this process (Li *et al.*, 2019). Highly relevant intersections between the JA/ethylene/SA axis and cytokinin/auxin axis include regulation of *ENHANCER OF SHOOT REGENERATION 2* and an auxin biosynthesis gene by *ETHYLENE RESPONSE FACTOR 1* (Mao *et al.*, 2016; Ikeuchi *et al.*, 2018). Competing and antagonistic roles of JA/ethylene and SA likely evolved in response to a need for distinct defense mechanisms against biotrophic and necrotrophic pathogens, in both cases with precise control of necrosis of infected and nearby tissues being essential for the plant to contain infection (Spoel *et al.*, 2007). In the context of *in vitro* regeneration, the wounding of explants (excised plant tissues) may lead to either necrosis or regeneration at the site of wounding, with the patterns of necrosis and regeneration being highly genotype-dependent, including in poplar (Ma *et al.*, 2022a,b). Moreover, a third key phytohormone axis involves the competitive actions of gibberellins and abscisic acid (ABA), often studied in the context of their roles in seed and vegetative development, as well as in abiotic stress response (Shu *et al.*, 2018). This axis is also highly relevant to *in vitro* regeneration for reasons including the action of downstream DELLA transcription factors that mediate cross talk with the auxin-cytokinin and JA/ethylene/SA axes (Davière & Achard, 2016).

Numerous genes have been identified that can enhance the ability of plants to undergo regeneration when their expression is manipulated via overexpression or targeted expression; they are often referred to as 'morphogenic regulators' (MRs). These are typically transcription factors closely connected to the auxin-cytokinin axis (Ikeuchi *et al.*, 2018), with genes including *WUSCHEL* (Gallois *et al.*, 2002; Arroyo-Herrera *et al.*, 2008; Bouchabké-Coussa *et al.*, 2013; Lowe *et al.*, 2016; Kadri *et al.*, 2021; Lou *et al.*, 2022; Nelson-Vasilchik *et al.*, 2022; Pan *et al.*, 2022) and the *WUSCHEL-ASSOCIATED HOMEOBOX* (*WOX*) family (Liu *et al.*, 2018; Hassani *et al.*, 2022; Pan *et al.*, 2022); AP2/ERF transcription factors such as *BABY BOOM* (Boutilier *et al.*, 2002; Srinivasan *et al.*, 2007; Deng *et al.*, 2009; Heidmann *et al.*, 2011; Lutz *et al.*, 2015; Lowe *et al.*, 2016; Horstman *et al.*, 2017; Nelson-Vasilchik *et al.*, 2022); and the related *PLETHORA (PLT)* gene family (Kareem *et al.*, 2015; Lian *et al.*, 2022), among others (reviewed by Lee & Wang, 2023). More recently, *GROWTH REGULATORY FACTOR* (*GRF*) family genes have demonstrated potential to enhance regeneration in at least seven species (Debernardi *et al.*, 2020; Luo & Palmgren, 2021; Lee & Wang, 2023), and are distinctive in their negative regulation by auxin (Beltramino *et al.*, 2021), induction downstream of gibberellins (van der Knaap *et al.*, 2000) via DELLA transcriptional repressors (Lantzouni *et al.*, 2020), and direct or indirect regulation of gene expression and phytohormone concentrations across diverse pathways (Piya *et al.*, 2020). Despite the many successes in using MRs to enhance regeneration, the efficacies of these treatments are genotype-dependent, sometimes with modest, negative, or undetectable benefit in certain genotypes or species (Klimaszewska *et al.*, 2010; Hu *et al.*, 2016; Lowe *et al.*, 2016; Mookkan *et al.*, 2017; Aregawi *et al.*, 2020; Hoerster *et al.*, 2020; Che *et al.*, 2022; Ryan, 2022).

The genetic discovery of additional MRs, particularly those involving phytohormone cross talk and signaling outside of the auxin-cytokinin axis, may be essential for TR systems that are robust across diverse species, genotypes, and tissue culture environments. Association mapping, such as with genome-wide association studies (GWAS), is a prominent approach to discover genetic variants controlling important agronomic traits (Cortes *et al.*, 2021). This approach has been used to a limited extent to study rates of *in vitro* regeneration; however, population sizes and thus the statistical power of those studies has been limited (Tuskan *et al.*, 2018; Lardon & Geelen, 2020; Nguyen *et al.*, 2020; Zhang *et al.*, 2020; Dai *et al.*, 2022), and none have studied regeneration of transgenic tissues.

Network analysis methods involving integration of multi-omic datasets can provide powerful insights into regulatory modules and hubs, such as was demonstrated by a recent study in wheat that revealed epigenetic regulators and transcription factors had major influences on hormone response during TR (Liu *et al.*, 2023). The study of transformation introduces additional challenges beyond those of regeneration alone, as the use of gene transfer treatments such as wounding-associated *Agrobacterium* co-cultivation followed by antibiotic selection produce additional complex biological responses. In addition, transgenic tissues often develop slowly through multiple stages (e.g. microcallus, macrocallus, and shoot), thus phenotypes must be determined over time while maintaining sterile conditions. In addition, transgenic tissue size and frequency is challenging to quantify, especially in early stages of development. For these reasons, we spent several years developing a phenomic system for measuring transgenic tissue growth and regeneration. We previously reported a GWAS of *in planta* regeneration of *Populus trichocarpa* (black cottonwood) (Tuskan *et al.*, 2006) absent transformation, utilizing a red, green, and blue (RGB) computer vision system for high-throughput measurement of callus and shoot during regeneration (M. F. Nagle *et al.*, 2024). In the present work, we built upon this RGB-based methodology by adding fluorescent-hyperspectral imaging, which enabled the efficient use of fluorescent proteins as positive markers of transformation and thus enabled us to determine when regenerating tissues were either transgenic or escapes from negative selection (i.e. from antibiotics used in tissue culture media).

Populus trichocarpa represents an ideal model to study *in vitro* TR in a GWAS context due to a wealth of genomic resources, as well as detailed *in vitro* culture (Kang *et al.*, 2009; Ma *et al.*, 2022a,b) and transformation methods (Han *et al.*, 1997; Ma *et al.*, 2004; Song *et al.*, 2006; Li *et al.*, 2015). A diverse population of 1323 re-sequenced poplar clones, collected from across its natural range in northwestern North America, provides high levels of genetic diversity and statistical power for analysis. A single-nucleotide polymorphism (SNP) dataset for this population features over 7 million SNPs with polymorphism in at least 5% of genotypes and nonmissing data in at least 90% of

genotypes. The density of these SNPs relative to genes is also high, with a mean frequency of one SNP every *c.* 50 bases on contiguous assembled chromosomes. In addition, linkage disequilibrium (LD) decays rapidly to 0.2 (R^2) within 2 kb, enabling powerful and high-resolution association mapping. Portions of this study population have been previously used to study diverse traits including *in vitro* callus regeneration (Tuskan *et al.*, 2018), *in planta* callus and shoot regeneration (M. F. Nagle *et al.*, 2024), cell wall development (Furches *et al.*, 2019), and metabolic traits (Zhang *et al.*, 2018), among others (Muchero *et al.*, 2018; Bdeir *et al.*, 2019; Chhetri *et al.*, 2019, 2020; Weighill *et al.*, 2019; Nagle *et al.*, 2023).

Here, we report over 400 candidate genes, with gene ontology overrepresentation of transcriptional regulators and protein-binding proteins (among other categories), providing evidence for the role of a complex genetic regulatory network (GRN) that controls regeneration of transgenic tissues. Integration of published high-throughput datasets, including differentially expressed genes, promoter binding, and protein–protein interactions, indicated that over 200 of our candidate genes are members of a GRN involving diverse phytohormone signaling pathways and cross talk thereof. Our network analysis, alongside epistasis tests that help to validate and extend our GWAS results, indicate that many candidate genes are members of contiguous GRNs that span diverse pathways involving hormone, wound and biotic stress response.

Materials and Methods

Plant materials

We examined *in vitro* TR in clones from a re-sequenced clone bank of *P. trichocarpa* (Tuskan *et al.*, 2018; Zhang *et al.*, 2018; Bdeir *et al.*, 2019; Chhetri *et al.*, 2019, 2020; Furches *et al.*, 2019; Weighill *et al.*, 2019; M. F. Nagle *et al.*, 2024), with 1307 clones available at a field location in Corvallis, OR. Cuttings from the field were propagated in glasshouse conditions, and plants grown in the glasshouse were used to provide explant material to undergo TR (Supporting Information Methods S1). Stem explants underwent sterilization (as detailed in Table S1) and were transformed using an enhanced Green Fluorescent Protein (eGFP) reporter plasmid with a minimal matrix attachment region (Fig. S1), introduced via a proprietary strain of *Agrobacterium* (LBA4404-derived) featuring thymidine auxotrophy (*thy*$^-$, provided courtesy of Corteva Agrisciences). All *Agrobacterium* was treated with an induction media featuring acetosyringone (AS), and each genotype was cultured both with and without AS added to the subsequent co-cultivation media. Following co-cultivation, stem explants underwent callus induction and shoot induction media treatments (Methods S2). Given the large number of genotypes, tissue culturing and phenotyping were conducted across 69 phases spanning roughly 20 months. Contamination of tissue cultures was an ongoing challenge that required our sterilization procedures to evolve over the course of this multiple-year study. Phase identifiers were recorded such that they could be used as covariates in downstream statistical models (Table S1; Methods S1–S9).

Phenotyping with RGB and hyperspectral computer vision

We developed a computer vision workflow where semantic segmentation of RGB images provided maps of tissue identity (e.g. callus, shoot and unregenerated stem) and CubeGLM analysis of fluorescence hyperspectral data provided maps of eGFP signal. These two data layers were aligned, stacked, and used to provide an array of TR statistics via the GMOdetector workflow (https://github.com/naglemi/GMOnotebook). We collected images using a custom-designed RGB and hyperspectral imaging platform (macroPhor Array imager from Middleton Spectral Vision, Middleton, WI, USA). Fluorescent-hyperspectral images were collected using a single-frequency 488 nm laser as an excitation source to enable excitation of reporter proteins (such as eGFP used here, and others) as well as chlorophyll (Methods S3). In addition, our models included a 'noise' coefficient that controlled for autofluorescence and enhanced our ability to distinguish low levels of reporter protein signal from autofluorescence, which might otherwise lead to false positives when relying solely on fluorescent microscopy with filters. RGB image segmentation was enabled by a training set prepared using our previously reported IDEAS web-based image annotation interface (Yuan *et al.*, 2022) and by transfer learning to retrain a DeepLabv3+ model (Chen *et al.*, 2018) previously used in our GWAS of *in planta* regeneration (M. F. Nagle *et al.*, 2024). The retrained model was then deployed to segment 19 961 Petri dish images into background, callus, shoot, contaminated/necrotic tissue, and unregenerated stem explant material. Of note, we believe that the category of unregenerated stem includes microcalli that are not visible at the macroscopic scale of imaging.

To provide an additional filter for contaminated material, we also trained a DenseNet binary classification model (Huang *et al.*, 2018) to filter markedly contaminated explants and spots with no explants (removed during subculturing due to contamination). eGFP signal was quantified on a per-pixel basis using our CubeGLM Python package (https://github.com/naglemi/cubeglm/) to project hypercubes (3D array representations of hyperspectral images) onto design matrices of eGFP, chlorophyll, and noise components represented by standard spectra obtained from KemoQuant (Middleton Spectral Vision).

To enable cross-referencing of RGB and hyperspectral data collected from different cameras, these two image layers were aligned using a homography transformation. This transformation involved the imaging of grids with both RGB and hyperspectral cameras, followed by the selection of matching points on complementary images and the computation of transformation matrices using the selected indices. Once computed, these transformation matrices were applied on a high-throughput scale for alignment of all images collected with the same camera positions and settings.

We sought to classify given tissues and explants as transgenic or not based on whether hyperspectral signal indicative of a fluorescent reporter protein met a given x threshold in at least y pixels of hyperspectral images. To determine appropriate parameters for x and y, we performed fluorescent microscopy to examine 4423 explants that underwent transformation and recorded whether each appeared to display fluorescent protein or not. Hyperspectral

weights were computed as previously described and images were divided into 12 portions, as each plate featured 12 explants in specific positions. We next constructed a cost function based on the difference between true positive rates and false positive rates and performed a parameter sweep, finding that parameters of 3 pixels meeting a reporter threshold of 38 performed well with a true positive rate of 79.0% and false positive rate of 10.4%. Other parameter sets with similarly low cost are given on the GitHub; the ideal parameter set for the user depends on their level of interest in very small areas of signal (few pixels with fluorescent reporter signal) and tolerance for false positives. Manual inspection of putatively errant classifications, via microscopy and false-color visualization of hyperspectral data with CUBEGLM, indicated that all observed disagreements were due to explant overlap, human error, or microcalli that were more readily visible under microscopy than with hyperspectral imaging (Fig. 1).

The various modules for image analysis were integrated to produce the GMODETECTOR pipeline for high-throughput TR phenotyping (Fig. 2; code and documentation at https://github.com/naglemi/GMOnotebook). The types of trait statistics obtained from this workflow are described in Table 1. Out of 19 961 samples, 16 107 passed quality control filtering criteria and were deemed suitable for downstream analysis. A total of 1266 genotypes were represented in at least one phenotype dataset (Table 1), while the remainder of the 1307 genotypes from the Corvallis replicate of the clone bank were not represented for reasons including failure to thrive in the glasshouse, significant fungal/bacterial contamination of all replicates, and/or technical errors during data collection. Further details for phenotyping are given in Methods S4. Visual examples of regeneration phenotypes across all experimental timepoints (week 3, week 7 and week 10) are given in Fig. S2.

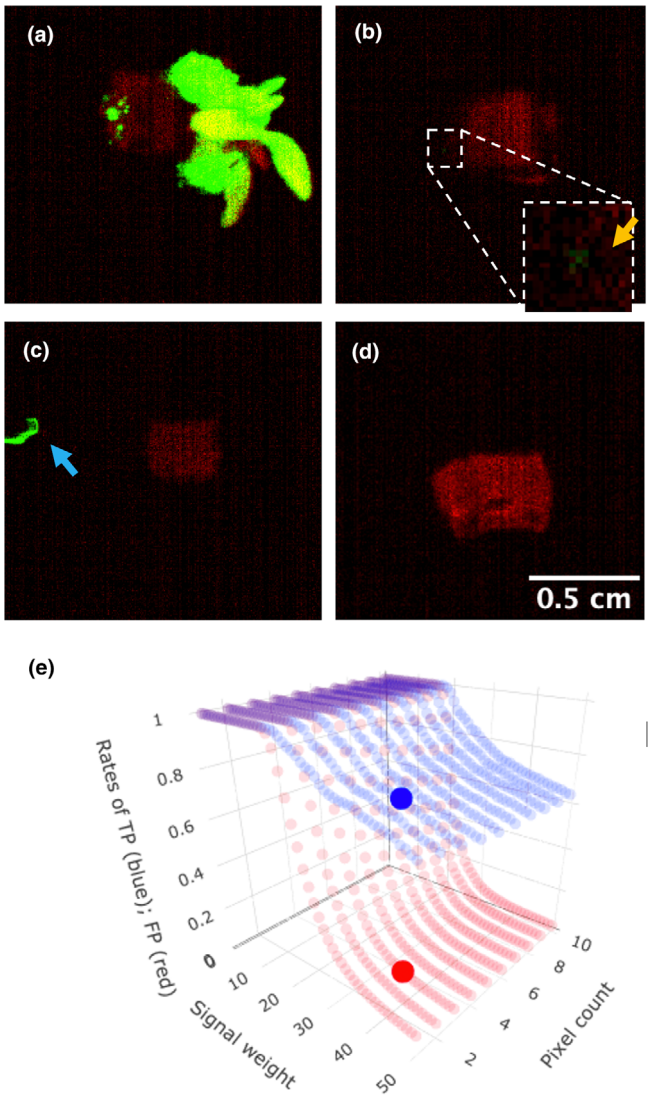

Fig. 1 Classification of explants as transgenic or not based on thresholding with fluorescent reporter weights. (a–d) Results are shown for selected parameters (transgenic if 3 pixels have fluorescent reporter weight of 38 or greater), which were first obtained by tuning for DsRed and then confirmed to also work for enhanced Green Fluorescent Protein (eGFP). Hyperspectral image data for explants is shown in false color, with DsRed weight in green and chlorophyll weight in red. The model was trained with a ground truth dataset based on human classifying explants as transgenic or not using fluorescent microscopy. (a) True positives typically displayed clear and strong reporter signal in callus and/or shoot. (b) False negatives typically featured very small areas of reporter signal, which were more readily identified by microscopy than our macroscopic hyperspectral system. A box (dotted line) displays a 'zoomed-in' view of an area with few pixels (orange arrow) displaying reporter signal. (c) False positives were found to represent cases of one explant growing into a grid position assigned to an adjacent explant on the same Petri dish (blue arrow) in all cases inspected. (d) True negatives displayed no fluorescent signal. (e) We performed a parameter sweep to identify parameters that worked well to provide high true positive rates (blue) with low false positive rates (red). Parameters include thresholds for weight of fluorescent signal intensity representative of the reporter protein (signal weight, *x*-axis) and for the number of pixels with this signal weight (pixel count, *z*-axis). Datapoints for the selected heuristic (pixel count of 3 and signal weight of 38) are enlarged and opaque. This plot was produced with PLOTLY in R.

Statistical analysis and network inference

We conducted a principal component analysis (PCA) across five trait groupings, employing unit vector scaling or Pareto scaling as appropriate. PCA was performed using PRCOMP in R (Methods S5) and scree plots (Figs S3–S7) were consulted to determine which principal components (PCs) explained major portions of variance and were thus used for downstream association mapping. Association mapping of computer vision traits (Table 1) and PCs derived from them was performed by mirroring methods from our prior studies (Nagle *et al.*, 2023; M. F. Nagle *et al.*, 2024). Our first study using these methods, for GWAS of *in planta* regeneration in poplar, includes a detailed explanation of the methods and comparisons of their relative advantages and disadvantages (M. F. Nagle *et al.*, 2024). To summarize, GWAS methods used include Genome-wide Efficient Mixed Model Association (GEMMA; Zhou & Stephens, 2012) for continuous traits transformed toward normality, Generalized Mixed Model Association Test (GMMAT; Chen *et al.*, 2016) for binarized traits, and SNP set (Sequence) Kernel Association Test (SKAT; Ionita-Laza *et al.*, 2013 with our Multi-Threaded Monte Carlo SKAT (MTMC-SKAT)) R extension to efficiently compute permutation-based empirical *P*-values for continuous traits without transformation. MTMC-SKAT was

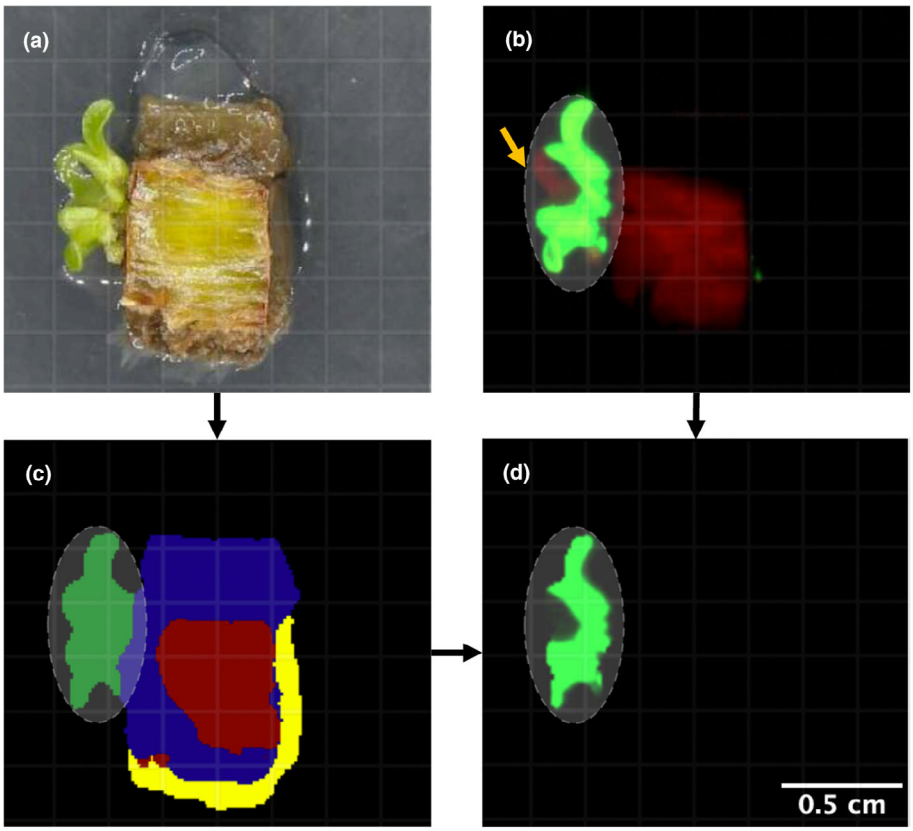

Fig. 2 GMODETECTOR integration of red, green, and blue (RGB)-based segmentation and hyperspectral analysis to measure frequencies of transformation and regeneration of specific tissues. (a) Raw RGB image data for a given explant. (b) False-color representation of hyperspectral image data for given explant, where green and red represent weights indicative of enhanced Green Fluorescent Protein (eGFP) and chlorophyll, respectively. Most shoot in this image is transgenic, although an 'escape' leaf with no eGFP is visible (orange arrow). (c) Segmentation output given RGB image data (segments: shoot (green), callus (blue), unregenerated stem (red), contamination (yellow)). (d) Cross-referencing of results from RGB-based segmentation and hyperspectral analysis to measure transgenic shoot. In phenomics outputs (b–d), shoot tissue is circled and highlighted with a manually added circle to aid visualization. A grid was also manually added to all these figures to aid visualization.

Table 1 Statistics collected from phenomics workflow and used for downstream analysis.

(a) Trait summary	(b) Trait computation	(c) Tissue (*X*)	(d) Timepoints
Frequency of *X* transgenic tissue	*n* explants with transgenic *X* tissue/*n* explants	Stem and callus	Week 3
		Callus	Week 3
		Callus and shoot	Week 7
		Shoot	Week 7, Week 10
Frequency of *X* tissue	*n* explants with *X* tissue/*n* explants	Callus	Week 3
		Callus and shoot	Week 7
Size of *X* transgenic tissue	*n* pixels of *X* tissue meeting eGFP significance threshold	Stem and callus	Week 3
		Callus	Week 3
		Callus and shoot	Week 7
		Shoot	Week 7, Week 10
Size of *X* tissue	*n* pixels of *X* tissue	Callus	Week 3
		Callus and shoot	Week 7

A given trait vector used for downstream analysis represents a specific trait (a, b) for a specific tissue or tissue group (c) at one of three timepoints (d), either with or without the acetosyringone (AS) treatment. For example, the first row corresponds to two trait vectors: (1) frequency of transgenic stem/callus at week 3 with AS; or (2) without AS in co-cultivation media. eGFP, enhanced green fluorescent protein.

deployed using high-performance cluster resources provided by the NSF XSEDE and ACCESS programs. To identify and rank significance of quantitative trait loci (QTLs), we applied multiple-testing corrections using the Bonferroni and Benjamini–Hochberg false discovery rate (FDR) methods, as well as applying the Augmented Rank Truncation (Vsevolozhskaya *et al.*, 2019) to identify associations represented by relatively weak signals (Methods S6). To obtain annotation data for candidate genes, we sourced gene annotations from PHYTOZOME (Tuskan *et al.*, 2006), Arabidopsis homolog information from The Arabidopsis Information Resource (TAIR), and predicted lincRNA data from the GreeNC database (Di Marsico *et al.*, 2022).

We performed gene ontology (GO) and pathway tests for groups of candidate genes categorized by their association with early or later TR stages, and used the PANTHER tool for overrepresentation tests (Mi *et al.*, 2019; Thomas *et al.*, 2022). Additionally, KEGG metabolic pathway enrichment tests were conducted using the CLUSTERPROFILER package in R (Methods

Table 2 Published datasets that were integrated for network analysis.

Type of interaction	Description	Reference
Transcriptional regulation (direct)	DNA sites bound by WUS (ChIP-seq)	Ma *et al*. (2019)
	DNA sites bound by GRF1 (ChIP-seq)	Piya *et al*. (2020)
	Y1H with selected meristematic regulators	Ikeuchi *et al*. (2018)
Transcriptional regulation (direct or indirect)	DEGs in *WUS* overexpression lines (RNA-Seq)	Ma *et al*. (2019)
	DEGs in *GRF1* overexpression lines (RNA-Seq)	Piya *et al*. (2020)
	DEGs in *pi-4kβ1, 2* loss-of-function lines (RNA-Seq)	Starodubtseva *et al*. (2022)
	DEGs in *nf-yc 3, 4, 9* loss-of-function lines (RNA-Seq)	X. Liu *et al*. (2016)
	DEGs in *HAN* overexpression lines (microarray)	Zhang *et al*. (2013)
	DEGs in *OBP1* overexpression lines (quantitative reverse transcription polymerase chain reaction)	Skirycz *et al*. (2008)
Protein–protein interaction	Y2H of selected phytohormone regulators	Altmann *et al*. (2020)

Descriptions and sources are provided for each curated dataset related to Arabidopsis homologs of candidate genes and their regulation. These published datasets provide information on differentially expressed genes (DEGs), and interactions supported by yeast one-hybrid (Y1H) and yeast two-hybrid (Y2H) assays. This published data was integrated to support the construction and analysis of a regulatory network comprising homologs of our genes and genes with which they interact.

S7). To construct a network model to inform candidate gene prioritization based on multi-omic lines of evidence, we integrated data from published studies (Table 2) involving Arabidopsis homologs of candidate genes. Using IGRAPH in R, we computed statistics for each node including degree and a Random Walk with Restart (RWR) score (Methods S8). The network was visualized using CYTOSCAPE, focusing on paths connecting candidate gene orthologs.

To support epistasis tests, we first used PLINK to produce a filtered SNP set in which 400 candidate genes were represented by *c.* five of the most significant corresponding SNPs. Epistasis tests were conducted using the epistasis function of the FAST-LMM PYTHON package (Lippert *et al.*, 2011) (Methods S9).

Results

PCA informs contrasts across stages and types of regeneration

For each of five PCA groupings, we obtained two components appearing before 'elbows', all of which accounted for > 10% of variance. For the three groupings of callus regeneration, callus transformation, and shoot transformation, inspection of loadings indicated that the first components of each represent a general positive trend across input traits, while the second components represent contrasts (e.g. between treatments, or frequency and size of tissues; Figs S3, S4, S6). For instance, for the callus regeneration trait PCA, PC1 displayed similar values across the four traits of callus size and callus frequency with and without added acetosyringone (AS) treatment, although callus size traits appear to contribute slightly more than callus frequency traits. PC2 displayed a clear contrast between callus size traits (positive values) and callus frequency traits (negative values), enabling genetic mapping of the contrast between initial callus emergence and subsequent proliferation.

For the two groupings with traits spanning multiple tissue types (transgenic frequency batch and comprehensive batch), PC1 and PC2 represented contrasts across stages of regeneration and/or types of phenotype measures. Although PCs for these groupings are not simple to interpret, clusters of values are clearly visible corresponding to distinct measures and/or stages of regeneration. For the transgenic tissue frequency PCA, both PC1 and PC2 display clusters of values for three groups: (1) frequencies of transgenic shoots; (2) frequencies of transgenic tissues labeled as callus; and (3) frequencies of transgenic tissues labeled as either callus or stem, which we determined to include very early or stunted stages of microcallus development that were not clearly visible and classifiable by our system as callus (Fig. S7). No significant PCs for any of the five PCA groupings appear to represent clear contrasts between media treatments with and without added AS.

Large number of associations supported by combination of GWAS methods

Among traits with a h^2_{SNP} above 0.10, all related to callus alone or combinations/contrasts of callus/shoot were investigated for candidate genes, with the exception of a borderline-heritable trait representing a contrast of callus and shoot regeneration with and without extra AS (Table S2). Traits representing shoot alone could not be mapped successfully due to problems with model-fitting for these highly sparse traits. Where multiple loci within 30 kb passed significance thresholds, the candidate locus was taken as that of greatest significance (henceforth, 'QTL peak'). MTMC-SKAT yielded 97 QTL peaks passing the conservative Bonferroni threshold and 319 passing this and/or the FDR ($\alpha = 0.10$) threshold. Only nine QTL peaks passing the FDR ($\alpha = 0.10$) threshold were identified by GEMMA, and none by GMMAT. The use of ART greatly improved statistical power for GEMMA and GMMAT, enabling the identification of 203 and 22 QTL peaks, respectively (Fig. 3; Table S3). Across methods, traits, and significance thresholds, we report a total of 409 unique poplar candidate genes implicated by a QTL peak within 5 kb, and 271 of these are annotated as being putatively orthologous with a specific Arabidopsis gene (Tables S4, S5).

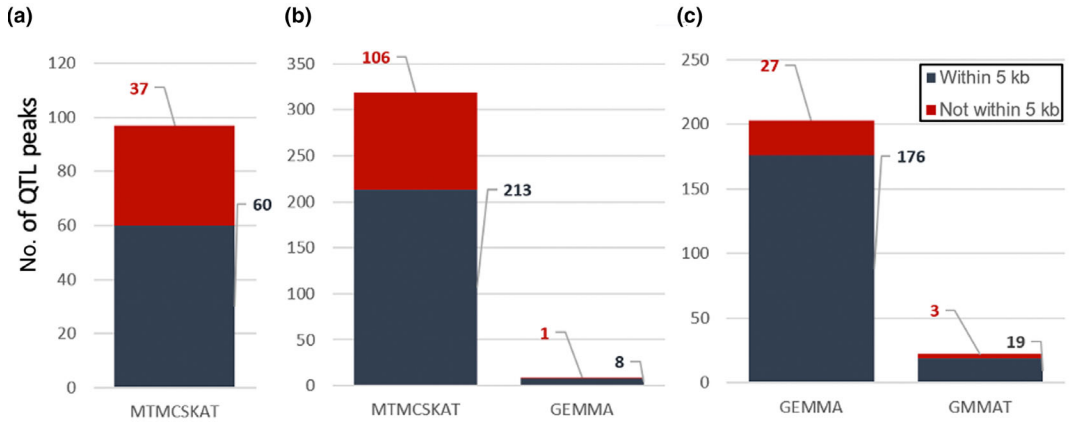

Fig. 3 Tallies of *in vitro* callus and shoot associations across genome-wide association study methods. (a) Conservative Bonferroni correction, given assumption of independence between all single-nucleotide polymorphisms (SNPs); Genome-wide Efficient Mixed Model Association (GEMMA); Generalized Mixed Model Association Test (GMMAT) or SNP windows for Multi-Threaded Monte Carlo SNP-Set Sequence Kernel Association Test (MTMC-SKAT). (b) False discovery rate ($\alpha = 0.10$) correction computed by Benjamini–Hochberg method, also given assumption of independence. (c) Augmented Rank Truncation (ART)-Bonferroni, where the number of tests is considered as the number of possible 1 kb ART windows in the genome. Additional statistics on candidate genes are given in Supporting Information Table S3.

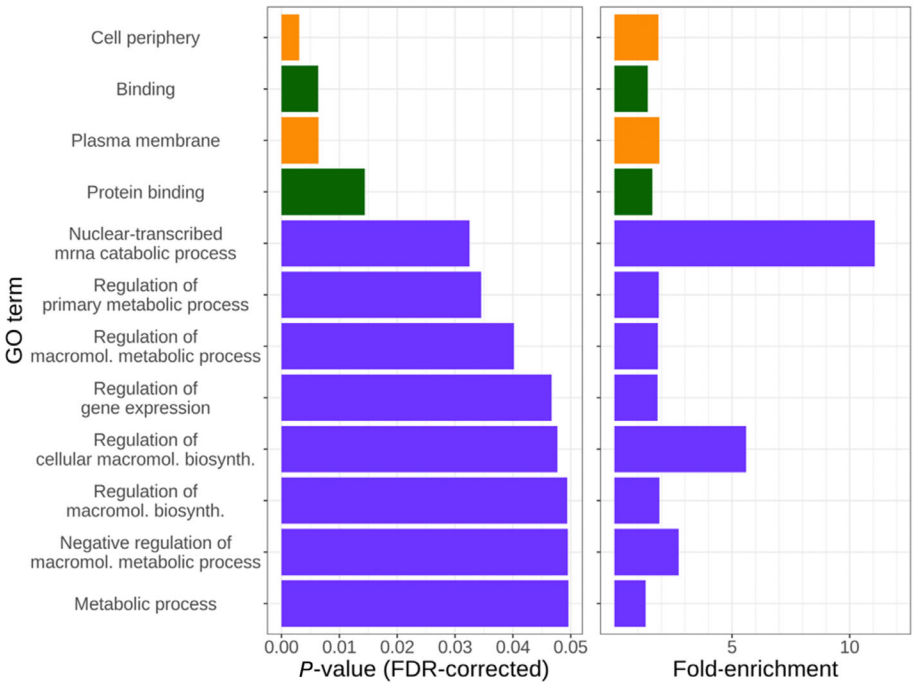

Fig. 4 Fold-enrichment and false discovery rate (FDR)-corrected *P*-values for overrepresented gene ontology (GO) terms found among candidate genes. Gene ontology overrepresentation analysis was performed using PANTHER (https://www.pantherdb.org/), using Arabidopsis genes labeled in the *Populus trichocarpa* genome annotation as being orthologous with candidate genes supported by a quantitative trait locus peak inside the gene or within 5 kb. Color represents GO term aspects: cellular component (orange), molecular function (green), and biological process (purple).

Gene ontology analysis reveals significantly overrepresented categories

The 271 Arabidopsis orthologs of candidate genes within 5 kb of a QTL peak were evaluated with GO and pathway enrichment tests. Of the three groups of candidates tested (week 3 machine vision traits, weeks 7–10 machine vision traits, and contrasts/trends from PCA), only the first group (week 3) yielded significantly overrepresented GO terms (Fig. 4). These indicate prominent roles for genes involved in outer parts of cells ('cell periphery' and 'plasma membrane'), protein–protein interactions ('protein binding'), 'gene expression' and metabolic, catabolic or biosynthetic processes, including terms suggesting involvement in RNA interference and protein degradation.

Network analysis supports the identification of likely regulatory hubs

Network analysis was performed using the 271 Arabidopsis orthologs of poplar candidate genes implicated by a QTL peak within 5 kb, with a network constructed given edges curated from literature (Table 2). Among these genes, 206 had at least one

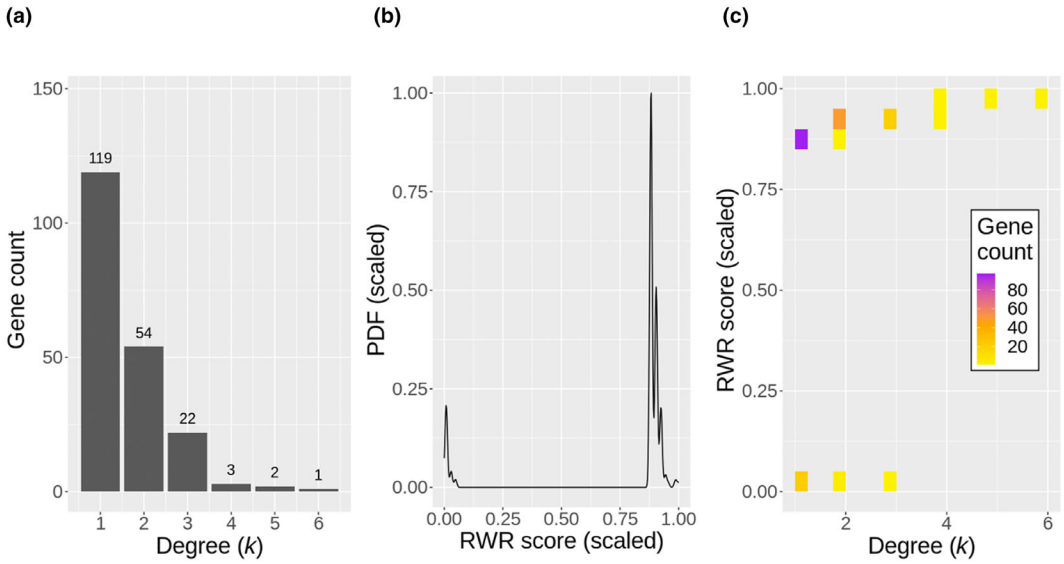

Fig. 5 Summary statistics from network analysis of candidate genes. Edges connecting genes (nodes) based on *in vitro* assays or their differential expression in transgenic lines were curated from literature (Table 2). Statistics are shown for 201 Arabidopsis orthologs of candidate genes that were connected into a contiguous network, except for *NF-YC9*, *HAN*, *OBP1* and *PI-4Kβ1* due to their outlier status as a result of being the focus of curated datasets (Table 1). All network statistics are shown in Supporting Information Tables S4 and S5. Summary statistics are shown for the collapsed version of the network in which parallel edges between any two given genes are reduced to single edges (the Materials and Methods section). (a) Degree (*k*) provides a measure of the number of genes that a given gene is connected to. (b) The random walk with restart (RWR) score provides measure of relevance of candidate genes with respect to WUS (used as seed). (c) Relationship between *k* and RWR score.

edge and 82 had more than one edge (degree > 1; Fig. 5). A contiguous network included 202 of these genes. Among these, 79 were differentially expressed in *pi-4kβ1, 2* mutants, 183 implicated as being regulated by *WUS*, *HAN* and/or *GRF1* via differential expression and/or promoter binding, and 62 appear in the region of overlap as being regulated by *WUS*/*HAN*/*GRF1* and the *P14Kβ* family (Fig. 6; Zhang *et al.*, 2013; Ma *et al.*, 2019; Piya *et al.*, 2020; Starodubtseva *et al.*, 2022). RWR scores were computed with *WUS* as a seed node and displayed two clusters, where genes in the cluster with greater scores have direct connections to *WUS* and others do not (Fig. 5b,c).

Epistasis testing provides evidence for interactions relevant to regeneration in poplar

We used Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) to screen for epistatic interactions among 400 candidate genes with respect to 19 traits (the Materials and Methods section). Results passing Bonferroni and/or FDR ($\alpha = 0.10$) significance thresholds were only found for a single trait, representing total callus size at week 3 under the control treatment (without added AS). We report 45 SNP–SNP interactions passing the Bonferroni threshold, representing 18 pairs of distinct gene–gene interactions involving 20 genes. With the stringency of multiple-testing correction relaxed to FDR ($\alpha = 0.10$), we report 294 SNP–SNP interactions, representing 115 gene–gene interactions involving 81 genes (Figs 7, 8; Tables S6, S7).

We cross-referenced results from epistasis testing together with the interaction network curated from literature review. Among the 81 candidate genes indicated by FaST-LMM to be involved in epistatic interactions, 59 were annotated as having Arabidopsis orthologs, 38 of which were represented in network data curated from literature, and 27 of which were in a contiguous network constructed from the literature-curated data (which featured 202 Arabidopsis orthologs of candidate genes in total). We observed no cases where parallel direct edges were found across the FaST-LMM epistasis and literature-curated network datasets (e.g. Gene X and Gene Y directly linked by our epistasis test in addition to published yeast one-hybrid interaction).

Discussion

Candidate genes imply important roles of hormone signaling and cross talk

We undertook an extensive literature review to investigate candidate genes prioritized based on h^2_{SNP} of traits, proximity of SNPs to genes, overrepresented GO categories, and multiple lines of evidence from network analysis and epistasis testing. We only considered 271 candidate genes which had an Arabidopsis homolog as indicated in the *P. trichocarpa* genome annotation.

Our most notable candidate genes represent regulators of hormone-regulated gene expression, biosynthesis of hormones and other metabolites, and wound response, pH, nutrient and hormone transport, cell structure and more. These associations and the body of literature on their Arabidopsis homologs demonstrate that capacity for *in vitro* regeneration depends on highly complex, nonlinear and interdependent hormonal pathways,

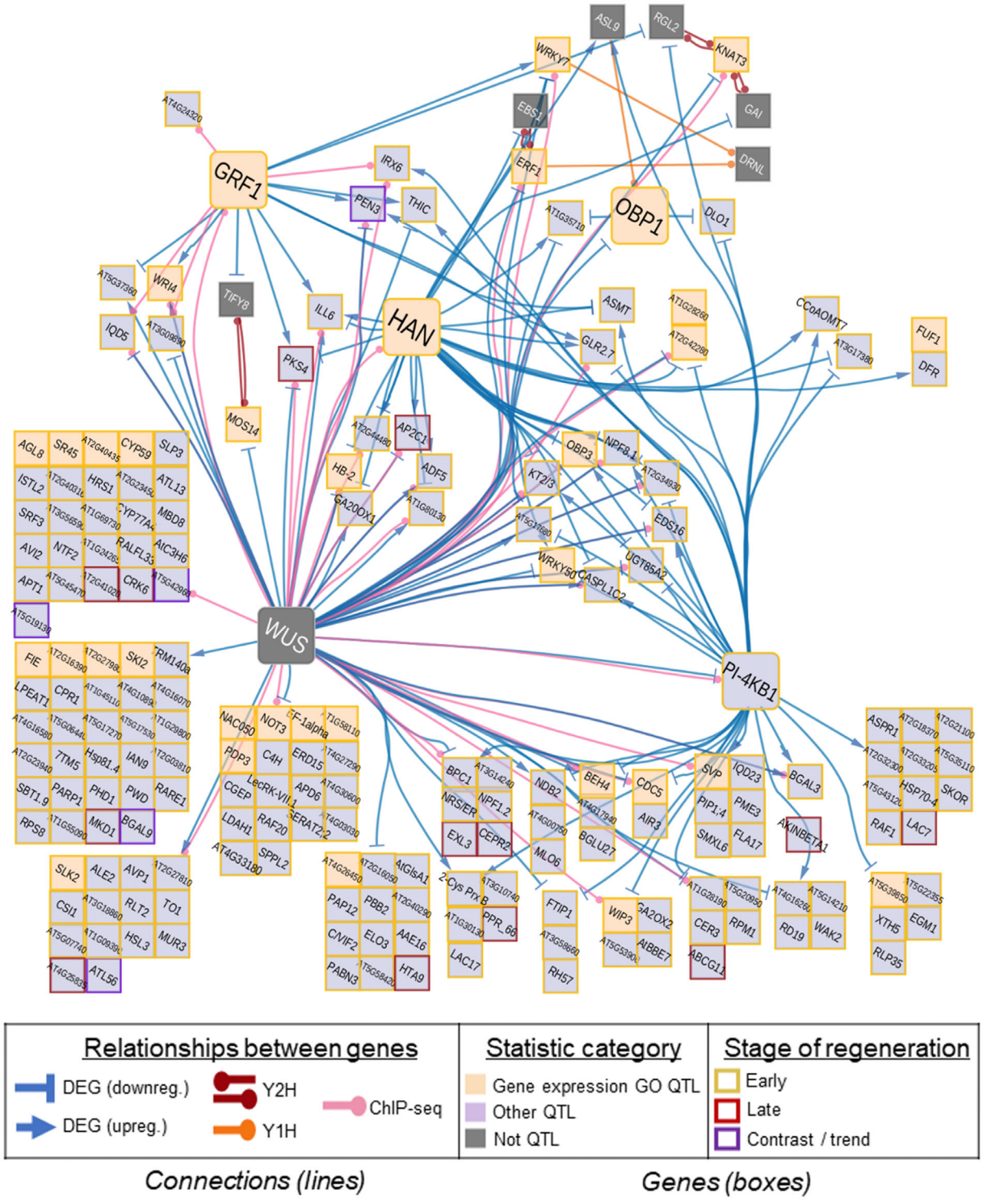

Fig. 6 Sub-network for candidate genes with Arabidopsis homologs. Arabidopsis homologs of candidate genes are presented along with connections from published transcriptomic, ChIP-seq, yeast one-hybrid (Y1H) and yeast two-hybrid (Y2H) datasets (Table 2). For simplicity and readability, this figure is limited to curated datasets closely linked to auxin/cytokinin signaling (those for *WUS*, *GRF1* and *HAN*), or salicylic acid signaling (via the *PI-4Kβ* family). The complete set of edges used in our upstream network construction and analysis are given in Supporting Information Tables S4 and S5. In addition to candidate genes (orange and purple), other genes relevant to pathways connecting candidates are also shown (gray). Genes represented by identical edges are placed adjacently and represented by with a single nonredundant edge or edges for the cluster. For genes represented by at least some degree of functional characterization and which have a common name on the Arabidopsis Information Resource (arabidopsis.org; Berardini *et al*., 2015), the common name acronym is given as a label. For genes for which a common name was not retrieved, typically due to little or no characterization of the gene, the accession ID is given instead. This figure was produced using Cytoscape.

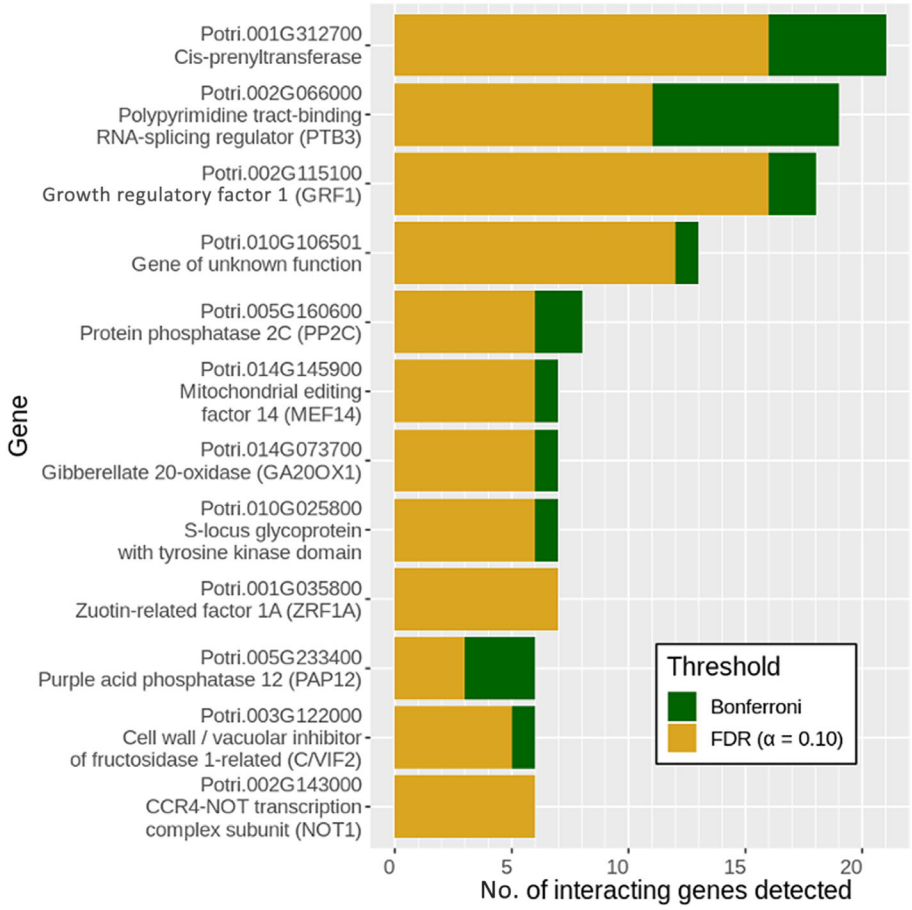

Fig. 7 Tallies of epistatic interactions detected, shown for the 12 candidate genes for which pairwise interactions were detected with six or more candidate genes. Interactions were tallied by the most stringent multiple-testing threshold each passed for a given gene pair, including Bonferroni (more stringent) and the false discovery rate (FDR; less stringent). All interactions passing the Bonferroni threshold also passed the FDR threshold, and are only tallied toward the more stringent Bonferroni threshold in this figure.

along with downstream regulators of cell division, expansion, and structure (Fig. 9). Here, we emphasize the candidate genes we identified that are most likely to act as regulatory hubs of hormone signaling, as transgenic perturbations of these may show the most promise in helping to improve regeneration of recalcitrant plants. Summaries of additional selected candidate genes are provided in Notes S1.

Auxin and cytokinin network

Key regulators of auxin-related epigenetic reprogramming are in the *GRF* family (Szczygieł-Sommer & Gaj, 2019), members of which appear to be involved in recruitment of SWITCH/SUCROSE NONFERMENTING (SWI/SNF) chromatin remodeling complex (Debernardi *et al.*, 2014) and include our candidate Potri.002G115100 (homolog of *GRF1*). This poplar homolog is a likely target of a stem-expressed microRNA transcriptionally regulated downstream of auxin (Yang *et al.*, 2023). In addition to acting both upstream (Szczygieł-Sommer & Gaj, 2019) and downstream (Beltramino *et al.*, 2021; Yang *et al.*, 2023) of auxin, the *GRF* family has been implicated downstream of gibberellin signaling (van der Knaap *et al.*, 2000) and upstream of cytokinin, JA and SA, among other phytohormone pathways (Piya *et al.*, 2020). A key role of *GRF1* as a regeneration regulatory hub is supported by our epistasis results, showing significant interactions with 18 other candidates.

ZUOTIN-RELATED FACTOR 1A (*ZRF1A*; homolog of Potri.001G035800) encodes an epigenetic regulator of auxin/cytokinin network genes including *WUS*, *STM*, and *CLV3*, with a mechanism likely involving ubiquitination of histone 2A proteins (such as one encoded by candidate gene Potri.006G249400; Feng *et al.*, 2016, 2021; Guzmán-López *et al.*, 2016). In Arabidopsis, loss-of-function mutants of *ZRF1* show aberrations in the meristems of root (Chen *et al.*, 2019) and shoot (Feng *et al.*, 2016; Guzmán-López *et al.*, 2016). Our epistasis tests support interactions of *ZRF1* with seven other candidate genes, including the splicing factor *ATO*, the cell expansion regulator *EXORDIUM-LIKE 3*, and the tumor-suppressor gene *INHIBITOR OF GROWTH 2* (*ING2*) (Ricordel *et al.*, 2021), a coordinator of epigenetic cross talk known to span histone 3 lysine 4 methylation and histone acetylation (Shi *et al.*, 2006).

HANABA TANARU (*HAN*, homolog of Potri.005G122700) represents another, likely major nexus between auxin/cytokinin signaling and other phytohormones. A transcriptomic study of wood development in *Populus alba* × *glandulosa* found the poplar homolog to be differentially expressed in cambium of mature stems (Kim *et al.*, 2019). Arabidopsis overexpression lines of *HAN* display differential expression of hundreds of genes representing diverse phytohormone pathways, including that of auxin, cytokinin, ethylene, JA, and SA, among others (Zhang *et al.*, 2013). These include orthologs of many of our candidate genes

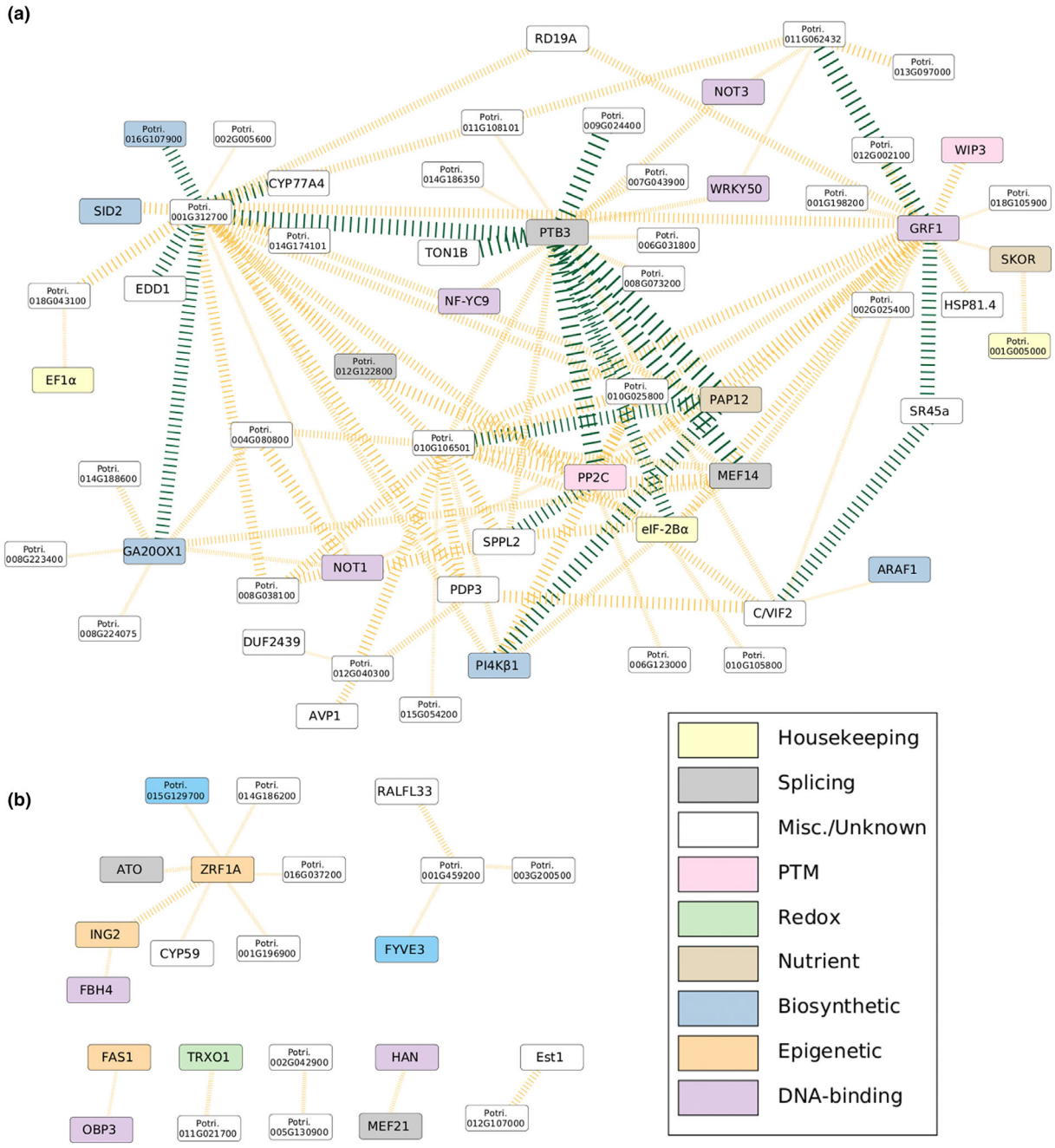

Fig. 8 Epistatic interactions detected by Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) are shown for (a) large contiguous network; and (b) islands disconnected from the larger network. The width of connections represents significance, while green connections represent passing the Bonferroni threshold and yellow represents passing the false discovery rate (FDR) threshold ($\alpha = 0.10$) computed by the Benjamini–Hochberg method. Genes of greatest degree (most connections) include homologs of phytohormone signaling regulators *GRF1*, *GA2OX1* and *ZRF1A*, as well as alternative splicing regulator *PTB3*, each of which were targeted for literature review (the Discussion section), as well as two genes of unknown function. Genes are labeled by common names for the *Populus trichocarpa* (poplar) gene and/or Arabidopsis homolog where such names are found in the poplar genome annotation, while uncharacterized or otherwise unnamed genes without annotated putative homologs are instead represented by accession IDs (Tuskan *et al.*, 2006).

(Fig. 6). These widespread effects on gene expression may result largely from direct transcriptional regulation of *CYTOKININ OXIDASE 3*, as well as through other targets (Ding *et al.*, 2015). *HAN* loss-of-function leads to disrupted apical/basal patterning and meristem formation during embryogenesis, indicated by meristematic development markers including *WOX5* (Nawy *et al.*, 2010), a poplar homolog of which influences adventitious rooting when overexpressed (Li *et al.*, 2018). Our epistasis results support an interaction of poplar *HAN* with Potri.010G161700 (homolog of *MITOCHONDRIAL EDITING FACTOR 21*),

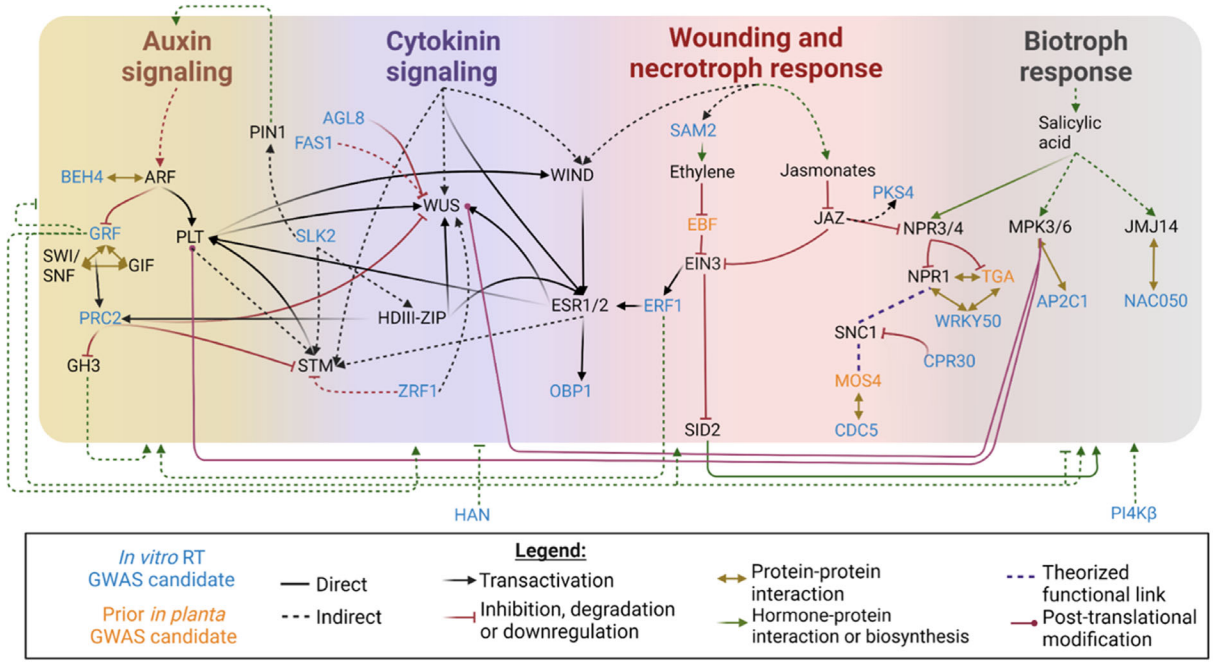

Fig. 9 Proposed model integrating roles of *in vitro* transformation and regeneration candidate genes in phytohormone signaling, crosstalk, and biosynthesis. Evidence for direct and indirect mechanistic relationships between genes was collected from studies in model organisms (Supporting Information Table S8). Interactions are assumed to be indirect except where strong functional evidence suggests a direct interaction. Key candidate genes with roles in gene expression regulation or hormone biosynthesis/metabolism are presented here, from both the current study and our prior genome-wide association study of *in planta* regeneration. This proposed model is simplified by combining connections with respect to seven gene families (*ARF*, *PLT*, *SWI/SNF*, *HDIII-ZIP*, *WIND*, *ESR1/2*, *JAZ*) with partially redundant and/or sequential roles. This model is simplified and nonexhaustive, intended to summarize likely roles of key candidate genes in these pathways rather than comprehensively explaining the pathways. This figure was produced using BioRender.

which may regulate RNA splicing in mitochondria in a manner responsive to abiotic stress and ABA signaling (Takenaka *et al.*, 2010; J-M. Liu *et al.*, 2016).

Ethylene, JA and SA networks

Plant tissue damage induces JA/ethylene, altering gene expression for tissue repair (Ikeuchi *et al.*, 2020), while SA signaling interacts antagonistically with JA/ethylene (Li *et al.*, 2019). Arabidopsis mutants in JA pathways show increased callus proliferation postwounding (Ikeuchi *et al.*, 2017).

In Arabidopsis, SA-dependent transcriptional regulation depends largely on a transcriptional complex consisting of NPR1 and TGA family proteins together with WRKY50 (homolog of Potri.006G224100). Across our GWAS reported here, and our prior GWAS of *in planta* regeneration (M. F. Nagle *et al.*, 2024), we report several likely regulators of *NPR1*, including *CPR30* (homolog of Potri.001G035500; Gou *et al.*, 2012), *MOS4* (implicated in our prior GWAS), and the MOS4-interacting *CELL DIVISION CYCLE 5* (*CDC5*; homolog of Potri.013G046300) (Monaghan *et al.*, 2009). In addition to *WRKY50*, a role for *WRKY7* (homolog of Potri.005G141400) in SA/JA/ethylene signaling is also supported by effects of *WRKY7* mutations on systemic acquired resistance (Arraño-Salinas *et al.*, 2018).

PI-4Kβ1 (homolog of Potri.007G109000) is believed to affect SA (Janda *et al.*, 2014; Šašek *et al.*, 2014) via biosynthesis of phosphoinositides, which function as cofactors for enzymes involved in membranes and vesicles responsible in part for signaling molecule transport (Mayinger, 2012; Starodubtseva *et al.*, 2022). In addition to regulating levels of SA itself, phosphoinositides may regulate the action of downstream SA-signaling genes, which are highly represented among targets of the phosphoinositide-dependent phospholipase C (Kalachova *et al.*, 2015). Double loss-of-function mutants of *PI-4Kβ1* and the related *PI-4Kβ2* (*pi-4kβ1, 2*) show phenotypes including overaccumulation of SA (Janda *et al.*, 2014; Šašek *et al.*, 2014) and insensitivity to exogenous auxin (Starodubtseva *et al.*, 2022). Our epistasis results support four interactions for *PI-4Kβ1*, including with two uncharacterized genes, an elongation factor and the phosphate scavenger *PURPLE ACID PHOSPATASE 12* (Hurley *et al.*, 2010; Robinson *et al.*, 2012) (Fig. 8; Notes S1), and our network integration supports the co-regulation of diverse genes by *PI-4Kβ1* and regulators of the auxin-cytokinin axis (Fig. 6).

Abscisic acid and gibberellin signaling

Other candidate genes suggest a role for ABA and gibberellins. *NF-YC9* (homolog of Potri.005G035800) encodes a member of a transcriptional complex responsible for ABA and gibberellin signaling, with loss-of-function and overexpression lines displaying altered sensitivity to one or both hormones (X. Liu *et al.*, 2016; Bi *et al.*, 2017). In addition, one other candidate

gene (putative RNA-splicing regulator *PTB3*; Potri.002G066000) was indicated by our epistasis tests to interact with poplar *NF-YC9*. We identified two candidate genes involved in metabolism of gibberellins. *GIBBERELLIN 2-OXIDASE 2* (homolog of Potri.004G065000) and *GIBBERELLIN 20-OXIDASE 1* (homolog of Potri.014G073700) encode multiple steps of a gibberellin metabolic pathway (Fig. S8; Kanehisa *et al.*, 2023). Roles for these genes in TR may involve the regulation of *GRF* family genes by gibberellin signaling (van der Knaap *et al.*, 2000), and/or other functions of gibberellins in regulating plant growth and development.

Cell periphery regulators or structural proteins

In addition to transcriptional and metabolic regulators, components and regulators of the cell periphery were shown by GO analysis to be significantly overrepresented among our candidates. Cell expansion is also believed to be influenced by *EXORDIUM* (*EXO*) and related genes (including *EXORDIUM-LIKE 3*, a homolog of Potri.015G129700), which appear to bridge brassinosteroid signaling with downstream expression of expansins and other growth regulators (Schröder *et al.*, 2009). Overexpression of *EXO* was reported to enhance *in vivo* root and shoot growth (Coll-Garcia *et al.*, 2004). Additionally, cell expansion can be affected by stiffening of cell walls with lignin; Potri.016G107900 is a homolog of *LACCASE 7* (*LAC7*), known along with other *LACCASE* genes to catalyze the biosynthesis of lignin in diverse plants including poplar (Sterjiades *et al.*, 1992; Zhao *et al.*, 2015; Qin *et al.*, 2020; Yu *et al.*, 2020) while producing ROS byproducts (relevance discussed below; Perna *et al.*, 2020).

Prioritization and functional validation of candidate genes and combinations thereof

Functional characterization of morphogenic regulators that can enhance regeneration when overexpressed has shown that these genes tend to function as regulatory hubs that control cascades of transcriptional regulation closely associated with auxin and cytokinin (Ikeuchi *et al.*, 2018; Nagle *et al.*, 2018; Gordon-Kamm *et al.*, 2019; Lee & Wang, 2023), for example in the cases of *BBM* and the downstream *LEAFY COTYLEDON* family (Horstman *et al.*, 2017), *WUS* (Ma *et al.*, 2019; Jha *et al.*, 2020), the *WIND* family and downstream *ESR* family (Iwase *et al.*, 2017, 2021), the *PLT* family (Kareem *et al.*, 2015), and the *GRF* family (Beltramino *et al.*, 2021). Together with our GO evidence for a significant role of transcription factors, we propose that transgenic perturbation of other candidate genes that function as auxin/cytokinin-signaling transcription factors, including *OBP1*, *SLK2*, *AGL8*, and *HAN*, may enhance regeneration.

The diverse roles of Arabidopsis homologs of our candidate genes across phytohormone pathways suggests that transgenic perturbation of ethylene, JA and SA signaling via candidates such as *WRKY7* and *WRKY50*, or other transcription factors, may help to improve TR. Stress and wounding response downstream of these hormones may act as 'bottlenecks' of regeneration via cross talk with auxin/cytokinin signaling, and/or through roles in necrosis, ROS homeostasis, growth and metabolism.

Reliable validation of highly polygenic traits, especially those marked by complex gene interactions and cell- or temporal-specific expression, presents a challenge for the commonly used functional validation approaches that employ simple gain- or loss-of-function methods. As an alternative and supplement to mutant studies, multi-omic integration and network analysis have become key in GWAS candidate gene validation, with a recent meta-analysis underscoring the roles of these *in silico* approaches to provide added support for noncoding variants in human GWAS (Alsheikh *et al.*, 2022). In addition, the availability of many published high-throughput datasets from Arabidopsis enabled us to construct a network model that informs possible functional relationships for over 200 of our candidate genes.

However, as published datasets and genome annotations are limited to known and characterized genes, there is likely to be confirmation bias with respect to known pathways. Among candidate genes putatively implicated by 409 QTLs in this work, 138 lacked known Arabidopsis homologs in the genome annotation. This risk of bias can be mitigated in future work via comprehensive and untargeted genome-wide gene/protein interaction datasets, along with the continued characterization and annotation of genes in reference genomes.

The key contributions from our epistasis and network analyses are the insights they provide into prioritizing candidate genes for further research, particularly with regard to design of experimental strategies to confirm or characterize their functions. For instance, *in planta* epistasis mutant studies and *in vitro* interaction assays can take advantage of putative gene–gene interactions to clarify conservation of gene function across poplar and model species homologs. In addition to functional validation, analysis of gene interactions in diverse genotypes should offer clues regarding the high genetic diversity in amenability to TR, and thus may help to develop 'gene cocktails' that will enhance TR in diverse genotypes.

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support (IOS no. 1546900, Analysis of genes affecting plant regeneration and transformation in poplar), and members of GREAT TREES Research Cooperative at OSU for its support of the Strauss laboratory.

Support for the Poplar GWAS dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER) via the Center for Bioenergy Innovation (CBI) under Contract No. DE-PS02-06ER64304. The Poplar GWAS Project used resources of the Oak Ridge Leadership Computing Facility and the Compute and Data Environment for Science at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

This work used EXPANSE at the San Diego Supercomputer Center at the University of California, San Diego. An EXPANSE allocation was provided first via allocation MCB200174 from the Extreme Science and Engineering Discovery Environment

(XSEDE), which was supported by National Science Foundation grant no. 1548562. This allocation was then transitioned to the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grant nos. 2138259, 2138286, 2138307, 2137603, and 2138296.

Plasmid material and thymidine-auxotroph *Agrobacterium* was provided courtesy of Corteva Agrisciences, from whom these materials are available to researchers upon request.

Competing interests

None declared.

Author contributions

SHS, LF, WM and YJ designed and directed the overall study, and obtained funding for its execution; WM provided updated SNP data and advised on its analysis; CM and EP designed and/or executed the phenotypic analyses; GSG cloned the transformation reporter construct and designed protocols for its use; SHS and AM supported the custom specification of imaging hardware produced for this project; MFN, JY, DK, YJ, JYL and LF created, adapted and executed the machine vision, computation and data analysis pipelines; MFN investigated candidate genes; MFN wrote the manuscript with editing from SHS, and all others contributed further edits and revisions.

ORCID

Li Fuxin https://orcid.org/0000-0001-7578-9622
Greg S. Goralogia https://orcid.org/0000-0003-4222-3264
Yuan Jiang https://orcid.org/0000-0001-6409-9159
Damanpreet Kaur https://orcid.org/0000-0001-8220-9083
Jia Yi Li https://orcid.org/0009-0001-3423-0619
Cathleen Ma https://orcid.org/0000-0002-5387-4719
Anna Magnuson https://orcid.org/0009-0001-0042-6939
Wellington Muchero https://orcid.org/0000-0002-0200-9856
Michael F. Nagle https://orcid.org/0000-0002-0297-8999
Ekaterina Peremyslova https://orcid.org/0000-0002-2851-0227
Steven H. Strauss https://orcid.org/0000-0001-9670-3082
Jialin Yuan https://orcid.org/0000-0003-1390-2233

Data availability

RGB and hyperspectral image data is available publicly (M. Nagle *et al.*, 2024) and linked at the main GitHub for the GMO-DETECTOR phenomics software (https://github.com/naglemi/GMOnotebook).

References

Alsheikh AJ, Wollenhaupt S, King EA, Reeb J, Ghosh S, Stolzenburg LR, Tamim S, Lazar J, Davis JW, Jacob HJ. 2022. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Medical Genomics* **15**: 74.

Altmann M, Altmann S, Rodriguez PA, Weller B, Elorduy Vergara L, Palme J, Marín-de la Rosa N, Sauer M, Wenig M, Villaécija-Aguilar JA *et al.* 2020. Extensive signal integration by the phytohormone protein network. *Nature* **583**: 271–276.

Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson DP, Kausch AP *et al.* 2016. Advancing crop transformation in the era of genome editing. *Plant Cell* **28**: 1510–1520.

Aregawi K, Shen J, Pierroz G, Bucheli C, Sharma M, Dahlberg J, Owiti J, Lemaux PG. 2020. Pathway to validate gene function in key bioenergy crop, *Sorghum bicolor*. *bioRxiv*. doi: 10.1101/2020.12.08.416347.

Argueso CT, Kieber JJ. 2024. Cytokinin: from autoclaved DNA to two-component signaling. *Plant Cell* koad327.

Arraño-Salinas P, Domínguez-Figueroa J, Herrera-Vásquez A, Zavala D, Medina J, Vicente-Carbajosa J, Meneses C, Canessa P, Moreno AA, Blanco-Herrera F. 2018. WRKY7, -11 and -17 transcription factors are modulators of the bZIP28 branch of the unfolded protein response during PAMP-triggered immunity in *Arabidopsis thaliana*. *Plant Science* **277**: 242–250.

Arroyo-Herrera A, Gonzalez AK, Moo RC, Quiroz-Figueroa FR, Loyola-Vargas VM, Rodriguez-Zapata LC, D'Hondt CB, Suárez-Solís VM, Castaño E. 2008. Expression of WUSCHEL in *Coffea canephora* causes ectopic morphogenesis and increases somatic embryogenesis. *Plant Cell, Tissue and Organ Culture* **94**: 171–180.

Bdeir R, Muchero W, Yordanov Y, Tuskan GA, Busov V, Gailing O. 2019. Genome-wide association studies of bark texture in *Populus trichocarpa*. *Tree Genetics & Genomes* **15**: 14.

Beltramino M, Debernardi JM, Ferela A, Palatnik JF. 2021. ARF2 represses expression of plant GRF transcription factors in a complementary mechanism to microRNA miR396. *Plant Physiology* **185**: 1798–1812.

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. 2015. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**: 474–485.

Bi C, Ma Y, Wang X-F, Zhang D-P. 2017. Overexpression of the transcription factor NF-YC9 confers abscisic acid hypersensitivity in Arabidopsis. *Plant Molecular Biology* **95**: 425–439.

Bouchabké-Coussa O, Obellianne M, Linderme D, Montes E, Maia-Grondard A, Vilaine F, Pannetier C. 2013. Wuschel overexpression promotes somatic embryogenesis and induces organogenesis in cotton (*Gossypium hirsutum* L.) tissues cultured *in vitro*. *Plant Cell Reports* **32**: 675–686.

Bouchez O, Huard C, Lorrain S, Roby D, Balagué C. 2007. Ethylene is one of the key elements for cell death and defense response control in the Arabidopsis lesion mimic mutant vad1. *Plant Physiology* **145**: 465–477.

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu C-M, van Lammeren AAM, Miki BLA *et al.* 2002. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* **14**: 1737–1749.

Che P, Wu E, Simon MK, Anand A, Lowe K, Gao H, Sigmund AL, Yang M, Albertsen MC, Gordon-Kamm W *et al.* 2022. Wuschel2 enables highly efficient CRISPR/Cas-targeted genome editing during rapid *de novo* shoot regeneration in sorghum. *Communications Biology* **5**: 344.

Chen D, Wang Q, Feng J, Ruan Y, Shen W-H. 2019. Arabidopsis ZUOTIN RELATED FACTOR1 proteins are required for proper embryonic and post-embryonic root development. *Frontiers in Plant Science* **10**: 1498.

Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, Szpiro AA, Chen W, Brehm JM, Celedón JC *et al.* 2016. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics* **98**: 653–666.

Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2018. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40**: 834–848.

Chhetri HB, Furches A, Macaya-Sanz D, Walker AR, Kainer D, Jones P, Harman-Ware AE, Tschaplinski TJ, Jacobson D, Tuskan GA *et al.* 2020. Genome-wide association study of wood anatomical and morphological traits in *Populus trichocarpa*. *Frontiers in Plant Science* **11**: 545748.

Chhetri HB, Macaya-Sanz D, Kainer D, Biswal AK, Evans LM, Chen J-G, Collins C, Hunt K, Mohanty SS, Rosenstiel T *et al.* 2019. Multitrait genome-wide association analysis of *Populus trichocarpa* identifies key polymorphisms controlling morphological and physiological traits. *New Phytologist* **223**: 293–309.

Coll-Garcia D, Mazuch J, Altmann T, Müssig C. 2004. EXORDIUM regulates brassinosteroid-responsive genes. *FEBS Letters* **563**: 82–86.

Cortes LT, Zhang Z, Yu J. 2021. Status and prospects of genome-wide association studies in plants. *The Plant Genome* **14**: e20077.

Dai L, Han S, Zhang Y, Hao D. 2022. Genetic architecture of embryogenic callus induction in maize from the perspective of population genomics. *Plant Cell, Tissue and Organ Culture* **150**: 345–359.

Davière J-M, Achard P. 2016. A pivotal role of DELLAs in regulating multiple hormone signals. *Molecular Plant* **9**: 10–20.

Debernardi JM, Mecchia MA, Vercruyssen L, Smaczniak C, Kaufmann K, Inze D, Rodriguez RE, Palatnik JF. 2014. Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. *The Plant Journal* **79**: 413–426.

Debernardi JM, Tricoli DM, Ercoli MF, Hayta S, Ronald P, Palatnik JF, Dubcovsky J. 2020. A GRF–GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nature Biotechnology* **38**: 1274–1279.

Deng W, Luo K, Li Z, Yang Y. 2009. A novel method for induction of plant regeneration via somatic embryogenesis. *Plant Science* **177**: 43–48.

Di Marsico M, Paytuvi Gallart A, Sanseverino W, Aiese Cigliano R. 2022. GreeNC 2.0: a comprehensive database of plant long non-coding RNAs. *Nucleic Acids Research* **50**: D1442–D1447.

Ding L, Yan S, Jiang L, Zhao W, Ning K, Zhao J, Liu X, Zhang J, Wang Q, Zhang X. 2015. HANABA TARANU (HAN) bridges meristem and organ primordia boundaries through PINHEAD, JAGGED, BLADE-ON-PETIOLE2 and CYTOKININ OXIDASE 3 during flower development in Arabidopsis. *PLoS Genetics* **11**: e1005479.

Feng J, Chen D, Berr A, Shen W-H. 2016. ZRF1 chromatin regulators have polycomb silencing and independent roles in development. *Plant Physiology* **172**: 1746–1759.

Feng J, Gao Y, Wang K, Jiang M. 2021. A novel epigenetic regulator ZRF1: insight into its functions in plants. *Genes* **12**: 1245.

Furches A, Kainer D, Weighill D, Large A, Jones P, Walker AM, Romero J, Gazolla JGFM, Joubert W, Shah M *et al.* 2019. Finding new cell wall regulatory genes in *Populus trichocarpa* using multiple lines of evidence. *Frontiers in Plant Science* **10**: 1249.

Gallois J-L, Woodward C, Reddy GV, Sablowski R. 2002. Combined SHOOT MERISTEMLESS and WUSCHEL trigger ectopic organogenesis in Arabidopsis. *Development* **129**: 3207–3217.

Gordon-Kamm B, Sardesai N, Arling M, Lowe K, Hoerster G, Betts S, Jones T. 2019. Using morphogenic genes to improve recovery and regeneration of transgenic plants. *Plants* **8**: 38.

Gou M, Shi Z, Zhu Y, Bao Z, Wang G, Hua J. 2012. The F-box protein CPR1/CPR30 negatively regulates R protein SNC1 accumulation. *The Plant Journal* **69**: 411–420.

Guzmán-López JA, Abraham-Juárez MJ, Lozano-Sotomayor P, de Folter S, Simpson J. 2016. *Arabidopsis thaliana* gonidialess A/Zuotin related factors (GlsA/ZRF) are essential for maintenance of meristem integrity. *Plant Molecular Biology* **91**: 37–51.

Hager A. 2003. Role of the plasma membrane H^+-ATPase in auxin-induced elongation growth: historical and new aspects. *Journal of Plant Research* **116**: 483–505.

Han KH, Gordon MP, Strauss SH. 1997. High-frequency transformation of cottonwoods (genus *Populus*) by *Agrobacterium rhizogenes*. *Canadian Journal of Forest Research* **27**: 464–470.

Hassani SB, Trontin J-F, Raschke J, Zoglauer K, Rupps A. 2022. Constitutive overexpression of a conifer WOX2 homolog affects somatic embryo development in *Pinus pinaster* and promotes somatic embryogenesis and organogenesis in Arabidopsis seedlings. *Frontiers in Plant Science* **13**: 838421.

Heidmann I, de Lange B, Lambalk J, Angenent GC, Boutilier K. 2011. Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. *Plant Cell Reports* **30**: 1107–1115.

Hoerster G, Wang N, Ryan L, Wu E, Anand A, McBride K, Lowe K, Jones T, Gordon-Kamm B. 2020. Use of non-integrating Zm-Wus2 vectors to enhance maize transformation. *In Vitro Cellular & Developmental Biology – Plant* **56**: 265–279.

Horstman A, Li M, Heidmann I, Weemen M, Chen B, Muiño JM, Angenent GC, Boutilier K. 2017. The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. *Plant Physiology* **175**: 848–857.

Hu W, Li W, Xie S, Fagundez S, McAvoy R, Deng Z, Li Y. 2016. Kn1 gene overexpression drastically improves genetic transformation efficiencies of citrus cultivars. *Plant Cell, Tissue and Organ Culture* **125**: 81–91.

Huang G, Liu Z, van der Maaten L, Weinberger KQ. 2018. Densely connected convolutional networks. *arXiv*: 1608.06993.

Hurley BA, Tran HT, Marty NJ, Park J, Snedden WA, Mullen RT, Plaxton WC. 2010. The dual-targeted purple acid phosphatase isozyme AtPAP26 is essential for efficient acclimation of Arabidopsis to nutritional phosphate deprivation. *Plant Physiology* **153**: 1112–1122.

Ikeuchi M, Iwase A, Rymen B, Lambolez A, Kojima M, Takebayashi Y, Heyman J, Watanabe S, Seo M, De Veylder L *et al.* 2017. Wounding triggers callus formation via dynamic hormonal and transcriptional changes. *Plant Physiology* **175**: 1158–1174.

Ikeuchi M, Rymen B, Sugimoto K. 2020. How do plants transduce wound signals to induce tissue repair and organ regeneration? *Current Opinion in Plant Biology* **57**: 72–77.

Ikeuchi M, Shibata M, Rymen B, Iwase A, Bågman A-M, Watt L, Coleman D, Favero DS, Takahashi T, Ahnert SE *et al.* 2018. A gene regulatory network for cellular reprogramming in plant regeneration. *Plant and Cell Physiology* **59**: 770–782.

Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. 2013. Sequence kernel association tests for the combined effect of rare and common variants. *American Journal of Human Genetics* **92**: 841–853.

Iwase A, Harashima H, Ikeuchi M, Rymen B, Ohnuma M, Komaki S, Morohashi K, Kurata T, Nakata M, Ohme-Takagi M *et al.* 2017. WIND1 promotes shoot regeneration through transcriptional activation of *ENHANCER OF SHOOT REGENERATION1* in Arabidopsis. *Plant Cell* **29**: 54–69.

Iwase A, Kondo Y, Laohavisit A, Takebayashi A, Ikeuchi M, Matsuoka K, Asahina M, Mitsuda N, Shirasu K, Fukuda H *et al.* 2021. WIND transcription factors orchestrate wound-induced callus formation, vascular reconnection and defense response in Arabidopsis. *New Phytologist* **232**: 734–752.

Janda M, Šašek V, Ruelland E. 2014. The Arabidopsis pi4kIIIβ1β2 double mutant is salicylic acid-overaccumulating: a new example of salicylic acid influence on plant stature. *Plant Signaling & Behavior* **9**: e977210.

Jha P, Ochatt SJ, Kumar V. 2020. WUSCHEL: a master regulator in plant growth signaling. *Plant Cell Reports* **39**: 431–444.

Kadri A, Grenier De March G, Guerineau F, Cosson V, Ratet P. 2021. WUSCHEL overexpression promotes callogenesis and somatic embryogenesis in *Medicago truncatula* Gaertn. *Plants* **10**: 715.

Kalachova T, Kravets V, Zachowski A, Ruelland E. 2015. Importance of phosphoinositide-dependent signaling pathways in the control of gene expression in resting cells and in response to phytohormones. *Plant Signaling & Behavior* **10**: e1019983.

Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* **51**: D587–D592.

Kang B, Osburn L, Kopsell D, Tuskan GA, Cheng Z-M. 2009. Micropropagation of *Populus trichocarpa* 'Nisqually-1': the genotype deriving the *Populus* reference genome. *Plant Cell, Tissue and Organ Culture* **99**: 251–257.

Kareem A, Durgaprasad K, Sugimoto K, Du Y, Pulianmackal AJ, Trivedi ZB, Abhayadev PV, Pinon V, Meyerowitz EM, Scheres B *et al.* 2015. PLETHORA genes control regeneration by a two-step mechanism. *Current Biology* **25**: 1017–1030.

Kim M-H, Cho J-S, Jeon H-W, Sangsawang K, Shim D, Choi Y-I, Park E-J, Lee H, Ko J-H. 2019. Wood transcriptome profiling identifies critical pathway genes of secondary wall biosynthesis and novel regulators for vascular cambium development in *Populus*. *Genes* **10**: 690.

Klimaszewska K, Pelletier G, Overton C, Stewart D, Rutledge RG. 2010. Hormonally regulated overexpression of Arabidopsis WUS and conifer LEC1 (CHAP3A) in transgenic white spruce: implications for somatic embryo development and somatic seedling growth. *Plant Cell Reports* **29**: 723–734.

van der Knaap E, Kim JH, Kende H. 2000. A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. *Plant Physiology* **122**: 695–704.

Kotov AA, Kotova LM. 2023. Auxin/cytokinin antagonism in shoot development: from moss to seed plants. *Journal of Experimental Botany* **74**: 6391–6395.

Lantzouni O, Alkofer A, Falter-Braun P, Schwechheimer C. 2020. GROWTH-REGULATING FACTORS interact with DELLAs and regulate growth in cold stress. *Plant Cell* **32**: 1018–1034.

Lardon R, Geelen D. 2020. Natural variation in plant pluripotency and regeneration. *Plants* **9**: 1261.

Lee K, Wang K. 2023. Strategies for genotype-flexible plant transformation. *Current Opinion in Biotechnology* **79**: 102848.

Li J, Zhang J, Jia H, Liu B, Sun P, Hu J, Wang L, Lu M. 2018. The *WUSCHEL-related homeobox 5a* (*PtoWOX5a*) is involved in adventitious root development in poplar. *Tree Physiology* **38**: 139–153.

Li N, Han X, Feng D, Yuan D, Huang L-J. 2019. Signaling crosstalk between salicylic acid and ethylene/jasmonate in plant defense: do we understand what they are whispering? *International Journal of Molecular Sciences* **20**: 671.

Li Q, Yeh T-F, Yang C, Song J, Chen Z-Z, Sederoff RR, Chiang VL. 2015. Populus trichocarpa. In: Wang K, ed. *Methods in molecular biology.* Agrobacterium *protocols, vol. 2*. New York, NY, USA: Springer, 357–363.

Lian Z, Nguyen CD, Liu L, Wang G, Chen J, Wang S, Yi G, Wilson S, Ozias-Akins P, Gong H *et al.* 2022. Application of developmental regulators to improve *in planta* or *in vitro* transformation in plants. *Plant Biotechnology Journal* **20**: 1622–1635.

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. 2011. FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**: 833–835.

Liu B, Zhang J, Yang Z, Matsui A, Seki M, Li S, Yan X, Kohnen MV, Gu L, Prasad K *et al.* 2018. *PtWOX11* acts as master regulator conducting the expression of key transcription factors to induce *de novo* shoot organogenesis in poplar. *Plant Molecular Biology* **98**: 389–406.

Liu J-M, Zhao J-Y, Lu P-P, Chen M, Guo C-H, Xu Z-S, Ma Y-Z. 2016. The E-subgroup pentatricopeptide repeat protein family in *Arabidopsis thaliana* and confirmation of the responsiveness PPR96 to abiotic stresses. *Frontiers in Plant Science* **7**: 1825.

Liu X, Bie XM, Lin X, Li M, Wang H, Zhang X, Yang Y, Zhang C, Zhang XS, Xiao J. 2023. Uncovering the transcriptional regulatory network involved in boosting wheat regeneration and transformation. *Nature Plants* **9**: 908–925.

Liu X, Hu P, Huang M, Tang Y, Li Y, Li L, Hou X. 2016. The NF-YC–RGL2 module integrates GA and ABA signalling to regulate seed germination in Arabidopsis. *Nature Communications* **7**: 12768.

Lou H, Huang Y, Wang W, Cai Z, Cai H, Liu Z, Sun L, Xu Q. 2022. Overexpression of the AtWUSCHEL gene promotes somatic embryogenesis and lateral branch formation in birch (*Betula platyphylla* Suk.). *Plant Cell, Tissue and Organ Culture* **150**: 371–383.

Lowe K, Wu E, Wang N, Hoerster G, Hastings C, Cho M-J, Scelonge C, Lenderts B, Chamberlin M, Cushatt J *et al.* 2016. Morphogenic regulators baby boom and Wuschel improve monocot transformation. *Plant Cell* **28**: 1998–2015.

Luo G, Palmgren M. 2021. GRF-GIF chimeras boost plant regeneration. *Trends in Plant Science* **26**: 201–204.

Lutz KA, Martin C, Khairzada S, Maliga P. 2015. Steroid-inducible BABY BOOM system for development of fertile *Arabidopsis thaliana* plants after prolonged tissue culture. *Plant Cell Reports* **34**: 1849–1856.

Ma C, Duan C, Jiang Y, Nagle M, Peremyslova E, Goddard A, Strauss SH. 2022a. Factors affecting *in vitro* regeneration in the model tree *Populus trichocarpa*: II. Heritability estimates, correlations among explant types, and genetic interactions with treatments among wild genotypes. *In Vitro Cellular & Developmental Biology – Plant* **58**: 853–864.

Ma C, Goddard A, Peremyslova E, Duan C, Jiang Y, Nagle M, Strauss SH. 2022b. Factors affecting *in vitro* regeneration in the model tree *Populus trichocarpa* I. Medium, environment, and hormone controls on organogenesis. *In Vitro Cellular & Developmental Biology – Plant* **58**: 837–852.

Ma C, Strauss SH, Meilan R. 2004. *Agrobacterium*-mediated transformation of the genome-sequenced poplar clone, nisqually-1 (*Populus trichocarpa*). *Plant Molecular Biology Reporter* **22**: 311–312.

Ma Y, Miotk A, Šutiković Z, Ermakova O, Wenzl C, Medzihradszky A, Gaillochet C, Forner J, Utan G, Brackmann K *et al.* 2019. WUSCHEL acts as an auxin response rheostat to maintain apical stem cells in Arabidopsis. *Nature Communications* **10**: 5093.

Majda M, Robert S. 2018. The role of auxin in cell wall expansion. *International Journal of Molecular Sciences* **19**: 951.

Mao J-L, Miao Z-Q, Wang Z, Yu L-H, Cai X-T, Xiang C-B. 2016. Arabidopsis ERF1 mediates cross-talk between ethylene and auxin biosynthesis during primary root elongation by regulating ASA1 expression. *PLoS Genetics* **12**: e1005760.

Mayinger P. 2012. Phosphoinositides and vesicular membrane traffic. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids* **1821**: 1104–1113.

Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD. 2019. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols* **14**: 703–721.

Monaghan J, Xu F, Gao M, Zhao Q, Palma K, Long C, Chen S, Zhang Y, Li X. 2009. Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathogens* **5**: e1000526.

Mookkan M, Nelson-Vasilchik K, Hague J, Zhang ZJ, Kausch AP. 2017. Selectable marker independent transformation of recalcitrant maize inbred B73 and sorghum P898012 mediated by morphogenic regulators BABY BOOM and WUSCHEL2. *Plant Cell Reports* **36**: 1477–1491.

Muchero W, Sondreli KL, Chen J-G, Urbanowicz BR, Zhang J, Singan V, Yang Y, Brueggeman RS, Franco-Coronado J, Abraham N *et al.* 2018. Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant–pathogen interactions in a tree. *Proceedings of the National Academy of Sciences, USA* **115**: 11573–11578.

Nagle M, Déjardin A, Pilate G, Strauss SH. 2018. Opportunities for innovation in genetic transformation of forest trees. *Frontiers in Plant Science* **9**: 1443.

Nagle M, Yuan J, Kaur D, Ma C, Peremyslova E, Jiang Y, Goralogia G, Magnuson A, Li JY, Muchero W *et al.* 2024. *RGB and hyperspectral phenomics dataset for in vitro transformation and regeneration of* Populus trichocarpa [Data set]. Oak Ridge, TN, USA: Oak Ridge National Laboratory (ORNL).

Nagle MF, Yuan J, Kaur D, Ma C, Peremyslova E, Jiang Y, Niño de Rivera A, Jawdy S, Chen J-G, Feng K *et al.* 2024. GWAS supported by computer vision identifies large numbers of candidate regulators of *in planta* regeneration in *Populus trichocarpa*. *G3: Genes, Genomes, Genetics* jkae026.

Nagle MF, Yuan J, Kaur D, Ma C, Peremyslova E, Jiang Y, Zahl B, de Rivera AN, Muchero W, Fuxin L *et al.* 2023. GWAS identifies candidate genes controlling adventitious rooting in *Populus trichocarpa*. *Horticulture Research* **10**: uhad125.

Nawy T, Bayer M, Mravec J, Friml J, Birnbaum KD, Lukowitz W. 2010. The GATA factor HANABA TARANU is required to position the proembryo boundary in the early Arabidopsis embryo. *Developmental Cell* **19**: 103–113.

Nelson-Vasilchik K, Hague JP, Tilelli M, Kausch AP. 2022. Rapid transformation and plant regeneration of sorghum (*Sorghum bicolor* L.) mediated by altruistic Baby boom and Wuschel2. *In Vitro Cellular & Developmental Biology – Plant* **58**: 331–342.

Nguyen THN, Winkelmann T, Debener T. 2020. Genetic analysis of callus formation in a diversity panel of 96 rose genotypes. *Plant Cell, Tissue and Organ Culture* **142**: 505–517.

Pan C, Li G, Malzahn AA, Cheng Y, Leyson B, Sretenovic S, Gurel F, Coleman GD, Qi Y. 2022. Boosting plant genome editing with a versatile CRISPR-Combo system. *Nature Plants* **8**: 513–525.

Perna V, Meyer AS, Holck J, Eltis LD, Eijsink VGH, Wittrup Agger J. 2020. Laccase-catalyzed oxidation of lignin induces production of H_2O_2. *ACS Sustainable Chemistry & Engineering* **8**: 831–841.

Piya S, Liu J, Burch-Smith T, Baum TJ, Hewezi T. 2020. A role for Arabidopsis growth-regulating factors 1 and 3 in growth–stress antagonism. *Journal of Experimental Botany* **71**: 1402–1417.

Qin S, Fan C, Li X, Li Y, Hu J, Li C, Luo K. 2020. LACCASE14 is required for the deposition of guaiacyl lignin and affects cell wall digestibility in poplar. *Biotechnology for Biofuels* **13**: 197.

Ricordel C, Chaillot L, Blondel A, Archambeau J, Jouan F, Mouche A, Tiercin M, Burel A, Lena H, Desrues B *et al.* 2021. ING2 tumor suppressive protein translocates into mitochondria and is involved in cellular metabolism homeostasis. *Oncogene* **40**: 4111–4123.

Robinson WD, Park J, Tran HT, Del Vecchio HA, Ying S, Zins JL, Patel K, McKnight TD, Plaxton WC. 2012. The secreted purple acid phosphatase isozymes AtPAP12 and AtPAP26 play a pivotal role in extracellular phosphate-scavenging by *Arabidopsis thaliana*. *Journal of Experimental Botany* **63**: 6531–6542.

Ryan NW. 2022. *Overexpression of the GROWTH REGULATING FACTOR 4-GRF-INTERACTING FACTOR 1 transcription factor chimera modifies transformation and regeneration efficiency in* Populus *and* Eucalyptus. Master's thesis, Oregon State University, Corvallis, OR, USA. ScholarsArchive@OSU. ID: s4655q40j.

Šašek V, Janda M, Delage E, Puyaubert J, Guivarc'h A, López Maseda E, Dobrev PI, Caius J, Bóka K, Valentová O *et al.* 2014. Constitutive salicylic acid accumulation in pi4kIIIβ1β2 Arabidopsis plants stunts rosette but not root growth. *New Phytologist* **203**: 805–816.

Schröder F, Lisso J, Lange P, Müssig C. 2009. The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biology* **9**: 20.

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Peña P, Lan F, Kaadige MR *et al.* 2006. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**: 96–99.

Shu K, Zhou W, Chen F, Luo X, Yang W. 2018. Abscisic acid and gibberellins antagonistically mediate plant development and abiotic stress responses. *Frontiers in Plant Science* **9**: 416.

Skirycz A, Radziejwoski A, Busch W, Hannah MA, Czeszejko J, Kwaśniewski M, Zanor M-I, Lohmann JU, De Veylder L, Witt I *et al.* 2008. The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. *The Plant Journal* **56**: 779–792.

Song J, Lu S, Chen Z-Z, Lourenco R, Chiang VL. 2006. Genetic transformation of *Populus trichocarpa* genotype nisqually-1: a functional genomic tool for woody plants. *Plant and Cell Physiology* **47**: 1582–1589.

Spoel SH, Johnson JS, Dong X. 2007. Regulation of tradeoffs between plant defenses against pathogens with different lifestyles. *Proceedings of the National Academy of Sciences, USA* **104**: 18842–18847.

Srinivasan C, Liu Z, Heidmann I, Supena EDJ, Fukuoka H, Joosen R, Lambalk J, Angenent G, Scorza R, Custers JBM *et al.* 2007. Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (*Nicotiana tabacum* L.). *Planta* **225**: 341–351.

Starodubtseva A, Kalachova T, Retzer K, Jelínková A, Dobrev P, Lacek J, Pospíchalová R, Angelini J, Guivarc'h A, Pateyron S *et al.* 2022. An Arabidopsis mutant deficient in phosphatidylinositol-4-phosphate kinases β1 and β2 displays altered auxin-related responses in roots. *Scientific Reports* **12**: 6947.

Sterjiades R, Dean JFD, Eriksson K-EL. 1992. Laccase from sycamore maple (*Acer pseudoplatanus*) polymerizes monolignols. *Plant Physiology* **99**: 1162–1168.

Szczygieł-Sommer A, Gaj MD. 2019. The miR396–GRF regulatory module controls the embryogenic response in Arabidopsis via an auxin-related pathway. *International Journal of Molecular Sciences* **20**: 5221.

Taiz L, Zeiger E, Møller I, Murphy A. 2015. *Plant physiology & development*. Sunderland, CT, USA: Sinauer Associates.

Takenaka M, Verbitskiy D, Zehrmann A, Brennicke A. 2010. Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *Journal of Biological Chemistry* **285**: 27122–27129.

Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. 2022. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Science* **31**: 8–22.

Thorpe TA. 2007. History of plant tissue culture. *Molecular Biotechnology* **37**: 169–180.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Tuskan GA, Mewalal R, Gunter LE, Palla KJ, Carter K, Jacobson DA, Jones PC, Garcia BJ, Weighill DA, Hyatt PD *et al.* 2018. Defining the genetic components of callus formation: a GWAS approach. *PLoS ONE* **13**: e0202519.

Vsevolozhskaya OA, Hu F, Zaykin DV. 2019. Detecting weak signals by combining small p-values in genetic association studies. *Frontiers in Genetics* **10**: 1051.

Weighill D, Tschaplinski TJ, Tuskan GA, Jacobson D. 2019. Data integration in poplar: 'omics layers and integration strategies. *Frontiers in Genetics* **10**: 874.

Yang L, Ping T, Lu W, Song S, Wang J, Wang Q, Chai G, Bai Y, Chen Y. 2023. Genome-wide identification of auxin-responsive microRNAs in the poplar stem. *Genes & Genomics* **45**: 1073–1083.

Yu X, Gong H, Cao L, Hou Y, Qu S. 2020. MicroRNA397b negatively regulates resistance of *Malus hupehensis* to *Botryosphaeria dothidea* by modulating MhLAC7 involved in lignin biosynthesis. *Plant Science* **292**: 110390.

Yuan J, Kaur D, Zhou Z, Nagle M, Kiddle NG, Doshi NA, Behnoudfar A, Peremyslova E, Ma C, Strauss SH *et al.* 2022. Robust high-throughput phenotyping with deep segmentation enabled by a web-based annotator. *Plant Phenomics* **2022**: 9893639.

Zhang G, Liu W, Gu Z, Wu S, E Y, Zhou W, Lin J, Xu L. 2023. Roles of the wound hormone jasmonate in plant regeneration. *Journal of Experimental Botany* **74**: 1198–1206.

Zhang J, Yang Y, Zheng K, Xie M, Feng K, Jawdy SS, Gunter LE, Ranjan P, Singan VR, Engle N *et al.* 2018. Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in *Populus*. *New Phytologist* **220**: 502–516.

Zhang Q, Su Z, Guo Y, Zhang S, Jiang L, Wu R. 2020. Genome-wide association studies of callus differentiation for the desert tree, *Populus euphratica*. *Tree Physiology* **40**: 1762–1777.

Zhang X, Zhou Y, Ding L, Wu Z, Liu R, Meyerowitz EM. 2013. Transcription repressor HANABA TARANU controls flower development by integrating the actions of multiple hormones, floral organ specification genes, and GATA3 family genes in Arabidopsis. *Plant Cell* **25**: 83–101.

Zhao Y, Lin S, Qiu Z, Cao D, Wen J, Deng X, Wang X, Lin J, Li X. 2015. MicroRNA857 is involved in the regulation of secondary growth of vascular tissues in Arabidopsis. *Plant Physiology* **169**: 2539–2552.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.

Nagle, M., Yuan, J., Kaur, D., Ma, C., Peremyslova, E., Jiang, Y., Goralogia, G., Magnuson, A., Li, J. Y., Muchero, W., Li, F. X., & Strauss, S. (2024). RGB and hyperspectral phenomics dataset for in vitro transformation and regeneration of Populus trichocarpa [Data set]. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). https://doi.org/10.11578/2331261

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Plasmid used for transformation treatment in genome-wide association study.

Fig. S2 Examples of explants across experimental timepoints displaying varying levels of transgenic callus and shoot.

Fig. S3 Callus regeneration principal component analysis.

Fig. S4 Callus transformation principal component analysis.

Fig. S5 Comprehensive principal component analysis.

Fig. S6 Shoot transformation principal component analysis.

Fig. S7 Transgenic tissue frequency principal component analysis.

Fig. S8 Portion of gibberellin metabolic pathway regulated by homologs of two candidate genes.

Methods S1 GWAS population.

Methods S2 Assays of *in vitro* regeneration and transformation.

Methods S3 Image acquisition.

Methods S4 Computer vision workflows.

Methods S5 Principal component analysis.

Methods S6 Association mapping.

Methods S7 Gene ontology and pathway overrepresentation testing.

Methods S8 Network analysis.

Methods S9 Epistasis analysis.

Notes S1 Literature review of additional candidate genes.

Table S1 Overview of three sterilization methods used for tissue culture and listing of sterilization methods for each experimental phase.

Table S2 Metadata for *in vitro* traits, including methods and results of transformations and SNP heritability as computed by Genome-wide Efficient Mixed Model Association.

Table S3 Summary statistics of *in vitro* associations with respect to methods, significance thresholds, and genes most closely implicated by quantitative trait locus peaks.

Table S4 *In vitro* callus and shoot regeneration and transformation associations passing Bonferroni and/or Benjamini–Hochberg False Discovery Rate (FDR; $\alpha = 0.10$) thresholds.

Table S5 *In vitro* callus and shoot regeneration and transformation associations passing Augmented Rank Truncation-Bonferroni threshold.

Table S6 Numbers of significant epistatic interactions detected by Factored Spectrally Transformed Linear Mixed Models, tallied by candidate gene and significance threshold.

Table S7 Detailed results from Factored Spectrally Transformed Linear Mixed Models tests for epistatic interactions among candidate genes from genome-wide association study.

Table S8 Interactions among Arabidopsis homologs of gene candidates and their intermediates.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.