# Pleiotropic and Epistatic Network-Based Discovery: Integrated Networks for Target Gene Discovery

Deborah Weighill[1,2], Piet Jones[1,2], Manesh Shah[2], Priya Ranjan[2], Wellington Muchero[2], Jeremy Schmutz[4,5], Avinash Sreedasyam[5], David Macaya-Sanz[6], Robert Sykes[3], Nan Zhao[7], Madhavi Z. Martin[2], Stephen DiFazio[6], Timothy J. Tschaplinski[2], Gerald Tuskan[2] and Daniel Jacobson[*1,2]

[1]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, Knoxville, TN, USA
[2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[3]National Renewable Energy Laboratory, Golden, CO, USA
[4]Department of Energy Joint Genome Institute, Walnut Creek, CA, USA
[5]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
[6]Department of Biology, West Virginia University, Morgantown, WV, USA
[7]The University of Tennessee Institute of Agriculture, University of Tennessee, Knoxville, Knoxville, TN, USA
[*]jacobsonda@ornl.gov

## Abstract

*Biological organisms are complex systems that are composed of functional networks of interacting molecules and macromolecules. Complex phenotypes are the result of orchestrated, hierarchical, heterogeneous collections of expressed genomic variants. However, the effects of these variants are the result of historic selective pressure and current environmental and epigenetic signals, and, as such, their co-occurrence can be seen as genome-wide correlations in a number of different manners. Biomass recalcitrance (i.e., the resistance of plants to degradation or deconstruction, which ultimately enables access to a plant's sugars) is a complex polygenic phenotype of high importance to biofuels initiatives. This study makes use of data derived from the re-sequenced genomes from over 800 different Populus trichocarpa genotypes in combination with metabolomic and pyMBMS data across this population, as well as co-expression and co-methylation networks in order to better understand the molecular interactions involved in recalcitrance, and identify target genes involved in lignin biosynthesis/degradation. A Lines Of Evidence (LOE) scoring system is developed to integrate the information in the different layers and quantify the number of lines of evidence linking genes to lignin-related lignin-phenotypes across the network layers. The resulting Genome Wide Association Study networks, integrated with Single Nucleotide Polymorphism (SNP) correlation, co-methylation and co-expression networks through the LOE scores are proving to be a powerful approach to determine the pleiotropic and epistatic relationships underlying cellular functions and, as such, the molecular basis for complex phenotypes, such as recalcitrance.*

## 1. Keywords:

*Multi-omic data layering, LOE Scores, Lines of Evidence Scores, GWAS, SNP correlation, networks, lignin, recalcitrance, bioenergy, co-expression, co-methylation, metabolomics, pyMBMS*

## 2. Introduction

*Populus* species are promising sources of cellulosic biomass for biofuels because of their fast growth rate, high cellulose content and moderate lignin content (Sannigrahi et al., 2010). Ragauskas et al. (2006) outline areas of

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

research needed "to increase the impact, efficiency, and sustainability of bio-refinery facilities" (Ragauskas et al., 2006), such as research into modifying plants to enhance favorable traits, including altered cell wall structure leading to increased sugar release, as well as resilience to biotic and abiotic stress. One particular research target in *Populus* species is the decrease/alteration of the lignin content of cell walls.

A large collection of different data types has been generated for *Populus trichocarpa*. The genome has been sequenced and annotated (Tuskan et al., 2006), and the assembly is currently in its third version of revision. A collection of 1,100 accessions of *P. trichocarpa* that have been clonally propagated in four different common gardens (Tuskan et al., 2011; Slavov et al., 2012; Evans et al., 2014) have been resequenced, which has provided a large set of ∼ 28,000,000 Single Nucleotide Polymorphisms (SNPs) that has recently been publicly released (`http://bioenergycenter.org/besc/gwas/`). Many molecular phenotypes, such as untargetted metabolomics and pyMBMS phenotypes, that have been measured in this population provide an unparalleled resource for Genome Wide Association Studies (for example, see McKown et al. (2014)). DNA methylation data in the form of MeDIP (Methyl-DNA immunoprecipitation)-seq has been performed on 10 different *P. trichocarpa* tissues (Vining et al., 2012), and gene expression has been measured across various tissues and conditions.

This study involves integrating these various data types in order to identify new possible candidate genes involved in lignin biosynthesis/degradation/regulation. Integrating Genome Wide Association Study (GWAS) data with other data types has previously been done to help provide context and identify relevant subnetworks/modules (Calabrese et al., 2017; Bunyavanich et al., 2014). Ritchie et al. (2015) reviewed techniques for integrating various data types for the aim of investigating gene-phenotype associations. Integrating multiple lines of evidence is a useful strategy as the more lines of evidence that connect a gene to a phenotype lowers the chance of false positives. Ritchie et al. (2015) categorized data integration approaches into two main classes, namely multi-staged analysis and meta-dimensional analysis. Multi-staged analysis analyses aims to enrich a biological signal through various steps of analysis. Meta-dimensional analysis involves the concurrent analysis of various data types, and is divided into three subcategories (Ritchie et al., 2015): Concatenation-based integration concatenates the data matrices of different data types into a single matrix on which a model is constructed (for example, see Fridley et al. (2012)). Model-based integration involves constructing a separate model for each dataset and then constructing a final model from the results of the separate models (for example, see Kim et al. (2013)). Transformation-based integration involves transforming transforming each data type into a common form (e.g. a network) before combining them (see for example, Kim et al. (2012)).

This study involves the development of an approach which can be seen as a type of transformation-based integration. Association networks for various different data types were constructed, including a pyMBMS GWAS network, a metabolomics GWAS network, as well as co-expression, co-methylation and SNP correlation networks, and subsequently the information in the different networks was integrated through the calculation of the newly developed Lines Of Evidence (LOE) scores defined in this study. These scores quantify the number of lines of evidence connecting each gene to lignin-related genes and phenotypes. This multi-omic data integration approach allowed for the identification of new possible candidate genes involved in lignin biosynthesis/regulation through multiple lines of evidence.

# 3. Methods

## 3.1. Overview

This approach involved combining various data types in order to identify new possible target genes involved in lignin biosynthesis/degradation/regulation. Figure 1 summarizes the overall approach. First, association networks were constructed including metabolomics and pyMBMS GWAS networks, co-expression, co-methylation and SNP correlation networks. Known lignin-related genes and phenotypes were then identified, and used as seeds to select lignin-related subnetworks from these various networks. The Lines Of Evidence (LOE) scoring technique was developed, and each gene was then scored based on its Lines Of Evidence linking it to lignin-related genes and phenotypes.

## 3.2. GWAS Network Construction

### 3.2.1 Metabolomics Data

The *P. trichocarpa* leaf samples for 851 unique clones were collected over three consecutive sunny days in July 2012. For 200 of those clones, a second biological replicate was also sampled. Typically, leaves (leaf plastocron index

9 plus or minus 1) on a south facing branch from the upper canopy of each tree were quickly collected, wiped with a wet tissue to clean both surfaces and the leaf then fast frozen under dry ice. Leaves were kept on dry ice and shipped back to the lab and stored at -80°C until processed for analyses. Metabolites from leaf samples were lyophilized and then ground in a micro-Wiley mill (1 mm mesh size). Approximately 25 mg of each sample was twice extracted in 2.5 mL 80% ethanol (aqueous) for 24 hr with the extracts combined, and 0.5 ml dried in a helium stream. Sorbitol (75 µl of a 1 mg/mL aqueous solution) was added before extraction as an internal standard to correct for differences in extraction efficiency, subsequent differences in derivatization efficiency and changes in sample volume during heating. Metabolites in the dried sample extracts were converted to their trimethylsilyl (TMS) derivatives, and analyzed by gas chromatography-mass spectrometry, as described previously (Tschaplinski et al., 2012; Li et al., 2012). Briefly, dried extracts of metabolites were dissolved in acetonitrile followed by the addition of N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS), and samples then heated for 1 h at 70°C to generate TMS derivatives. After 2 days, aliquots were injected into an Agilent 5975C inert XL gas chromatograph-mass spectrometer (GCMS). The standard quadrupole GCMS is operated in the electron impact (70 eV) ionization mode, targeting 2.5 full-spectrum (50-650 Da) scans per second, as described previously (Tschaplinski et al., 2012). Metabolite peaks were extracted using a key selected ion, characteristic m/z fragment, rather than the total ion chromatogram, to minimize integrating co-eluting metabolites. The peak areas were normalized to the amount of internal standard (sorbitol) injected and the amount of sample extracted. A large user-created database (>2400 spectra) of mass spectral electron impact ionization (EI) fragmentation patterns of TMS-derivatized metabolites, as well as the Wiley Registry 10th Edition combined with NIST 2014 mass spectral library, are used to identify the metabolites of interest to be quantified.

### 3.2.2 pyMBMS Data

A commercially available molecular beam mass spectrometer (MBMS) designed specifically for biomass analysis was used for pyrolysis vapor analysis (Evans and Milne, 1987; Sykes et al., 2009; Tuskan et al., 1999). Approximately 4 mg of air dried 20 mesh biomass was introduced into the quartz pyrolysis reactor via 80 uL deactivated stainless steel Eco-Cups provided with the autosampler. Mass spectral data from m/z 30-450 were acquired on a Merlin Automation data system version 3.0 using 17 eV electron impact ionization.

The pyMBMS mz peaks were annotated as described in (Sykes et al., 2009), as done previously in (Muchero et al., 2015).

### 3.2.3 Single Nucleotide Polymorphism Data

A dataset consisting of 28,342,758 SNPs called across 882 *P. trichocarpa* (Tuskan et al., 2006) genotypes was obtained from `http://bioenergycenter.org/besc/gwas/`. This dataset is derived from whole genome sequencing of undomesticated *P. trichocarpa* genotypes collected from the U.S. and Canada, and clonally replicated in common gardens (Tuskan et al., 2011). Genotypes from this population have previously been used for population genomics (Evans et al., 2014) and GWAS studies in *P. trichocarpa* (McKown et al., 2014) as well as for investigating linkage disequilibrium in the population (Slavov et al., 2012).

Whole genome resequencing was carried out on a sample 882 P. trichocarpa natural individuals to an expected median coverage of 15x using Illumina Genome Analyzer, HiSeq 2000, and HiSeq 2500 sequencing platforms at the DOE Joint Genome Institute. Alignments to the *P. trichocarpa* Nisqually-1 v.3.0 reference genome were performed using BWA v0.5.9-r16 with default parameters, followed by post-processing with the picard FixMateInformation and MarkDuplicates tools. Genetic variants were called by means of the Genome Analysis Toolkit v. 3.5.0 (GATK; Broad Institute, Cambridge, MA, USA) (McKenna et al., 2010; Van der Auwera et al., 2013). Briefly, variants were called independently for each individual using the concatenation of RealignerTargetCreator, IndelRealigner and HaplotypeCaller tools, and the whole population was combined using GenotypeGVCFs, obtaining a dataset with all the variants detected across the sample population. Biallelic SNPs were extracted using the SelectVariants tool and quality-filtered using the GATKâĂŹs machine-learning implementation Variant Quality Score Recalibration (VQSR). To this end, the tool VariantRecalibrator was used to create the recalibration file and the sensitivity tranches file. As a "truth" dataset, we used SNP calls from a population of seven female and seven male *P. trichocarpa* that had been crossed in a half diallel design. "True" SNPs were identified by the virtual absence of segregation distortion and Mendelian violations in the progeny of these 49 crosses (ca. 500 offspring in total). As a "non-true" dataset, we used the SNP calls of seven open-pollinated crosses from these 7 females ($n = 90$), filtered using hard-filtering methods recommended in the GATK documentation (tool: VariantFiltration; quality thresholds: $QD < 1.5$, $FS > 75.0$, $MQ < 35.0$, missing alleles $< 0.5$ and $MAF > 0.05$). The prior likelihoods for the true and non-true datasets were Q = 15 and Q = 10, respectively, and the variant quality annotations to define the variant recalibration space were DP, QD, MQ, MQRankSum, ReadPosRankSum, FS, SOR and InbreedingCoeff. Finally, we used the ApplyRecalibration tool on the full GWAS dataset to assign SNPs to tranches representing different levels of confidence. We selected SNPs in the tranche with true sensitivity $< 90$, which minimizes false positives, but at an expected cost of 10% false negatives. The final filtered dataset had a transition/transversion ratio of 2.07, compared to 1.88 for the unfiltered SNPs. To further validate the quality of these SNP calls, we compared them to an Illumina Infinium BeadArray that had been generated from a subset of this population dataset (Geraldes et al., 2013). The average match rate was 96% (±2% SD) for 641 individuals across 20,723 loci.

SNPs in this dataset were divided into different Tranches, indicating the percentage of "true" SNPs recovered. For further analysis in this study, we made use of the PASS SNPs, corresponding to the most stringent Tranche, recovering 90% of the true SNPs [ see `http://gatkforums.broadinstitute.org/gatk/discussion/39/variant-quality-score-recalibration-vqsr`].
VCFtools (Danecek et al., 2011) was used to extract the desired Tranche of SNPs from the VCF file and reformat it into .tfam and .tped files.

### 3.2.4 GWAS Analysis

The metabolomics and pyMBMS data was used as phenotypes in a genome wide association analysis. The respective phenotype measured over all the genotypes were analyzed to account for potential outliers. A median absolute deviation (MAD) from the median (Leys et al., 2013) cutoff was applied to determine if a particular measurement of a given phenotype was an outlier with respect to all measurements of that phenotype across the population. To account for asymmetry, the deviation values were estimated separately for values below and above the median, respectively. The distribution of the measured values together with the distribution of their estimated deviation was analyzed and a cutoff of 5 was determined to identify putative outlier values. Phenotypes that had non-outlier measurements in at least 20 percent of the population were retained for further analysis, this was to ensure sufficient signal for the genome wide association model. This resulted in 1262 pyMBMS derived phenotypes and 818 metabolomics derived phenotypes.

To estimate the statistical significant associations between the respective phenotypes and the SNPs called across the population, we applied a linear mixed model using EMMAX Kang et al. (2010). Taking into account population structure estimated from a kinship matrix, we tested each of the respective 2080 phenotypes against the high-confidence SNPs and corrected for multiple hypotheses bias using the Benjamini-Hochberg control for false-discovery rate of 0.1 Benjamini and Hochberg (1995). This was done in parallel with a python wrapper that utilized the schwimmbad python package (Price-Whelan and Foreman-Mackey, 2017).

SNP-Phenotype GWAS networks were then pruned to only include SNPs that resided within genes, and SNPs were mapped to their respective genes, resulting in a gene-phenotype network. SNPs were determined to be within genes using the gene boundaries defined in the `Ptrichocarpa_210_v3.0.gene.gff3` from the *P. trichocarpa* version 3.0 genome assembly on Phytozome (Goodstein et al., 2012).

## 3.3. Co-Expression Network Construction

*Populus trichocarpa* (Nisqually-1) RNA-seq dataset from JGI Plant Gene Atlas project (Sreedasyam et al., unpublished) was obtained from Phytozome. This dataset consists of samples for standard tissues (leaf, stem, root and bud tissue) and libraries generated from nitrogen source study. List of sample descriptions was accessed from: `https://phytozome.jgi.doe.gov/phytomine/aspect.do?name=Expression`.

### 3.3.1 Plant growth and treatment conditions

*Populus trichocarpa* (Nisqually-1) cuttings were potted in 4″ X 4″ X 5″ containers containing 1:1 mix of peat and perlite. Plants were grown under 16-h-light/8-h-dark conditions, maintained at 20-23 °C and an average of 235 $\mu$mol m$^{-2}$s$^{-1}$ to generate tissue for (1) standard tissues and (2) nitrogen source study. Plants for standard tissue experiment were watered with McCownâĂŹs woody plant nutrient solution and plants for nitrogen experiment were supplemented with either 10mM KNO3 (NO3− plants) or 10mM NH4Cl (NH4+ plants) or 10 mM urea (urea plants). Once plants reached leaf plastochron index 15 (LPI-15), leaf, stem, root and bud tissues were harvested and immediately flash frozen in liquid nitrogen and stored at -80°C until further processing was done. Every harvest involved at least three independent biological replicates for each condition and a biological replicate consisted of tissue pooled from 3 plants.

### 3.3.2 RNA extraction and sequencing

Tissue was ground under liquid nitrogen and high quality RNA was extracted using standard Trizol-reagent based extraction (Li and Trick, 2005). The integrity and concentration of the RNA preparations were checked initially using Nano-Drop ND-1000 (Nano-Drop Technologies) and then by BioAnalyzer (Agilent Technologies). Plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina's TruSeq Stranded mRNA HT sample prep kit utilizing poly-A selection of mRNA following the protocol outlined by Illumina in their user guide: `http://support.illumina.com/sequencing/sequencing_kits/truseq_stranded_mrna_ht_sample_prep_kit.html`, and with the following conditions: total RNA starting material was 1 ug per sample and 8 cycles of PCR was used for library amplification. The prepared libraries were then quantified by qPCR using the Kapa SYBR Fast Illumina Library Quantification Kit (Kapa Biosystems) and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4, and IlluminaâĂŹs cBot instrument to generate a clustered flowcell for sequencing. Sequencing of the flowcell was performed on the Illumina HiSeq2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2x150 indexed run recipe.

### 3.3.3 Correlation Analysis

Gene expression atlas data for *P. trichocarpa* consisting of 63 different samples were used to construct a co-expression network. Reads were trimmed using Skewer (Jiang et al., 2014). Star (Dobin et al., 2013) was then used to align the reads to the *P. trichocarpa* reference genome (Tuskan et al., 2006) obtained from Phytozome (Goodstein et al., 2012). TPM (Transcripts Per Million) expression values (Wagner et al., 2012) were then calculated for each gene. This resulted in a gene expression matrix $E$ in which rows represented genes, columns represented samples and each entry $ij$ represented the expression (TPM) of gene $i$ in sample $j$. The Spearman correlation coefficient was then calculated between the expression profiles of all pairs of genes (i.e. all pairs of rows of the matrix $E$) using the mcxarray and mcxdump programs from the MCL-edge package (Van Dongen, 2008, 2001) available from `http://micans.org/mcl/`. This was performed in parallel using Perl wrappers making use of the Parallel::MPI::Simple Perl module, (Alex Gough, `http://search.cpan.org/~ajgough/Parallel-MPI-Simple-0.03/Simple.pm`) using compute resources at the Oak Ridge Leadership Computing Facility (OLCF).

Supplementary Figure S1A shows the distribution of Spearman correlation values for the co-expression network. An absolute threshold of 0.85 was applied.

## 3.4. Co-Methylation Network Construction

Methylation data for *P. trichocarpa* (Vining et al., 2012) re-aligned to the version 3.0 assembly of *P. trichocarpa* was obtained from Phytozome (Goodstein et al., 2012). This data consisted of MeDIP-seq (Methyl-DNA immunoprecipitation-seq) reads from 10 different *P. trichocarpa* tissues, including bud, callus, female catkin, internode explant, leaf, male catkin, phloem, regenerated internode, root and xylem tissue.

BamTools stats (Barnett et al., 2011) was used to determine basic properties of the reads in each .bam file. Samtools (Li et al., 2009) was then used to extract only mapped reads. The number of reads which mapped to each gene feature was determined using htseq-count (Anders et al., 2014). These read counts were then converted to TPM values (Wagner et al., 2012), providing a methylation score for each gene in each tissue. The TPM value for a gene $g$ in a given sample was defined as:

$$TPM_g = \frac{\frac{c_g}{l_g} \times 10^6}{\sum_g \frac{c_g}{l_g}} \tag{1}$$

where $c_g$ is the number of reads mapped to gene $g$ and $l_g$ is the length of gene $g$ in kb, calculated by subtracting the gene start position from the gene end position, and dividing the resulting difference by 1,000. A methylation matrix $M$ was then formed, in which rows represented genes, columns represented tissues and each entry $ij$ represented the methylation score (TPM) of gene $i$ in tissue $j$. A co-methylation network (see references (Busch et al., 2016; Akulenko and Helms, 2013; Davies et al., 2012)) was then constructed by calculating the Spearman correlation coefficient between the methylation profiles of all pairs of genes using mcxarray and mcxdump programs from the MCL-edge package (Van Dongen, 2008, 2001) `http://micans.org/mcl/`. Supplementary Figure S1B shows the distribution of Spearman Correlation values. An absolute threshold of 0.95 was applied.

Read counting using htseq-count, as well as Spearman correlation calculations were performed in parallel using Perl wrappers making use of the Parallel::MPI::Simple Perl module, developed by Alex Gough and available on The Comprehensive Perl Archive Network (CPAN) at `www.cpan.org` and used compute resources at the Oak Ridge Leadership Computing Facility (OLCF).

## 3.5. SNP Correlation Network Construction

The Custom Correlation Coefficient (CCC) (Climer et al., 2014b,a) was used to calculate the correlation between the occurrence of pairs of SNPs across the 882 genotypes. The CCC between allele $x$ at position $i$ and allele $y$ and position $j$ is defined as:

$$CCC_{i_x j_y} = \frac{9}{2} R_{i_x j_y} \left(1 - \frac{1}{f_{i_x}}\right) \left(1 - \frac{1}{f_{j_y}}\right) \tag{2}$$

where $R_{i_x j_y}$ is the relative co-occurrence of allele $x$ at position $i$ and allele $y$ at position $j$, $f_{i_x}$ is the frequency of allele $x$ at position $i$ and $f_{j_y}$ is the frequency of allele $y$ at position $j$.

This was performed in a parallel fashion using similar computational approaches as described for the co-expression network above. The set of ~10 million SNPs was divided into 20 different blocks, and the CCC was calculated for each within-block and cross-block SNPs in separate jobs, to a total of 210 MPI jobs (Figure 2). A threshold of 0.7 was then applied. The resulting SNP correlation network was pruned to only include SNPs that resided within genes. Gene boundaries used were defined in the `Ptrichocarpa_210_v3.0.gene.gff3` from the *P. trichocarpa* version 3.0 genome assembly on Phytozome (Goodstein et al., 2012). A local LD filter was then set, retaining correlations between SNPs greater than 10kb apart. The distribution of CCC values can be seen in Supplementary Figure S1C (Supplementary Note 1).

## 3.6. Gene Annotation

*P. trichocarpa* gene annotations in the `Ptrichocarpa_210_v3.0.annotation_info.txt` file from the version 3.0 genome assembly were used, available on Phytozome (Goodstein et al., 2012). This included *Arabidopsis* best hits and corresponding gene descriptions, as well as GO terms (Gene Ontology Consortium, 2017; Ashburner et al., 2000) and Pfam domains (Finn et al., 2016). Genes were also assigned MapMan annotations using the Mercator tool (Lohse et al., 2014).

## 3.7. Scoring Lines of Evidence (LOE)

A scoring system was developed in order to quantify the Lines Of Evidence (LOE) linking each gene to lignin-related genes/phenotypes. The LOE scores quantify the number of lines linking each gene to lignin-related genes and phenotypes across the different network data layers. The process of defining and calculating LOE scores is described below.

### 3.7.1 Selection of Lignin-related Genes

Lignin building blocks (monolignols) are derived from phenylalanine in the phenylpropanoid and monolignol pathways, and phenylalanine itself is produced from the shikimate pathway (Vanholme et al., 2010). To compile a list of *P. trichocarpa* genes which are related to the biosynthesis of lignin, *P. trichocarpa* genes were assigned MapMan annotations using the Mercator tool (Lohse et al., 2014). Genes in the Shikimate (MapMan bins 13.1.6.1, 13.1.6.3 and 13.1.6.4), Phenylpropanoid (MapMan bin 16.2) and Lignin/Lignan (MapMan bin 16.2.1) pathways were then selected. A list of these lignin-related genes and their MapMan annotations can be seen in Supplementary Table S1.

### 3.7.2 Selection of Lignin-related Phenotypes

Lignin-related pyMBMS peaks, as described in Sykes et al. (2009), Davis et al. (2006) and Muchero et al. (2015) were identified among the pyMBMS GWAS hits, and are shown in Supplementary Table S2. Lignin-related metabolites and metabolites in the lignin pathway were also identified among the metabolomics GWAS hits, a list of which can be seen in Supplementary Table S3. For partially identified metabolites, additional RT and mz information can be seen in Supplementary Table S3.

#### 3.7.3 Extraction of Lignin-Related Subnetworks

Let $L_G$, $L_M$ and $L_P$ represent our sets of lignin-related genes, metabolites and pyMBMS peaks, respectively (Supplementary Tables S1, S2 and S3). A network can be defined as $N = (V, E)$ where $V$ is the set of nodes and $E$ is the set of edges connecting nodes in $V$. In particular, let the co-expression network be represented by $N_{coex} = (V_{coex}, E_{coex})$, the co-methylation network by $N_{cometh} = (V_{cometh}, E_{cometh})$ and the SNP correlation network by $N_{snp} = (V_{snp}, E_{snp})$. The GWAS networks can be represented as bipartite networks $N = (U, V, E)$ where $U$ is the set of phenotype nodes, $V$ is the set of gene nodes, and $E$ is the set of edges, with each edge $e_{ij}$ connecting node $i \in U$ with node $j \in V$. Let the metabolomics GWAS network be represented by $N_{metab} = (U_{metab}, V_{metab}, E_{metab})$ and the pyMBMS GWAS network by $N_{pymbms} = (U_{pymbms}, V_{pymbms}, E_{pymbms})$. We construct the *guilt by association* subnetworks of genes connected to lignin-related genes/phenotypes as follows:

$N^L_{coex}$ is the subnetwork of $N_{coex}$ including the lignin related genes $l \in L_G$ and their direct neighbors:

$$\begin{aligned} N^L_{coex} &= (V^L_{coex}, E^L_{coex}) \text{ where} && (3)\\ V^L_{coex} &= \{g | g \in (L_G \cap V_{coex})\} \cup \{g | (g \in V_{coex}) \wedge (\exists l \in L_G | \{l, g\} \in E_{coex})\} && (4)\\ E^L_{coex} &= \{e = \{i, j\} \in E_{coex} | i \in V^L_{coex} \wedge j \in V^L_{coex}\} && (5) \end{aligned}$$

$N^L_{cometh}$ is the subnetwork of $N_{cometh}$ including the lignin related genes $l \in L_G$ and their direct neighbors:

$$\begin{aligned} N^L_{cometh} &= (V^L_{cometh}, E^L_{cometh}) \text{ where} && (6)\\ V^L_{cometh} &= \{g | g \in (L_G \cap V_{cometh})\} \cup \{g | (g \in V_{cometh}) \wedge (\exists l \in L_G | \{l, g\} \in E_{cometh})\} && (7)\\ E^L_{cometh} &= \{e = \{i, j\} \in E_{cometh} | i \in V^L_{cometh} \wedge j \in V^L_{cometh}\} && (8) \end{aligned}$$

$N^L_{snp}$ is the subnetwork of $N_{snp}$ including the lignin related genes $l \in L_G$ and their direct neighbors:

$$\begin{aligned} N^L_{snp} &= (V^L_{snp}, E^L_{snp}) \text{ where} && (9)\\ V^L_{snp} &= \{g | g \in (L_G \cap V_{snp})\} \cup \{g | \left(g \in V_{snp}\right) \wedge \left(\exists l \in L_G | \{l, g\} \in E_{snp}\right)\} && (10)\\ E^L_{snp} &= \{e = \{i, j\} \in E_{snp} | i \in V^L_{snp} \wedge j \in V^L_{snp}\} && (11) \end{aligned}$$

$N^L_{metab}$ is the subnetwork of $N_{metab}$ including the lignin related metabolites $m \in L_M$ and their direct neighboring genes:

$$\begin{aligned} N^L_{metab} &= (U^L_{metab}, V^L_{metab}, E^L_{metab}) \text{ where} && (12)\\ U^L_{metab} &= \{m | m \in (L_M \cap U_{metab})\} && (13)\\ V^L_{metab} &= \{g | (g \in V_{metab}) \wedge (\exists m \in L_M | (m, g) \in E_{metab})\} && (14)\\ E^L_{metab} &= \{e = (i, j) \in E_{metab} | i \in U^L_{metab} \wedge j \in V^L_{metab}\} && (15) \end{aligned}$$

$N^L_{pymbms}$ is the subnetwork of $N_{pymbms}$ including the lignin related pyMBMS peaks $p \in L_P$ and their direct neighboring genes:

$$\begin{aligned} N^L_{pymbms} &= (U^L_{pymbms}, V^L_{pymbms}, E^L_{pymbms}) \text{ where} && (16)\\ U^L_{pymbms} &= \{p | p \in (L_P \cap U_{pymbms})\} && (17)\\ V^L_{pymbms} &= \{g | \left(g \in V_{pymbms}\right) \wedge \left(\exists p \in L_P | (p, g) \in E_{pymbms}\right)\} && (18)\\ E^L_{pymbms} &= \{e = (i, j) \in E_{pymbms} | i \in U^L_{pymbms} \wedge j \in V^L_{pymbms}\} && (19) \end{aligned}$$

#### 3.7.4 Calculating LOE Scores

For a given gene $g$, the *degree* of that gene $D(g)$ indicates the number of connections that the gene has in a given network. Let $D_{coex}(g)$, $D_{cometh}(g)$, $D_{snp}(g)$ , $D_{metab}(g)$ , $D_{pymbms}(g)$ represent the degrees of gene $g$ in the lignin subnetworks $N^L_{coex}$, $N^L_{cometh}$, $N^L_{snp}$, $N^L_{metab}$ and $N^L_{pymbms}$, respectively. The LOE *breadth* score $LOE_{breadth}(g)$ is then defined as

$$LOE_{breadth}(g) = \text{bin}\left(D_{coex}(g)\right) + \text{bin}\left(D_{cometh}(g)\right) + \text{bin}\left(D_{snp}(g)\right) + \text{bin}\left(D_{metab}(g)\right) + \text{bin}\left(D_{pymbms}(g)\right) \tag{20}$$

where

$$\text{bin}(x) = \begin{cases} 1 \text{ if } x \geq 1 \\ 0 \text{ otherwise} \end{cases} \tag{21}$$

The $LOE_{breadth}(g)$ score indicates the number of different types of lines of evidence that exist linking gene $g$ to lignin-related genes/phenotypes.

The LOE *depth* score $LOE_{depth}(g)$ represents the total number of lines of evidence exist linking gene $g$ to lignin-related genes/phenotypes, and is defined as

$$LOE_{depth}(g) = D_{coex}(g) + D_{cometh}(g) + D_{snp}(g) + D_{metab}(g) + D_{pymbms}(g) \tag{22}$$

The GWAS LOE score $LOE_{gwas}(g)$ indicates the number of lignin-related phenotypes (metabolomic or pyMBMS) that a gene is connected to, and is defined as:

$$LOE_{gwas}(g) = D_{metab}(g) + D_{pymbms}(g) \tag{23}$$

Distributions of the LOE scores can be seen in Supplementary Figure S2. Cytoscape version 3.4.0 (Shannon et al., 2003) was used for network visualization.

## 3.8. Packages Used

Networks were visualized using Cytoscape version 3.4.0 (Shannon et al., 2003). Expression, methylation, SNP correlation and GWAS diagrams were created using R (R Core Team, 2017) and various R libraries (de Vries and Ripley, 2016; Auguie, 2017; Wickham, 2007; Arnold, 2017; Wickham, 2009). Data parsing, wrappers and LOE score calculation was performed using Perl. Diagrams were edited to overlay certain text using Microsoft PowerPoint.

# 4. Results and Discussion

## 4.1. Layered Networks, LOE Scores and New Potential Targets

This study involved the construction of a set of networks providing different layers of information about the relationships between genes, and between genes and phenotypes, and the development of a Lines Of Evidence scoring system (LOE scores) which integrate the information in the different network layers and quantify the number of lines of evidence connecting genes to lignin-related genes/phenotypes. The GWAS network layers provide information as to which genes are potentially involved in certain functions because they contain genomic variants significantly associated with measured phenotypes. The co-methylation and co-expression networks provide information on different layers of regulatory mechanisms within the cell. The SNP correlation network provides information about possible co-evolution relationships between genes, through correlated variants across a population.

Marking known genes and phenotypes involved in lignin biosynthesis in these networks allowed for the calculation of a set of LOE (Lines Of Evidence) scores for each gene, indicating the strength of the evidence linking each gene to lignin-related functions. The breadth LOE score indicates the number of types of lines of evidence (number of layers) which connect the gene to lignin-related genes/phenotypes, whereas the depth LOE score indicates the total number of lignin-related genes/phenotypes the gene is associated with. Individual layer LOE scores (e.g. co-expression LOE score or GWAS LOE score) indicate the number of lignin-related associations the gene has within that layer.

To select the top set of potential new candidate genes involved in lignin biosynthesis, genes which showed a number of different lines of evidence connecting them to lignin-related functions were identified by selecting genes with a LOE breadth score $>= 3$. Since the GWAS networks provide the highest resolution, most direct connections to lignin-related functions, it was also required that our potential new targets had a GWAS score $>= 1$. This provides a set of 375 new candidate genes potentially involved in lignin biosynthesis, identified through multiple lines of evidence (Supplementary Table S4). This set of Potential New Target genes will be referred to as set of PNTs. A selection of these potential new candidates below and their annotations, derived from their *Arabidopsis* best hits, will be discussed below.

## 4.2. Agamous-like Genes

Genes in the AGAMOUS-LIKE gene family are MADS-box transcription factors, many of which which have been found to play important roles in floral development (Yoo et al., 2006; Fernandez et al., 2014; Yu et al., 2017, 2004, 2002; Lee et al., 2000). Three potential AGAMOUS-LIKE (AGL) genes are found in the set of PNTs, in particular, a homolog of *Arabidopsis* AGL8 (AT5G60910, also known as FRUITFUL), a homolog of *Arabidopsis* AGL12 (AT1G71692), and a homolog of *Arabidopsis* AGL24 (AT4G24540) and AGL22 (AT2G22540).

The first potential AGL gene in our set of PNTs is Potri.012G062300, with a breadth score of 3 and a GWAS score of 2 (Figure 3A), whose best *Arabidopsis thaliana* hit is AGL8 (AT5G60910). It has GWAS associations with a lignin-related metabolite (quinic acid) and a lignin pyMBMS peak (syringol) (Figure 3C, Table 1) and is co-methylated with three lignin-related genes (Figure 3B, Table 3). There is thus strong evidence for the involvement of *P. trichocarpa* AGL8 in the regulation of lignin-related functions. There is literature evidence that supports the hypothesis of AGL8's involvement in the regulation of lignin biosynthesis. A patent exists for the use of AGL8 expression in reducing the lignin content of plants (Yanofsky et al., 2004). The role of AGL8 (FUL) was described in Ferrándiz et al. (2000), in which they investigated the differences in lignin deposition in transgenic plants in which AGL8 is constitutively expressed, loss-of-function AGL8 mutants and wild-type *Arabidopsis* plants (Ferrándiz et al., 2000). In wild-type plants, a single layer of valve cells were lignified. In loss-of-function AGL8 mutants, all valve mesophyl cell layers were lignified, while in the transgenic plants, constitutive expression of AGL8 resulted in loss of lignified cells (Ferrándiz et al., 2000). This study thus showed the involvement of AGL8 in fruit lignification during fruit development.

There is evidence of other AGAMOUS-LIKE genes affecting lignin content. A study by Gimenez *et al.* (2010) investigated TALG1, an AGAMOUS-LIKE gene in tomato, and found that TAGL1 RNAi-silenced fruits showed increased lignin content, and increased expression levels of lignin biosynthesis genes (Giménez et al., 2010). A recent study by Cosio *et al.* (2017) showed that AGL15 in *Arabidopsis* is also involved in regulating lignin-related functions, in that AGL15 binds to the promotor of peroxidase PRX17, and regulates its expression (Cosio et al., 2017). In addition, PRX17 loss of function mutants had reduced lignin content (Cosio et al., 2017).

There is thus compelling evidence that various AGAMOUS-LIKE genes are involved in regulating lignin biosynthesis/deposition in plants. Two other AGAMOUS-like genes are seen in the set of PNTs, namely a homolog of *Arabidopsis* AGL12 (Potri.013G102600) and a homolog of *Arabidopsis* AGL22/AGL24 (Potri.007G115100). Potri.013G102600 (AGL12) has GWAS associations with three lignin-related metabolites, namely hydroxyphenyl lignan glycoside, coumaroyl-tremuloidin and 3-O-caffeoyl-quinate (Figure 4A, Figure 4B, Table 1). It is co-expressed with four lignin-related genes including two caffeoyl coenzyme A O-methyltransferases, a caffeate O-methyltransferase and a ferulic acid 5-hydroxylase (Figure 4A, Figure 4C, Table 2) and it is co-methylated with four other lignin-related genes (Figure 4A, Figure 4D, Table 3). Potri.007G115100 (AGL22/AGL24) has GWAS associations with the syringaldehyde pyMBMS phenotype and a caffeoyl conjugate metabolite (Figure 5A, Figure 5B, Table 1). It also has SNP correlations with a laccase and a nicotinamidase (Figure 5A, Figure 5C, Figure 5D, Table 4, Supplementary Table S5). The combination of the multiple lines of multi-omic evidence thus suggest the involvement of *P. trichocarpa* homologs of *A. thaliana* AGL22/AGL24 and AGL12 in regulating lignin biosynthesis.

## 4.3. MYB Transcription Factors

MYB proteins contain the conserved MYB DNA-binding domain, and usually function as transcription factors. R2R3-MYBs have been found to regulate various functions, including flavonol biosynthesis, anthocyanin biosynthesis, lignin biosynthesis, cell fate and developmental functions (Dubos et al., 2010). The set of PNTs contains several genes which are homologs of *Arabidopsis* MYB transcription factors, including homologs of *Arabidopsis* MYB66/MYB3, MYB46, MYB36 and MYB111.

There is already existing literature evidence for how some of these MYBs affect lignin biosynthesis. Liu et al. (2015) reviews the involvement of MYB transcription factors in the regulation of phenylpropanoid metabolism. MYB3 in *Arabidopsis* is known to repress phenylpropanoid biosynthesis (Zhou et al., 2017a), and a *P. trichocarpa* homolog of MYB3 is found in our set of potential new targets. Another potential new target is the *P. trichocarpa* homolog of *Arabidopsis* MYB36 (Potri.006G170800) which is connected to lignin-related functions through multiple lines of evidence (Figure 6). In *Arabidopsis*, MYB36 has been found to regulate the local deposition of lignin during casparian strip formation, and *myb36* mutants exhibit incorrectly localized lignin deposition (Kamiya et al., 2015).

MYB46 is known to be a regulator of secondary cell wall formation (Zhong et al., 2007). Overexpression of MYB46 in *Arabidopsis* activates lignin, cellulose and xylan biosynthesis pathways (Zhong et al., 2007). The MYB46 homolog in *P. trichocarpa*, Potri.009G053900, is connected to lignin-related functions through multiple lines of evidence (Figure 7A), including a GWAS association with a hydroxyphenyl lignan glycoside (Figure 7E, Table 1), co-expression with pinoresinol reductase 1 and caffeate O-methyltransferase 1 (Figure 7F, Table 2) and co-methylation with dehydroquinate-shikimate dehydrogenase enzyme, cinnamyl alcohol dehydrogenase 9, 4-coumarate-CoA ligase activity/4CL) and caffeoyl-CoA 3-O-methyltransferase (Figure 7G, Table 3).

A MYB transcription factor in the set of PNTs which has, to our knowledge, not yet been directly associated with lignin biosynthesis is MYB111 (Figure 7A-D). However, with existing literature evidence, one can hypothesize that MYB111 can alter lignin content by redirecting carbon flux from flavonoids to monolignols. There is evidence that MYB111 is involved in crosstalk between lignin and flavonoid pathways. Monolignols and flavonoids are both derived from phenylalanine through the phenylpropanoid pathway (Liu et al., 2015). There is crosstalk between the signalling pathways of ultraviolet-B (UV-B) stress and biotic stress pathways (Schenke et al., 2011). In the study by Schenke et al. (2011), it was shown that under UV-B light stress, *Arabidopsis* plants produce flavonols as a UV protectant. Also, simultaniously applying the bacterial elicitor flg22, which simulates biotic stress, repressed flavonol biosynthesis genes and induced production of defense compounds including camalexin and scopoletin, as well as lignin, which provides a physical barrier preventing pathogens' entry (Schenke et al., 2011). This crosstalk involved regulation by MYB12 and MYB4 (Schenke et al., 2011). This study by Schenke et al. (2011) was performed using cell cultures. A second study (Zhou et al., 2017b) used *Arabidopsis* seedlings, and found that MYB111 may be involved in the crosstalk in planta (Zhou et al., 2017b). The multiple lines of evidence connecting the *P. trichocarpa* homolog of *Arabidopsis* MYB111 (Potri.010G141000) to lignin related functions, in combination with the above literature evidence suggests the involvement this gene in the regulation of lignin biosynthesis by redirecting carbon flux from flavonol biosynthesis to monolignol biosynthesis, as part of the crosstalk between UV-B protection and biotic stress signalling pathways.

## 4.4. Chloroplast Signal Recognition Particle

Potri.016G078600, a homolog of the *Arabidopsis* chloropast signal recognition particle cpSRP54 occurs in the set of PNTs (Figure 8). It has a GWAS LOE score of 3, through GWAS associations with salicyl-coumaroyl-glucoside, a caffeoyl conjugate and a feruloyl conjugate (Figure 8B, Table 1, Supplementary Table S4). It also has a breadth score of 4, indicating that it is linked to lignin-related genes/phenotypes though 4 different types of associations (Figure 8). CpSRP54 gene has been found to regulate carotenoid accumulation in *Arabidopsis* (Yu et al., 2012). CpSRP54 and cpSRP43 form a "transit complex" along with a light-harvesting chlorophyll a/b-binding protein (LHCP) family member to transport it to the thylakoid membrane (Groves et al., 2001; Schünemann, 2004). A study in *Arabidopsis* found that cpSRP43 mutants had reduced lignin content (Klenell et al., 2005). Since CpSRP54 regulates carotenoid accumulation, and cpSRP43 appears to affect lignin content, it is possible that chloroplast signal recognition particles affect lignin and carotenoid content through flux through the phenylpropanoid pathway, the common origin of both of these compounds. In fact, a gene mutation *cue1* which causes LHCP underexpression also results in reduced aromatic amino acid biosynthesis (Streatfield et al., 1999). These multiple lines of evidence, combined with the above cited literature suggests that chloroplast signal recognition particles in *P. trichocarpa* could potentially influence lignin content.

## 4.5. Concluding Remarks

This study made use of high-resolution GWAS data, combined with co-expression, co-methylation and SNP correlation networks in a multi-omic, data layering approach which has allowed the identification of new potential target genes involved in lignin biosynthesis/regulation. Various literature evidence supports the involvement of many of these new target genes in lignin biosynthesis/regulation, and these are suggested for future validation for involvement in the regulation of lignin biosynthesis. The data layering technique and LOE scoring system developed can be applied to other omic data types to assist in the generation of new hypotheses surrounding various functions of interest.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author Contributions

DW calculated methylation TPM values, constructed the networks, developed the scoring technique, performed the data layering and scoring analysis and interpreted the results, PJ performed the outlier analysis and GWAS, MS mapped gene expression atlas reads and calculated gene expression TPM values, SD, GT and WM lead the effort on constructing the GWAS population, TJT led the leaf sample collection for GCMS-based metabolomic analyses, identified the peaks, and summarized the metabolomics data, MZM collected the leaf samples and manually extracted the metabolite data, NZ conducted leaf sample preparation, extracted and derivatized and analyzed the metabolites by GCMS, PR aided in peak extraction, JS and AS generated the gene expression atlas data, SD and DMS generated the SNP calls, RS generated the pyMBMS data, DJ conceived of and supervised the project, generated MapMan annotations and edited the manuscript, DW, PJ, SD, DMS, RS, TJT, JS and AS wrote the manuscript.

## Funding

*Funding provided by The BioEnergy Science (BESC) and The Center for Bioenergy Innovation (CBI). U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science.*

*This research was also supported by the Department of Energy Laboratory Directed Research and Development funding (7758), at the Oak Ridge National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US DOE under contract DE-AC05-00OR22725.*

*This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) and the Compute and Data Environment for Science (CADES) at the 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.*

*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 Bioenergy Science Center (BESC) 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.*

*The JGI Plant Gene Atlas project conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Full Gene Atlas data sets are available at http://phytozome.jgi.doe.gov.*

## Acknowledgments

The authors would like to acknowledge Nancy Engle, David Weston, Ryan Aug, KC Cushman, Lee Gunter and Sara Jawdy for the metabolomics sample collection, Carissa Bleker for working on the GWAS and outlier analysis, Mark Davis for the pyMBMS data and the Department of Energy Joint Genome Institute (JGI) for sequencing.

## References

Ruslan Akulenko and Volkhard Helms. DNA co-methylation analysis suggests novel functional associations between gene pairs in breast cancer samples. *Human molecular genetics*, page ddt158, 2013.

Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, page btu638, 2014.

Jeffrey B. Arnold. *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*, 2017. URL `https://CRAN.R-project.org/package=ggthemes`. R package version 3.4.0.

Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, Midori A Harris, David P Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C Matese, Joel E Richardson, Martin Ringwald, Gerald M Rubin, and Gavin Sherlock. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25–29, 2000.

Baptiste Auguie. *gridExtra: Miscellaneous Functions for "Grid" Graphics*, 2017. URL `https://CRAN.R-project.org/package=gridExtra`. R package version 2.3.

Derek W Barnett, Erik K Garrison, Aaron R Quinlan, Michael P Strömberg, and Gabor T Marth. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12):1691–1692, 2011.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

Supinda Bunyavanich, Eric E Schadt, Blanca E Himes, Jessica Lasky-Su, Weiliang Qiu, Ross Lazarus, John P Ziniti, Ariella Cohain, Michael Linderman, Dara G Torgerson, et al. Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC medical genomics*, 7(1):48, 2014.

Robert Busch, Weiliang Qiu, Jessica Lasky-Su, Jarrett Morrow, Gerard Criner, and Dawn DeMeo. Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. *Respiratory Research*, 17(1):143, 2016.

Gina M Calabrese, Larry D Mesner, Joseph P Stains, Steven M Tommasini, Mark C Horowitz, Clifford J Rosen, and Charles R Farber. Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. *Cell systems*, 4(1):46–59, 2017.

Sharlee Climer, Alan R Templeton, and Weixiong Zhang. Allele-Specific Network Reveals Combinatorial Interaction that Transcends Small Effects in Psoriasis GWAS. *PLoS Comput Biol*, 10(9):e1003766, 2014a.

Sharlee Climer, Wei Yang, Lisa Fuentes, Victor G Dávila-Román, and C Charles Gu. A Custom Correlation Coefficient (CCC) Approach for Fast Identification of Multi-SNP Association Patterns in Genome-Wide SNPs Data. *Genetic Epidemiology*, 38(7):610–621, 2014b.

Claudia Cosio, Philippe Ranocha, Edith Francoz, Vincent Burlat, Yumei Zheng, Sharyn E Perry, Juan-Jose Ripoll, Martin Yanofsky, and Christophe Dunand. The class III peroxidase PRX17 is a direct target of the MADS-box transcription factor AGAMOUS-LIKE15 (AGL15) and participates in lignified tissue formation. *New Phytologist*, 213(1):250–263, 2017.

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A Albers, Eric Banks, Mark A DePristo, Robert E Handsaker, Gerton Lunter, Gabor T Marth, Stephen T Sherry, et al. The variant call format and VCFtools. *Bioinformatics*, 27(15): 2156–2158, 2011.

Matthew N Davies, Manuela Volta, Ruth Pidsley, Katie Lunnon, Abhishek Dixit, Simon Lovestone, Cristian Coarfa, R Alan Harris, Aleksandar Milosavljevic, Claire Troakes, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome biology*, 13(6):R43, 2012.

Mark F Davis, Gerald A Tuskan, Peggy Payne, Timothy J Tschaplinski, and Richard Meilan. Assessment of *Populus* wood chemistry following the introduction of a Bt toxin gene. *Tree physiology*, 26(5):557–564, 2006.

Andrie de Vries and Brian D. Ripley. *ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'*, 2016. URL `https://CRAN.R-project.org/package=ggdendro`. R package version 0.1-20.

Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

Christian Dubos, Ralf Stracke, Erich Grotewold, Bernd Weisshaar, Cathie Martin, and Loïc Lepiniec. MYB transcription factors in *Arabidopsis*. *Trends in plant science*, 15(10):573–581, 2010.

Luke M Evans, Gancho T Slavov, Eli Rodgers-Melnick, Joel Martin, Priya Ranjan, Wellington Muchero, Amy M Brunner, Wendy Schackwitz, Lee Gunter, Jin-Gui Chen, et al. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics*, 46(10):1089–1096, 2014.

Robert J Evans and Thomas A Milne. Molecular Characterization of the Pyrolysis of Biomass. *Energy & Fuels*, 1(2): 123–137, 1987.

Donna E Fernandez, Chieh-Ting Wang, Yumei Zheng, Benjamin J Adamczyk, Rajneesh Singhal, Pamela K Hall, and Sharyn E Perry. The MADS-Domain Factors AGAMOUS-LIKE15 and AGAMOUS-LIKE18, along with SHORT VEGETATIVE PHASE and AGAMOUS-LIKE24, Are Necessary to Block Floral Gene Expression during the Vegetative Phase. *Plant physiology*, 165(4):1591–1603, 2014.

Cristina Ferrándiz, Sarah J Liljegren, and Martin F Yanofsky. Negative regulation of the *SHATTERPROOF* genes by FRUITFULL during *Arabidopsis* fruit development. *Science*, 289(5478):436–438, 2000.

Robert D. Finn, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Jaina Mistry, Alex L. Mitchell, Simon C. Potter, Marco Punta, Matloob Qureshi, Amaia Sangrador-Vegas, Gustavo A. Salazar, John Tate, and Alex Bateman. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44(D1):D279–D285, 2016. doi: 10.1093/nar/gkv1344. URL `+http://dx.doi.org/10.1093/nar/gkv1344`.

Brooke L Fridley, Steven Lund, Gregory D Jenkins, and Liewei Wang. A Bayesian integrative genomic model for pathway analysis of complex traits. *Genetic epidemiology*, 36(4):352–359, 2012.

Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research*, 45(D1):D331–D338, 2017.

Armando Geraldes, SP Difazio, GT Slavov, P Ranjan, W Muchero, J Hannemann, LE Gunter, AM Wymore, CJ Grassa, N Farzaneh, et al. A 34K SNP genotyping array for *Populus trichocarpa*: Design, application to the study of natural populations and transferability to other *Populus* species. *Molecular Ecology Resources*, 13(2):306–323, 2013.

Estela Giménez, Benito Pineda, Juan Capel, María Teresa Antón, Alejandro Atarés, Fernando Pérez-Martín, Begoña García-Sogo, Trinidad Angosto, Vicente Moreno, and Rafael Lozano. Functional Analysis of the *Arlequin* Mutant Corroborates the Essential Role of the *Arlequin/TAGL1* Gene during Reproductive Development of Tomato. *PLoS One*, 5(12):e14427, 2010.

David M Goodstein, Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D Hayes, Joni Fazo, Therese Mitros, William Dirks, Uffe Hellsten, Nicholas Putnam, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1):D1178–D1186, 2012.

Matthew R Groves, Alexandra Mant, Audrey Kuhn, Joachim Koch, Stefan Dübel, Colin Robinson, and Irmgard Sinning. Functional Characterization of Recombinant Chloroplast Signal Recognition Particle. *Journal of Biological Chemistry*, 276(30):27778–27786, 2001.

Hongshan Jiang, Rong Lei, Shou-Wei Ding, and Shuifang Zhu. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics*, 15(1):182, 2014.

Takehiro Kamiya, Monica Borghi, Peng Wang, John MC Danku, Lothar Kalmbach, Prashant S Hosmani, Sadaf Naseer, Toru Fujiwara, Niko Geldner, and David E Salt. The MYB36 transcription factor orchestrates Casparian strip formation. *Proceedings of the National Academy of Sciences*, 112(33):10533–10538, 2015.

Hyun Min Kang, Jae Hoon Sul, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara Sabatti, Eleazar Eskin, et al. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, 42 (4):348–354, 2010.

Dokyoon Kim, Hyunjung Shin, Young Soo Song, and Ju Han Kim. Synergistic effect of different levels of genomic data for cancer clinical outcome prediction. *Journal of biomedical informatics*, 45(6):1191–1198, 2012.

Dokyoon Kim, Ruowang Li, Scott M Dudek, and Marylyn D Ritchie. ATHENA: Identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. *BioData mining*, 6(1):23, 2013.

Markus Klenell, Shigeto Morita, Mercedes Tiemblo-Olmo, Per Mühlenbock, Stanislaw Karpinski, and Barbara Karpinska. Involvement of the Chloroplast Signal Recognition Particle cpSRP43 in Acclimation to Conditions Promoting Photooxidative Stress in *Arabidopsis*. *Plant and cell physiology*, 46(1):118–129, 2005.

Horim Lee, Sung-Suk Suh, Eunsook Park, Euna Cho, Ji Hoon Ahn, Sang-Gu Kim, Jong Seob Lee, Young Myung Kwon, and Ilha Lee. The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes & Development*, 14(18):2366–2376, 2000.

Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.

Yongchao Li, Timothy J Tschaplinski, Nancy L Engle, Choo Y Hamilton, Miguel Rodriguez, James C Liao, Christopher W Schadt, Adam M Guss, Yunfeng Yang, and David E Graham. Combined inactivation of the *Clostridium cellulolyticum* lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations. *Biotechnology for biofuels*, 5(1):2, 2012.

Zhiwu Li and Harold N Trick. Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *Biotechniques*, 38(6):872, 2005.

Jingying Liu, Anne Osbourn, and Pengda Ma. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Molecular plant*, 8(5):689–708, 2015.

Marc Lohse, Axel Nagel, Thomas Herter, Patrick May, Michael Schroda, Rita Zrenner, Takayuki Tohge, Alisdair R Fernie, Mark Stitt, and Björn Usadel. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, cell & environment*, 37(5):1250–1258, 2014.

Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9):1297–1303, 2010.

Athena D McKown, Jaroslav Klápště, Robert D Guy, Armando Geraldes, Ilga Porth, Jan Hannemann, Michael Friedmann, Wellington Muchero, Gerald A Tuskan, Jürgen Ehlting, et al. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist*, 203(2):535–553, 2014.

Wellington Muchero, Jianjun Guo, Stephen P DiFazio, Jin-Gui Chen, Priya Ranjan, Gancho T Slavov, Lee E Gunter, Sara Jawdy, Anthony C Bryan, Robert Sykes, et al. High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. *BMC genomics*, 16(1):24, 2015.

Adrian M. Price-Whelan and Daniel Foreman-Mackey. schwimmbad: A uniform interface to parallel processing pools in Python. *The Journal of Open Source Software*, 2(17), sep 2017. doi: 10.21105/joss.00357. URL `https://doi.org/10.21105/joss.00357`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

Arthur J Ragauskas, Charlotte K Williams, Brian H Davison, George Britovsek, John Cairney, Charles A Eckert, William J Frederick, Jason P Hallett, David J Leak, Charles L Liotta, et al. The Path Forward for Biofuels and Biomaterials. *science*, 311(5760):484–489, 2006.

Marylyn D Ritchie, Emily R Holzinger, Ruowang Li, Sarah A Pendergrass, and Dokyoon Kim. Methods of integrating data to uncover genotype-phenotype interactions. *Nature reviews. Genetics*, 16(2):85, 2015.

Poulomi Sannigrahi, Arthur J Ragauskas, and Gerald A Tuskan. Poplar as a feedstock for biofuels: A review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, 4(2):209–226, 2010.

Dirk Schenke, Christoph Boettcher, and Dierk Scheel. Crosstalk between abiotic ultraviolet-B stress and biotic (flg22) stress signalling in *Arabidopsis* prevents flavonol accumulation in favor of pathogen defence compound production. *Plant, cell & environment*, 34(11):1849–1864, 2011.

Danja Schünemann. Structure and function of the chloroplast signal recognition particle. *Current genetics*, 44(6): 295–304, 2004.

Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, 2003.

Gancho T Slavov, Stephen P DiFazio, Joel Martin, Wendy Schackwitz, Wellington Muchero, Eli Rodgers-Melnick, Mindie F Lipphardt, Christa P Pennacchio, Uffe Hellsten, Len A Pennacchio, et al. Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree *Populus trichocarpa*. *New Phytologist*, 196(3):713–725, 2012.

Stephen J Streatfield, Andreas Weber, Elizabeth A Kinsman, Rainer E Häusler, Jianming Li, Dusty Post-Beittenmiller, Werner M Kaiser, Kevin A Pyke, Ulf-Ingo Flügge, and Joanne Chory. The Phosphoenolpyruvate/Phosphate Translocator Is Required for Phenolic Metabolism, Palisade Cell Development, and Plastid-Dependent Nuclear Gene Expression. *The Plant Cell*, 11(9):1609–1621, 1999.

Robert Sykes, Matthew Yung, Evandro Novaes, Matias Kirst, Gary Peter, and Mark Davis. High-Throughput Screening of Plant Cell-Wall Composition Using Pyrolysis Molecular Beam Mass Spectroscopy. *Biofuels: Methods and protocols*, pages 169–183, 2009.

Timothy J Tschaplinski, Robert F Standaert, Nancy L Engle, Madhavi Z Martin, Amandeep K Sangha, Jerry M Parks, Jeremy C Smith, Reichel Samuel, Nan Jiang, Yunqiao Pu, Arthur J Ragauskas, Choo Y Hamilton, Chunxiang Fu, Zeng-Yu Wang, Brian H Davidson, Richard A Dixon, and Jonathan R Mielenz. Down-regulation of the caffeic acid *O*-methyltransferase gene in switchgrass reveals a novel monolignol analog. *Biotechnology for Biofuels*, 5(1):1, 2012.

Gerald Tuskan, Darrell West, Harvey D Bradshaw, David Neale, Mitch Sewell, Nick Wheeler, Bob Megraw, Keith Jech, Art Wiselogel, Robert Evans, et al. Two High-Throughput Techniques for Determining Wood Properties as Part of a Molecular Genetics Analysis of Hybrid Poplar and Loblolly Pine. In *Twentieth Symposium on Biotechnology for Fuels and Chemicals*, pages 55–65. Springer, 1999.

Gerald Tuskan, Gancho Slavov, Steve DiFazio, Wellington Muchero, Ranjan Pryia, Wendy Schackwitz, Joel Martin, Daniel Rokhsar, Robert Sykes, Mark Davis, et al. *Populus* resequencing: towards genome-wide association studies. In *BMC Proceedings*, volume 5, page I21. BioMed Central Ltd, 2011.

Gerald A Tuskan, S Difazio, Stefan Jansson, J Bohlmann, I Grigoriev, U Hellsten, N Putnam, S Ralph, Stephane Rombauts, A Salamov, et al. The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313 (5793):1596–1604, 2006.

Geraldine A Van der Auwera, Mauricio O Carneiro, Christopher Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, Khalid Shakir, David Roazen, Joel Thibault, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current protocols in bioinformatics*, pages 11–10, 2013.

Stijn Van Dongen. Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications*, 30(1):121–141, 2008.

Stijn Marinus Van Dongen. Graph clustering by flow simulation. 2001.

Ruben Vanholme, Brecht Demedts, Kris Morreel, John Ralph, and Wout Boerjan. Lignin Biosynthesis and Structure. *Plant physiology*, 153(3):895–905, 2010.

Kelly J Vining, Kyle R Pomraning, Larry J Wilhelm, Henry D Priest, Matteo Pellegrini, Todd C Mockler, Michael Freitag, and Steven H Strauss. Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics*, 13(1):1, 2012.

Günter P Wagner, Koryu Kin, and Vincent J Lynch. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences*, 131(4):281–285, 2012.

Hadley Wickham. Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), 2007. URL `http://www.jstatsoft.org/v21/i12/paper`.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. ISBN 978-0-387-98140-6. URL `http://ggplot2.org`.

Martin F Yanofsky, Sarah Liljegren, and Cristina Ferrandiz. Selective control of lignin biosynthesis in transgenic plants, July 27 2004. US Patent 6,768,042.

Seung Kwan Yoo, Jong Seob Lee, and Ji Hoon Ahn. Overexpression of *AGAMOUS-LIKE 28* (*AGL28*) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. *Biochemical and biophysical research communications*, 348(3):929–936, 2006.

Bianyun Yu, Margaret Y Gruber, George G Khachatourians, Rong Zhou, Delwin J Epp, Dwayne D Hegedus, Isobel AP Parkin, Ralf Welsch, and Abdelali Hannoufa. Arabidopsis cpSRP54 regulates carotenoid accumulation in *Arabidopsis* and Brassica napus. *Journal of experimental botany*, 63(14):5189–5202, 2012.

Hao Yu, Yifeng Xu, Ee Ling Tan, and Prakash P Kumar. AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. *Proceedings of the National Academy of Sciences*, 99(25):16336–16341, 2002.

Hao Yu, Toshiro Ito, Frank Wellmer, and Elliot M Meyerowitz. Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. *Nature genetics*, 36(2):157, 2004.

Xiaohui Yu, Guoping Chen, Xuhu Guo, Yu Lu, Jianling Zhang, Jingtao Hu, Shibing Tian, and Zongli Hu. Silencing *SlAGL6*, a tomato *AGAMOUS-LIKE6* lineage gene, generates fused sepal and green petal. *Plant Cell Reports*, pages 1–11, 2017.

Ruiqin Zhong, Elizabeth A Richardson, and Zheng-Hua Ye. The MYB46 Transcription Factor Is a Direct Target of SND1 and Regulates Secondary Wall Biosynthesis in *Arabidopsis*. *The Plant Cell*, 19(9):2776–2792, 2007.

Meiliang Zhou, Kaixuan Zhang, Zhanmin Sun, Mingli Yan, Cheng Chen, Xinquan Zhang, Yixiong Tang, and Yanmin Wu. LNK1 and LNK2 Corepressors Interact with the MYB3 Transcription Factor in Phenylpropanoid Biosynthesis. *Plant Physiology*, 174(3):1348–1358, 2017a.

Zheng Zhou, Dirk Schenke, Ying Miao, and Daguang Cai. Investigation of the crosstalk between the flg22 and the UV-B-induced flavonol pathway in *Arabidopsis thaliana* seedlings. *Plant, cell & environment*, 40(3):453–458, 2017b.

# Figures

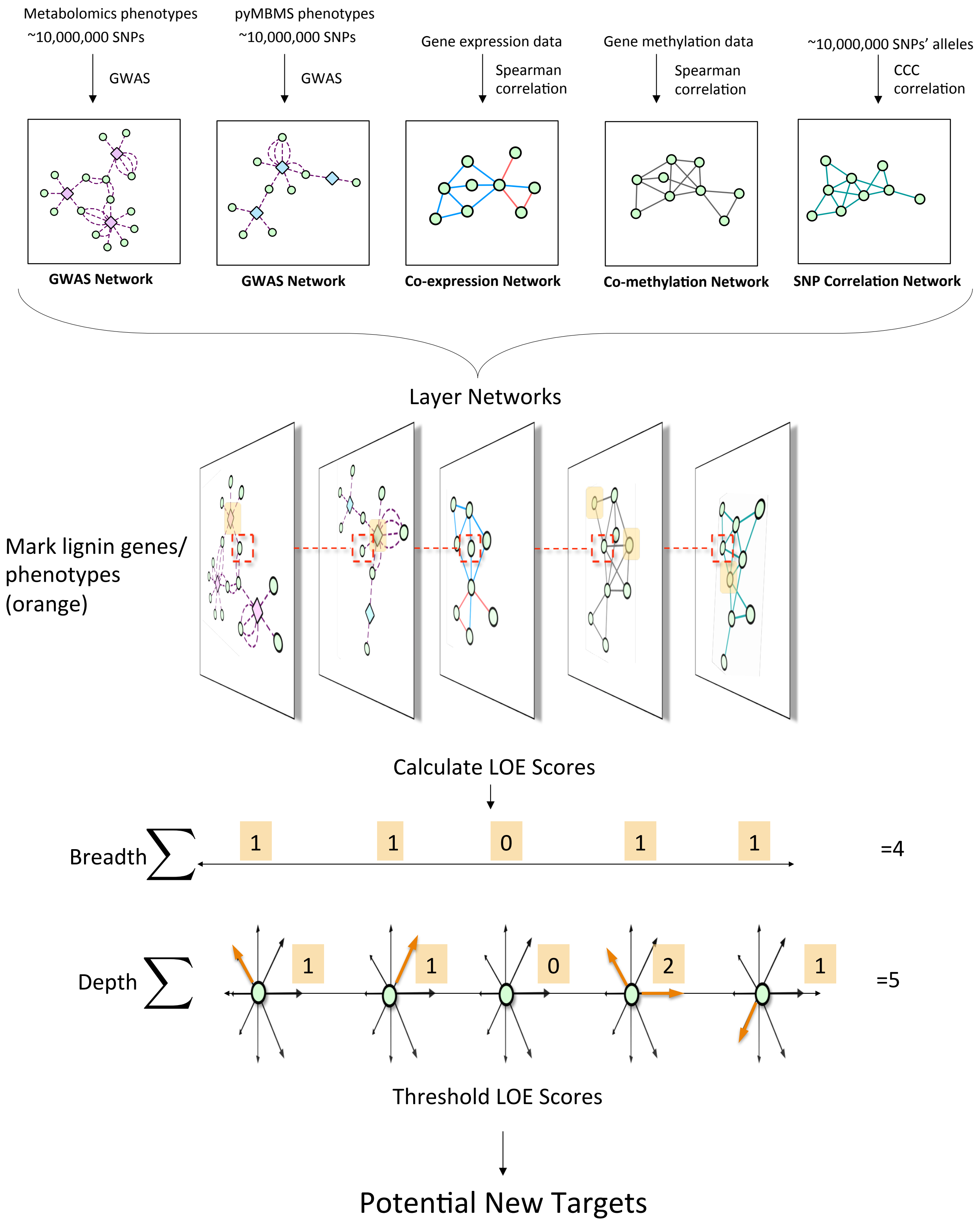

**Figure 1:** *Overview of pipeline for data layering and score calcualtion. First, the different network layers are constructed. Networks are layered, and lignin-related genes and phenotypes (orange) are identified. LOE scores are calculated for each gene. An example of the LOE score calculation for the red-boxed gene is shown. Thresholding the LOE scores results in a set of new potential target genes involved in lignin biosynthesis/degradation/regulation.*

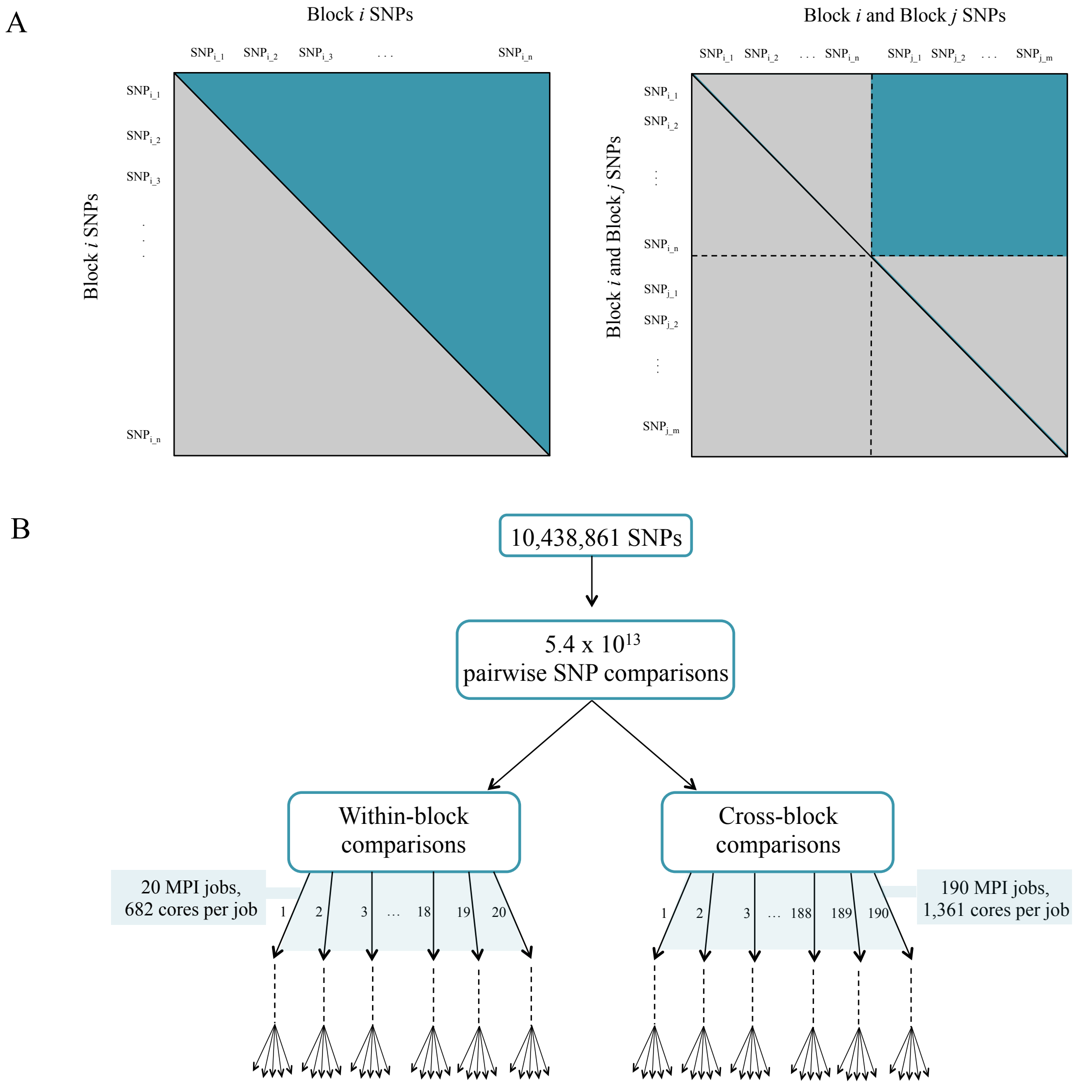

**Figure 2:** *(A) Parallelization strategy for ccc calculation between all pairs of SNPs. (B) MPI jobs for within and cross-block comparisons.*

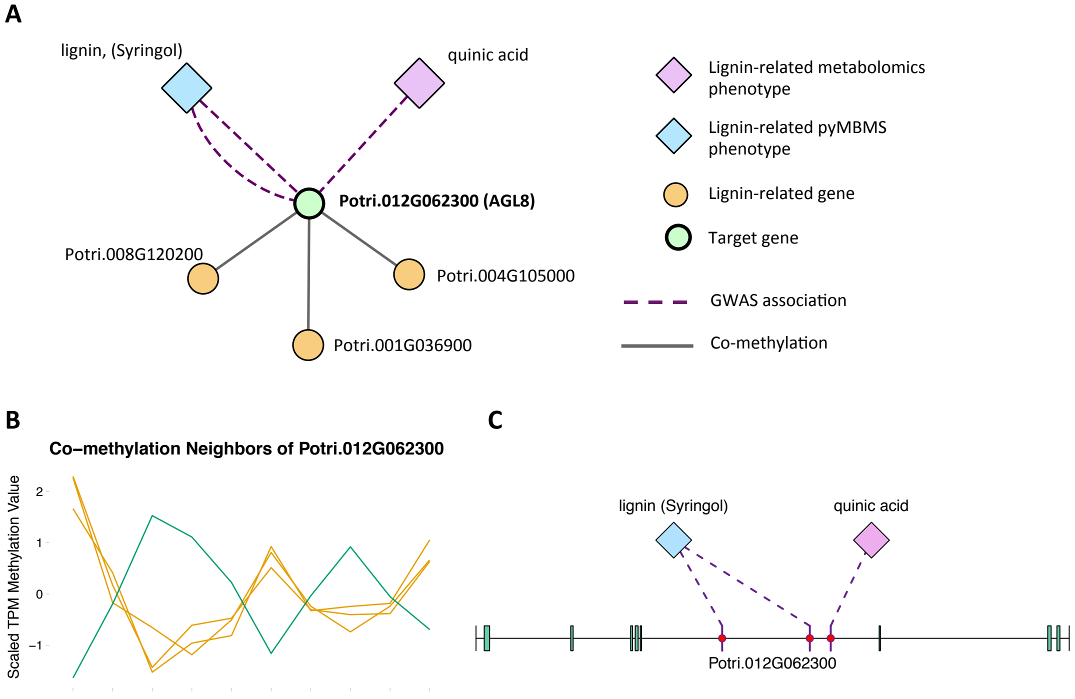

**Figure 3:** *(A) Lines of Evidence for Potri.012G062300 (homolog of Arabidopsis AGL8). (B) Co-methylation of Potri.012G062300 with three lignin-related genes (Table 3) The green line represents potential target Potri.012G062300 and yellow lines represent lignin-related genes. (C) GWAS associations of Potri.012G062300 with a lignin-related metabolite and a lignin-related pyMBMS peak (Table 1).*

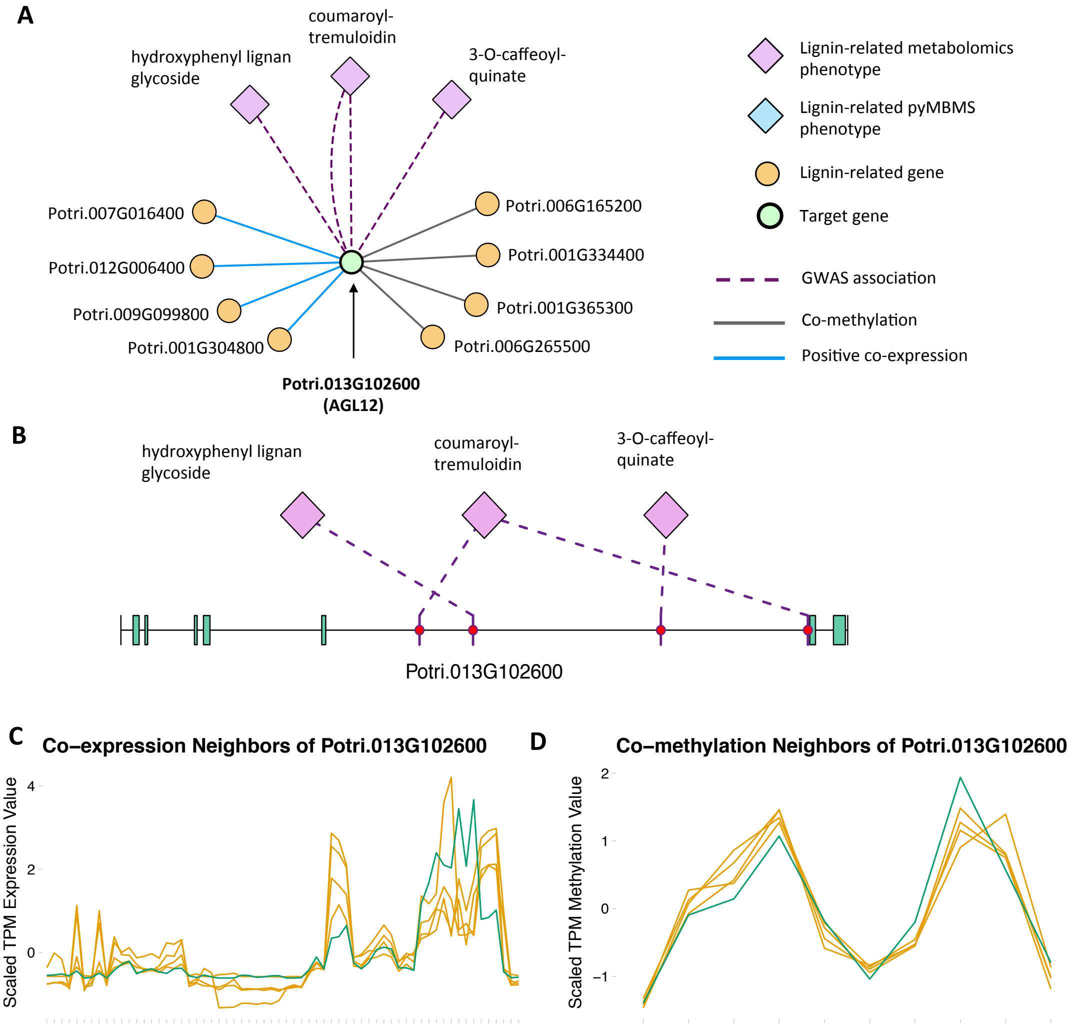

**Figure 4:** *(A) Lines of Evidence for Potri.013G102600 (homolog of Arabidopsis AGL12). (B) GWAS associations of Potri.013G102600 with three lignin-related metabolites (Table 1). (C) Co-expression of Potri.013G102600 with three lignin-related genes (Table 2). (D) Co-methylation of Potri.013G102600 with four lignin-related genes (Table 3). In line plots, the green lines represent potential target Potri.013G102600 and yellow lines represent lignin-related genes.*

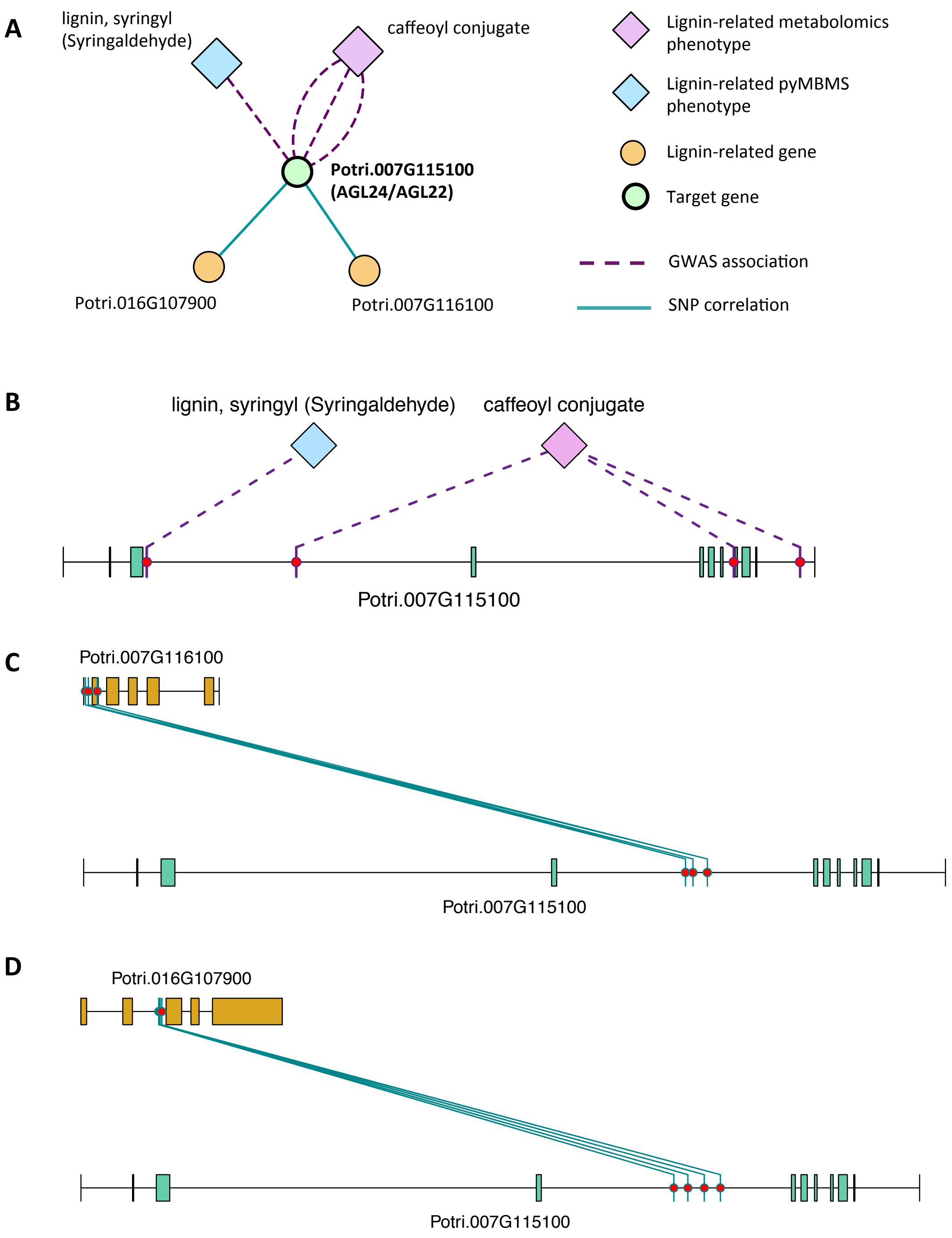

**Figure 5:** *(A) Lines of Evidence for Potri.007G115100 ( homolog of Arabidopsis AGL22/24). (B) GWAS associations of Potri.007G115100 with a lignin-related metabolite and a lignin-related pyMBMS peak (Table 1). (C,D) Correlations of SNPs in Potri.007G115100 with SNPs in two lignin-related genes (Table 4, Supplementary Table S5).*

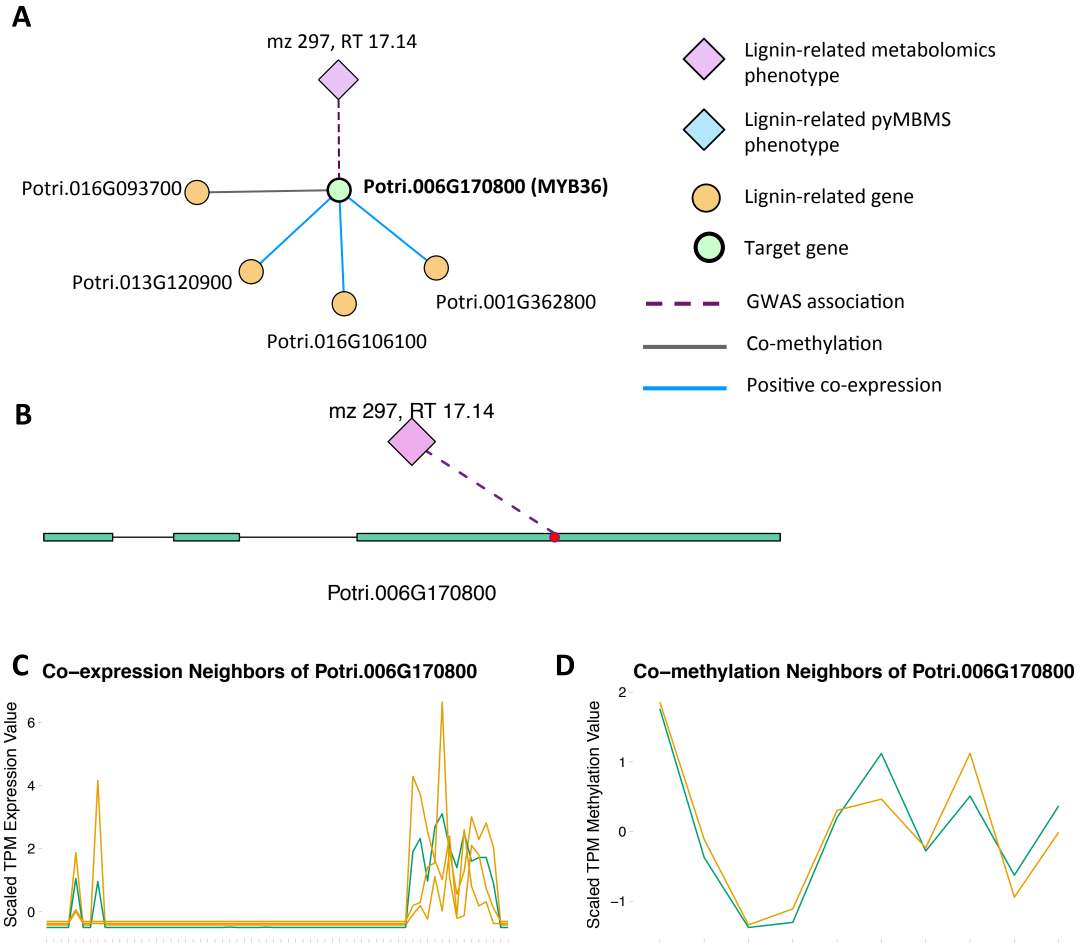

**Figure 6:** *(A) Lines of Evidence for Potri.006G170800 (homolog of Arabidopsis MYB36). (B) GWAS associations of Potri.006G170800 with a lignin-related metabolite (Table 1). (C) Co-expression of Potri.006G170800 with three lignin-related genes (Table 2). (D) Co-methylation of Potri.006G170800 with a lignin-related gene (Table 3). In line plots, the green lines represent potential target Potri.006G170800 and yellow lines represent lignin-related genes.*

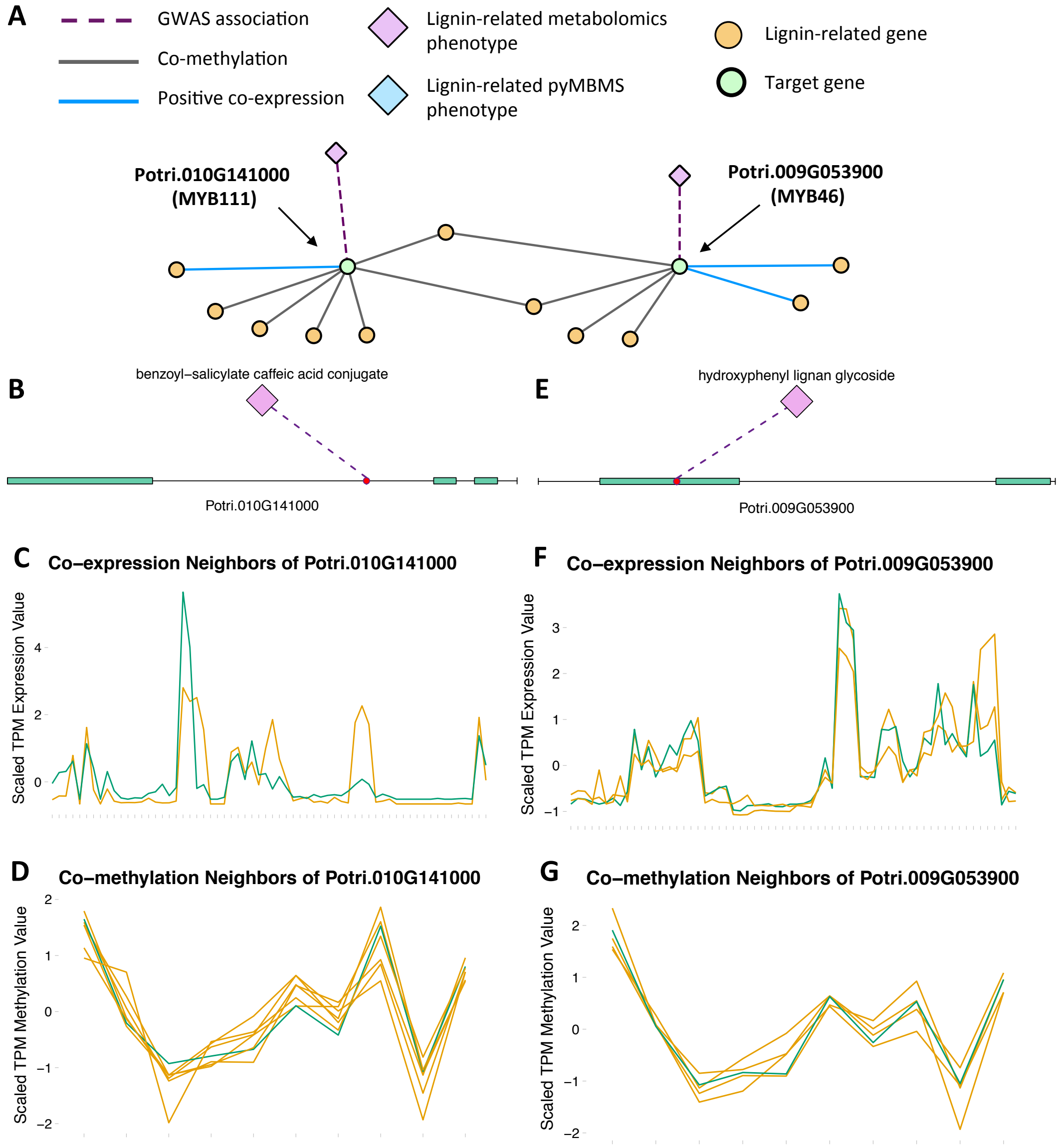

**Figure 7:** *(A) Lines of Evidence for Potri.009G053900 (homolog of Arabidopsis MYB46) and Potri.010G141000 (homolog of Arabidopsis MYB111). (B) GWAS associations of Potri.010G141000 with a lignin-related metabolite (Table 1). (C) Co-expression of Potri.010G141000 with a lignin-related gene (Table 2). (D) Co-methylation of Potri.010G141000 with six lignin-related genes (Table 3). (E) GWAS associations of Potri.009G053900 with a lignin-related metabolite (Table 1). (F) Co-expression of Potri.009G053900 with two lignin-related genes (Table 2). (G) Co-methylation of Potri.009G053900 with four lignin-related genes (Table 3). In line plots, the green lines represent potential targets Potri.009G053900/Potri.010G141000 and yellow lines represent lignin-related genes.*

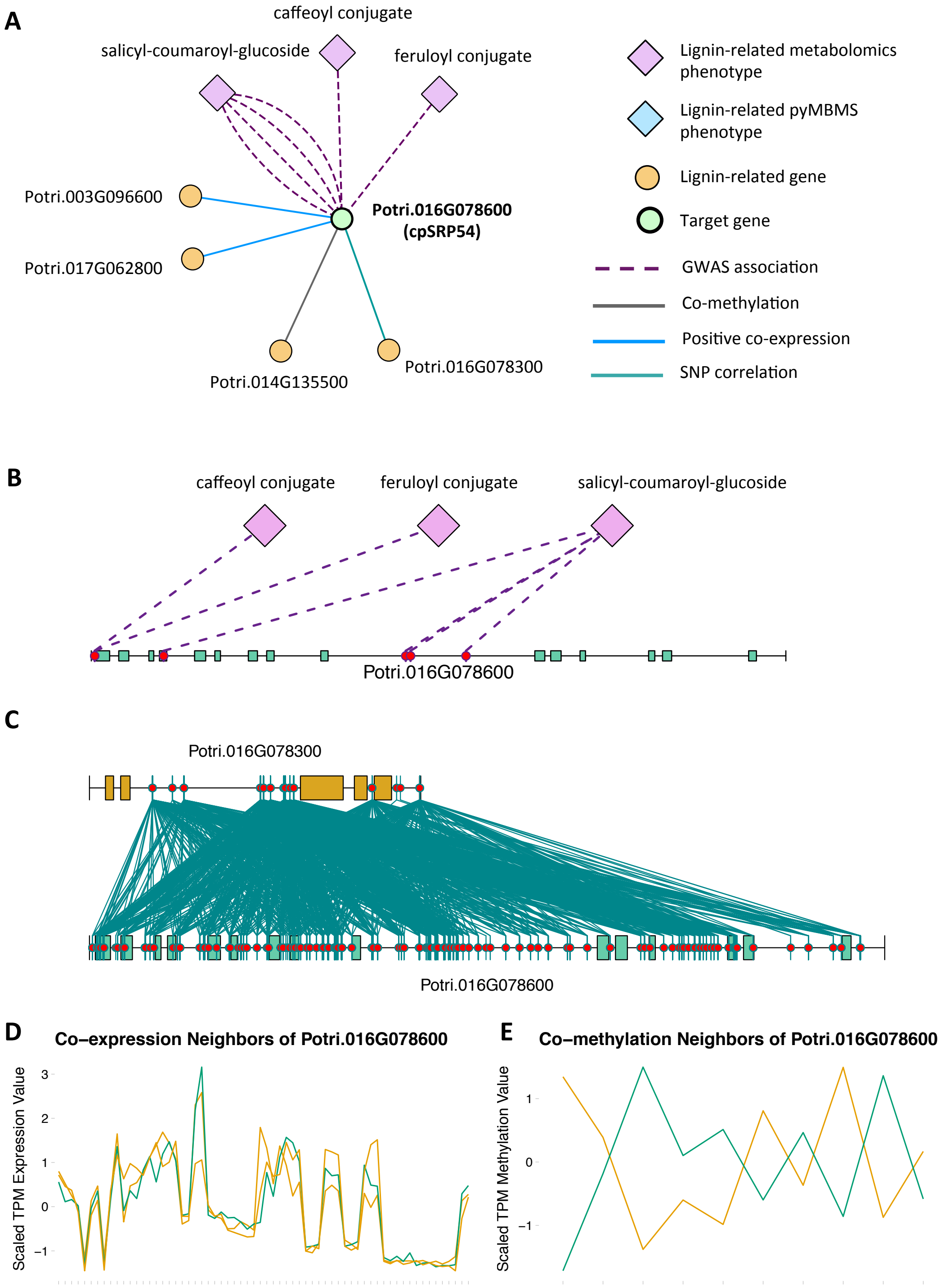

**Figure 8:** *(A) Lines of Evidence for Potri.016G078600 (homolog of Arabidopsis cpSRP54). (B) GWAS associations of Potri.016G078600 with three lignin-related metabolite (Table 1). (C) Correlations of SNPs within Potri.016G078600 with SNPs in a lignin-related gene (Table 4). (D) Co-expression of Potri.016G078600 with two lignin-related genes (Table 2). (E) Co-methylation of Potri.016G078600 with a lignin-related gene (Table 3). In line plots, the green lines represent potential target Potri.016G078600 and yellow lines represent lignin-related genes.*

## Tables

**Table 1:** *GWAS associations for select new potential target genes, indicating the SNP(s) within the potential new target gene which are associated with the lignin-related phenotype(s). Additional RT and mz information for partially identified metabolites can be seen in Supplementary Table S3.*

| Source SNP | Source Gene | Target Phenotype |
|---|---|---|
| **GWAS Associations for Potri.012G062300 (AGL8, AT5G60910)** | | |
| 12:6952245 | Potri.012G062300 | quinic acid |
| 12:6948543 | Potri.012G062300 | lignin (Syringol) |
| 12:6951532 | Potri.012G062300 | lignin (Syringol) |
| **GWAS Associations for Potri.013G102600 (AGL12, AT1G71692)** | | |
| 13:11604094 | Potri.013G102600 | 3-O-caffeoyl-quinate |
| 13:11606331 | Potri.013G102600 | coumaroyl-tremuloidin |
| 13:11600422 | Potri.013G102600 | coumaroyl-tremuloidin |
| 13:11601236 | Potri.013G102600 | hydroxyphenyl lignan glycoside |
| **GWAS Associations for Potri.007G115100 (AGL22, AT2G22540/AGL24, AT4G24540)** | | |
| 07:13650194 | Potri.007G115100 | caffeoyl conjugate |
| 07:13651354 | Potri.007G115100 | caffeoyl conjugate |
| 07:13642539 | Potri.007G115100 | caffeoyl conjugate |
| 07:13639923 | Potri.007G115100 | lignin, syringyl (Syringaldehyde) |
| **GWAS Associations for Potri.009G053900 (MYB46, AT5G12870)** | | |
| 09:5768381 | Potri.009G053900 | hydroxyphenyl lignan glycoside |
| **GWAS Associations for Potri.010G141000 (MYB111, AT5G49330)** | | |
| 10:15273000 | Potri.010G141000 | benzoyl-salicylate caffeic acid conjugate |
| **GWAS Associations for Potri.006G170800 (MYB36, AT5G57620)** | | |
| 06:17847162 | Potri.006G170800 | mz 297, RT 17.14 |
| **GWAS Associations for Potri.016G078600 (CPSRP54, AT5G03940)** | | |
| 16:5995136 | Potri.016G078600 | caffeoyl conjugate |
| 16:5995136 | Potri.016G078600 | feruloyl conjugate |
| 16:5996083 | Potri.016G078600 | salicyl-coumaroyl-glucoside |
| 16:5999408 | Potri.016G078600 | salicyl-coumaroyl-glucoside |
| 16:5999474 | Potri.016G078600 | salicyl-coumaroyl-glucoside |
| 16:6000236 | Potri.016G078600 | salicyl-coumaroyl-glucoside |

**Table 2:** *Co-expression associations for select new potential target genes. Annotations are derived from best Arabidopsis hit descriptions and GO terms and in some cases MapMan annotations.*

| **Source Gene** | **Target Gene** | **Target *Arabidopsis* best hit** | **Annotation** |
|---|---|---|---|
| **Co-expression Associations for Potri.013G102600 (AGL12, AT1G71692)** | | | |
| Potri.013G102600 | Potri.001G304800 | AT4G34050 | Caffeoyl Coenzyme A O-Methyltransferase 1 |
| Potri.013G102600 | Potri.009G099800 | AT4G34050 | Caffeoyl Coenzyme A O-Methyltransferase 1 |
| Potri.013G102600 | Potri.012G006400 | AT5G54160 | Caffeate O-Methyltransferase 1 |
| Potri.013G102600 | Potri.007G016400 | AT4G36220 | Ferulic acid 5-hydroxylase 1 |
| **Co-expression Associations for Potri.009G053900 (MYB46, AT5G12870)** | | | |
| Potri.009G053900 | Potri.003G100200 | AT1G32100 | pinoresinol reductase 1 |
| Potri.009G053900 | Potri.012G006400 | AT5G54160 | Caffeate O-Methyltransferase 1 |
| **Co-expression Associations for Potri.010G141000 (MYB111, AT5G49330)** | | | |
| Potri.010G141000 | Potri.007G030300 | AT3G50740 | UDP-glucosyl transferase 72E1 |
| **Co-expression Associations for Potri.006G170800 (MYB36, AT5G57620)** | | | |
| Potri.006G170800 | Potri.001G362800 | AT3G26300 | cytochrome P450, family 71, subfamily B, polypeptide 34/F5H |
| Potri.006G170800 | Potri.016G106100 | AT3G09220 | laccase 7 |
| Potri.006G170800 | Potri.013G120900 | AT4G35160 | N-acetylserotonin O-methyltransferase |
| **Co-expression Associations for Potri.016G078600 (CPSRP54, AT5G03940)** | | | |
| Potri.016G078600 | Potri.003G096600 | AT2G35500 | shikimate kinase like 2 |
| Potri.016G078600 | Potri.017G062800 | AT3G26900 | shikimate kinase like 1 |

**Table 3:** *Co-methylation associations for select new potential target genes. Annotations are derived from best Arabidopsis hit descriptions and GO terms and in some cases MapMan annotations.*

| **Source Gene** | **Target Gene** | **Target *Arabidopsis* best hit** | **Annotation** |
|---|---|---|---|
| **Co-methylation Associations for Potri.012G062300 (AGL8, AT5G60910)** | | | |
| Potri.012G062300 | Potri.001G036900 | AT3G21240 | 4-coumarate:CoA ligase 2 |
| Potri.012G062300 | Potri.008G120200 | AT1G68540 | Cinnamoyl CoA reductase-like 6 |
| Potri.012G062300 | Potri.004G105000 | AT5G14700 | (NAD(P)-binding Rossmann-fold superfamily protein, cinnamoyl-CoA reductase activity/CCR1 |
| **Co-methylation Associations for Potri.013G102600 (AGL12, AT1G71692)** | | | |
| Potri.013G102600 | Potri.001G334400 | AT5G63380 | 4-coumarate-CoA ligase activity /4CL |
| Potri.013G102600 | Potri.001G365300 | AT3G26300 | cytochrome P450, family 71, subfamily B, polypeptide 34/F5H |
| Potri.013G102600 | Potri.006G265500 | AT5G10820 | Major facilitator superfamily protein/Phenylpropanoid pathway |
| Potri.013G102600 | Potri.006G165200 | AT2G19070 | spermidine hydroxycinnamoyl transferase |
| **Co-methylation Associations for Potri.009G053900 (MYB46, AT5G12870)** | | | |
| Potri.009G053900 | Potri.008G196100 | AT3G06350 | bi-functional dehydroquinate-shikimate dehydrogenase enzyme |
| Potri.009G053900 | Potri.002G018300 | AT4G39330 | cinnamyl alcohol dehydrogenase 9 |
| Potri.009G053900 | Potri.004G102000 | AT4G05160 | 4-coumarate-CoA ligase activity/4CL) |
| Potri.009G053900 | Potri.008G136600 | AT1G67980 | caffeoyl-CoA 3-O-methyltransferase |
| **Co-methylation Associations for Potri.010G141000 (MYB111, AT5G49330)** | | | |
| Potri.010G141000 | Potri.008G196100 | AT3G06350 | bi-functional dehydroquinate-shikimate dehydrogenase enzyme |
| Potri.010G141000 | Potri.004G102000 | AT4G05160 | 4-coumarate-CoA ligase activity/4CL |
| Potri.010G141000 | Potri.008G074500 | AT5G34930 | arogenate dehydrogenase |
| Potri.010G141000 | Potri.005G028000 | AT5G48930 | hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase |
| Potri.010G141000 | Potri.018G100500 | AT2G23910 | NAD(P)-binding Rossmann-fold superfamily protein, cinnamoyl-CoA reductase activity/CCR1 |
| Potri.010G141000 | Potri.010G230200 | AT1G20510 | OPC-8:0 CoA ligase1, 4-coumarate-CoA ligase activity/4CL |
| **Co-methylation Associations for Potri.006G170800 (MYB36, AT5G57620)** | | | |
| Potri.006G170800 | Potri.016G093700 | AT4G05160 | AMP-dependent synthetase and ligase family, 4-coumarate-CoA ligase activity/4CL |
| **Co-methylation Associations for Potri.016G078600 (CPSRP54, AT5G03940)** | | | |
| Potri.016G078600 | Potri.014G135500 | AT3G06350 | bi-functional dehydroquinate-shikimate dehydrogenase enzyme |

**Table 4:** *SNP correlation associations for select new potential target genes. Annotations are derived from best Arabidopsis hit descriptions and GO terms and in some cases MapMan annotations.*

| **Source gene** | **Target gene** | **Target *Arabidopsis* best hit** | **Annotation** |
|---|---|---|---|
| **SNP Correlations for Potri.007G115100 (AGL22, AT2G22540/AGL24, AT4G24540)** | | | |
| Potri.007G115100 | Potri.007G116100 | AT2G22570 | nicotinamidase 1 |
| Potri.007G115100 | Potri.016G107900 | AT3G09220 | laccase 7 |
| **SNP Correlations for Potri.016G078600 (CPSRP54, AT5G03940)** | | | |
| Potri.016G078600 | Potri.016G078300 | AT4G37970 | cinnamyl alcohol dehydrogenase 6 |

# Supplementary Material

## 1. Supplementary Notes

### Note S1: Constructing Samples CCC Distribution

Printing out the complete result set of all possible pairwise comparisons of ∼10,000,000 SNPs would require more disk space than was possibly available. In order to construct an approximate distribution of the CCC values, we selected a random subset of 100,000 SNPs and calculated the CCC correlation between all pairs of these SNPs, storing all correlation values. This sampled set of correlations was used to compute the CCC distribution. Thereafter, the CCC was calculated between all pairs of all ∼10,000,000 SNPs. Only correlations meeting a threshold of 0.7 were stored.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## 2. Supplementary Figures

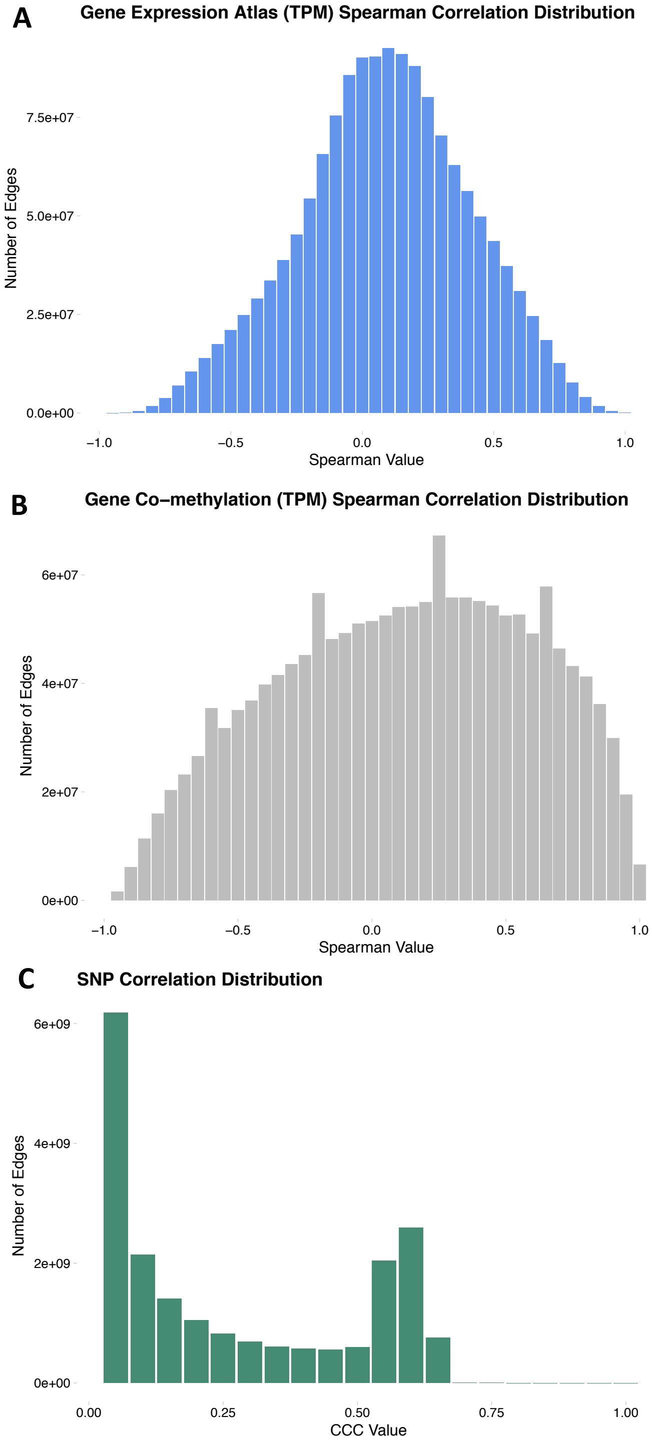

**Figure S 1:** *(A) Distribution of Spearman Correlation values in the co-expression network. (B) Distribution of Spearman Correlation values in the co-methylation network. (C) Sampled distribution of the CCC SNP correlation network. See Supplementary Note 1 for details on the construction of the sampled distribution.*

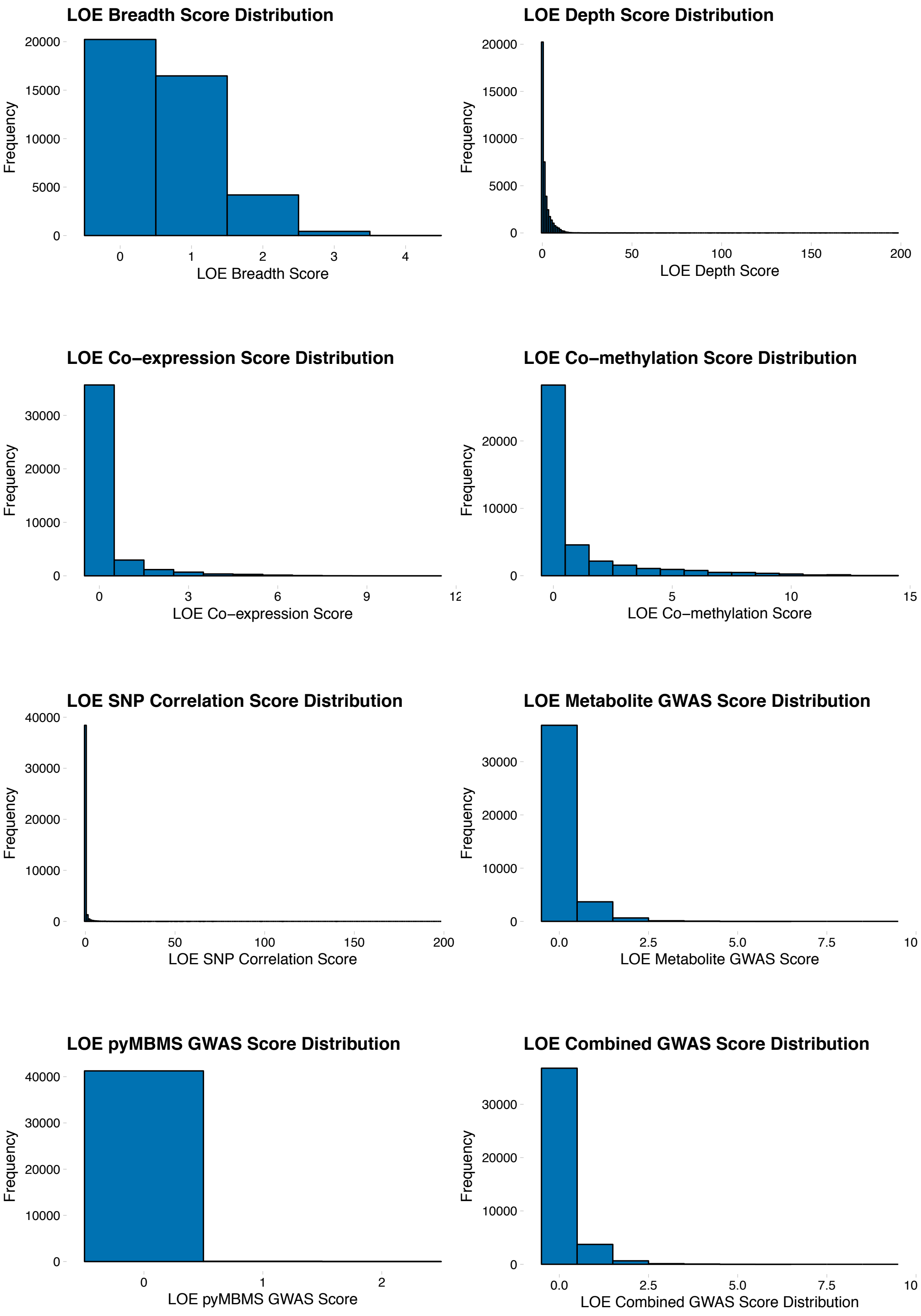

**Figure S 2:** *Lines Of Evidence (LOE) score distributions*

## 3. Supplementary Tables

**Table S 1:** *MapMan annotations of lignin genes.*

| Gene | MapMan Name |
|---|---|
| Potri.001G133200.v3.0 | secondary metabolism.flavonoids.isoflavones.isoflavone reductase : secondary metabolism.phenylpropanoids |
| Potri.003G196700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.012G094900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.001G372400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.006G097500.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.flavonoids.anthocyanins |
| Potri.T178300.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.unspecified |
| Potri.001G304800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |
| Potri.001G045000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.001G045100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.001G268600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.004G230900.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.007G029800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis : misc.UDP glucosyl and glucoronyl transferases |
| Potri.003G183900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.005G243700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.002G025700.v3.0 | misc.cytochrome P450 : secondary metabolism.phenylpropanoids.lignin biosynthesis.C3H |
| Potri.007G083000.v3.0 | misc.cytochrome P450 : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase |
| Potri.003G096600.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.shikimate kinase |
| Potri.017G033600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.013G029800.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.011G148100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.016G107900.v3.0 | secondary metabolism.simple phenols : secondary metabolism.phenylpropanoids |
| Potri.019G078100.v3.0 | secondary metabolism.flavonoids.isoflavones.isoflavone reductase : secondary metabolism.phenylpropanoids |
| Potri.007G030300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis : misc.UDP glucosyl and glucoronyl transferases |
| Potri.002G003200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.010G224100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL |
| Potri.009G062800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.014G025500.v3.0 | secondary metabolism.unspecified : secondary metabolism.phenylpropanoids |
| Potri.004G188100.v3.0 | amino acid metabolism.synthesis.aromatic aa.phenylalanine.arogenate dehydratase / prephenate dehydratase |
| Potri.001G045800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.006G199100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.001G046400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.007G083500.v3.0 | misc.cytochrome P450 : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.014G041900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis : secondary metabolism.flavonoids.dihydroflavonols : misc.UDP glucosyl and glucoronyl transferases |

| | |
|---|---|
| Potri.003G057000.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.008G038200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL |
| Potri.010G019000.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.001G307200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.016G065300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.018G100500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 : secondary metabolism.phenylpropanoids |
| Potri.008G040700.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.chorismate synthase |
| Potri.007G003800.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.001G045900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.004G053500.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.005G248500.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.010G020600.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.013G157900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C4H |
| Potri.002G004100.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.014G124100.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.008G195500.v3.0 | amino acid metabolism.synthesis.aromatic aa.phenylalanine.arogenate dehydratase / prephenate dehydratase |
| Potri.017G035100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL : secondary metabolism.phenylpropanoids |
| Potri.017G112800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.002G012800.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.016G091100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL |
| Potri.007G116100.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.001G036900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.T107000.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.007G085000.v3.0 | misc.cytochrome P450 : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.007G083200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : misc.cytochrome P450 |
| Potri.006G265500.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.003G030600.v3.0 | amino acid metabolism.synthesis.aromatic aa.tyrosine.prephenate dehydrogenase : amino acid metabolism.synthesis.aromatic aa.tyrosine.arogenate dehydrogenase & prephenate dehydrogenase |
| Potri.009G063300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.016G112400.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.flavonoids.anthocyanins |
| Potri.014G068300.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.5-enolpyruvylshikimate-3-phosphate synthase |
| Potri.013G120900.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.015G127000.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.005G028400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.007G083300.v3.0 | misc.cytochrome P450 : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase |
| Potri.007G084700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : misc.cytochrome P450 |
| Potri.005G110900.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.3-dehydroquinate synthase |

| | |
|---|---|
| Potri.004G102000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.006G048200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis : misc.UDP glucosyl and glucoronyl transferases |
| Potri.014G135500.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.010G230200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL : secondary metabolism.phenylpropanoids |
| Potri.007G030200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis : secondary metabolism.flavonoids.dihydroflavonols : misc.UDP glucosyl and glucoronyl transferases |
| Potri.008G136600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |
| Potri.016G031100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C3H |
| Potri.004G105000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.002G061100.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.shikimate kinase |
| Potri.007G084800.v3.0 | misc.cytochrome P450 : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.005G028200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.007G049200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.001G451100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.COMT : misc.O-methyl transferases |
| Potri.007G082900.v3.0 | misc.cytochrome P450 : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.003G003300.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.005G162800.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.3-deoxy-D-arabino-heptulosonate 7-phosphate synthase |
| Potri.015G003100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.COMT |
| Potri.018G104700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.005G043400.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.001G140700.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.008G196100.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.002G018300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.016G101500.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.002G076800.v3.0 | misc.O-methyl transferases : secondary metabolism.phenylpropanoids.lignin biosynthesis.COMT |
| Potri.001G334400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.003G093700.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.001G167800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C3H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : misc.cytochrome P450 |
| Potri.012G095000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.005G175400.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.001G300000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.T134100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.016G057300.v3.0 | misc.UDP glucosyl and glucoronyl transferases : secondary metabolism.flavonoids.flavonols.flavonol 3-O-glycosyltransferase : stress.biotic : secondary metabolism.phenylpropanoids.lignin biosynthesis |
| Potri.007G095700.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.3-deoxy-D-arabino-heptulosonate 7-phosphate synthase |
| Potri.010G104400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |

| | |
|---|---|
| Potri.006G169700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.006G094100.v3.0 | secondary metabolism.simple phenols : secondary metabolism.phenylpropanoids |
| Potri.007G081000.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.shikimate kinase |
| Potri.016G023300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.001G032800.v3.0 | hormone metabolism.brassinosteroid.synthesis-degradation.BRs.metabolic regulation : misc.cytochrome P450 : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.isoprenoids.carotenoids.carotenoid epsilon ring hydroxylase |
| Potri.009G099800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |
| Potri.011G004700.v3.0 | amino acid metabolism.synthesis.aromatic aa.phenylalanine.arogenate dehydratase / prephenate dehydratase |
| Potri.008G074500.v3.0 | amino acid metabolism.synthesis.aromatic aa.tyrosine.prephenate dehydrogenase : amino acid metabolism.synthesis.aromatic aa.tyrosine.arogenate dehydrogenase & prephenate dehydrogenase |
| Potri.005G084600.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.shikimate kinase |
| Potri.006G024400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.006G169600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.003G099700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.T161300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.012G006400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.COMT |
| Potri.001G128100.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.013G079500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 : secondary metabolism.phenylpropanoids |
| Potri.016G106100.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.simple phenols |
| Potri.010G125400.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.015G092300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.T149600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.005G043300.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |
| Potri.018G017400.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.009G148800.v3.0 | amino acid metabolism.synthesis.aromatic aa.phenylalanine.arogenate dehydratase / prephenate dehydratase |
| Potri.006G165200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.T071600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.004G161600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.001G133300.v3.0 | secondary metabolism.flavonoids.isoflavones.isoflavone reductase : secondary metabolism.phenylpropanoids |
| Potri.006G062600.v3.0 | amino acid metabolism.synthesis.aromatic aa.tyrosine.arogenate dehydrogenase & prephenate dehydrogenase |
| Potri.002G146400.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.5-enolpyruvylshikimate-3-phosphate synthase |
| Potri.016G106300.v3.0 | secondary metabolism.simple phenols : secondary metabolism.phenylpropanoids |
| Potri.005G147400.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.010G221600.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.chorismate synthase |
| Potri.018G104800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.001G042900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.005G028000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.004G017900.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.005G028100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.008G120200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.010G186300.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.018G109900.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.010G224200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL |
| Potri.007G016400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |

| | |
|---|---|
| Potri.001G362800.v3.0 | misc.cytochrome P450 : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.006G126800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL |
| Potri.008G082300.v3.0 | secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : misc.cytochrome P450 |
| Potri.T149400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.010G054200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.002G004500.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.001G150500.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.3-deoxy-D-arabino-heptulosonate 7-phosphate synthase |
| Potri.004G013400.v3.0 | amino acid metabolism.synthesis.aromatic aa.phenylalanine.arogenate dehydratase / prephenate dehydratase |
| Potri.016G093700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.009G095800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.009G063400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.003G100200.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.flavonoids.isoflavones.isoflavone reductase |
| Potri.018G021200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.018G105400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.002G183600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |
| Potri.007G083700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : misc.cytochrome P450 |
| Potri.009G062900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.018G146100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C4H |
| Potri.001G045300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.018G094200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.017G034900.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL : secondary metabolism.phenylpropanoids |
| Potri.019G048200.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.shikimate kinase |
| Potri.005G257700.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.018G070300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |
| Potri.002G086000.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.008G031500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL : secondary metabolism.phenylpropanoids |
| Potri.001G201100.v3.0 | amino acid metabolism.synthesis.aromatic aa.tyrosine.prephenate dehydrogenase |
| Potri.007G083600.v3.0 | secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : misc.cytochrome P450 |
| Potri.012G094800.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.016G078300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.007G030400.v3.0 | misc.UDP glucosyl and glucoronyl transferases : secondary metabolism.phenylpropanoids.lignin biosynthesis |
| Potri.003G057200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.008G071200.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.016G031000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C3H |
| Potri.003G188500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.008G136700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCoAOMT |
| Potri.001G045500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.006G141400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : misc.cytochrome P450 |

| | |
|---|---|
| Potri.005G073300.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.3-deoxy-D-arabino-heptulosonate 7-phosphate synthase |
| Potri.003G181400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.014G106600.v3.0 | misc.O-methyl transferases : secondary metabolism.phenylpropanoids.lignin biosynthesis.COMT |
| Potri.002G026000.v3.0 | misc.cytochrome P450 : secondary metabolism.phenylpropanoids.lignin biosynthesis.C3H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase |
| Potri.019G126400.v3.0 | polyamine metabolism : secondary metabolism.phenylpropanoids |
| Potri.006G033300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C3H |
| Potri.017G062800.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.shikimate kinase |
| Potri.018G105500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.HCT |
| Potri.001G363900.v3.0 | misc.cytochrome P450 : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase |
| Potri.007G084400.v3.0 | secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : misc.cytochrome P450 |
| Potri.T149300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.006G078100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C4H |
| Potri.001G365300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H : misc.cytochrome P450 |
| Potri.019G130700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.C4H |
| Potri.008G024800.v3.0 | secondary metabolism.flavonoids.dihydroflavonols : misc.UDP glucosyl and glucoronyl transferases : secondary metabolism.phenylpropanoids.lignin biosynthesis : hormone metabolism.salicylic acid.synthesis-degradation : secondary metabolism.flavonoids.anthocyanins.anthocyanidin 3-O-glucosyltransferase |
| Potri.009G076300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.019G049500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.005G175600.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.003G210700.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.014G025600.v3.0 | secondary metabolism.phenylpropanoids : secondary metabolism.unspecified |
| Potri.009G123600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.001G045700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.001G046100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.003G057100.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.003G210600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL : secondary metabolism.phenylpropanoids : lipid metabolism.FA synthesis and FA elongation.acyl coa ligase |
| Potri.001G045600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.006G178700.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 : secondary metabolism.phenylpropanoids |
| Potri.005G117500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.006G024300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.001G365100.v3.0 | misc.cytochrome P450 : secondary metabolism.flavonoids.dihydroflavonols.flavonoid 3-monooxygenase : secondary metabolism.phenylpropanoids.lignin biosynthesis.F5H |
| Potri.010G057000.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.4CL |
| Potri.001G045400.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.007G030500.v3.0 | misc.UDP glucosyl and glucoronyl transferases : secondary metabolism.flavonoids.dihydroflavonols : secondary metabolism.phenylpropanoids.lignin biosynthesis |
| Potri.013G029900.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.dehydroquinate/shikimate dehydrogenase |

| | |
|---|---|
| Potri.017G110500.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.002G099200.v3.0 | amino acid metabolism.synthesis.aromatic aa.chorismate.3-deoxy-D-arabino-heptulosonate 7-phosphate synthase |
| Potri.001G055700.v3.0 | secondary metabolism.phenylpropanoids |
| Potri.019G084300.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.COMT |
| Potri.011G148200.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |
| Potri.001G349600.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CCR1 |
| Potri.009G063100.v3.0 | secondary metabolism.phenylpropanoids.lignin biosynthesis.CAD |

**Table S 2:** *Mass/Charge (mz) ratio for Lignin pyMBMS Peaks.*

| mz | Annotation (Sykes et al., 2009) |
|---|---|
| 120 | lignin (vinylphenol) |
| 124 | lignin, guaiacyl |
| 137 | lignin,guaiacyl (Ethylguaiacol, homovanillin,coniferyl alcohol) |
| 138 | lignin,guaiacyl (Methylguaiacol) |
| 150 | lignin,guaiacyl (Vinylguaiacol) |
| 152 | lignin |
| 154 | lignin,syringyl (Syringol) |
| 168 | syringyl (4-Methyl-2,6-dimethoxyphenol) |
| 180 | lignin (Coniferyl alcohol, syringylethene) |
| 182 | lignin,syringyl (Syringaldehyde) |
| 210 | lignin,syringyl (Sinapylalcohol) |

**Table S 3:** *Lignin-related metabololites from the metabolomics analysis. For partially identified metabolites, additional RT and mz information is provided.*

See attached excel file.

**Table S 4:** *LOE Scores, Arabidopsis best hits and MapMan annotations of genes for which* $LOE_{breadth}(g) \geq 3$ *and* $LOE_{gwas}(g) \geq 1$.

See attached excel file.

**Table S 5:** *Positions of SNPs involved in SNP correlations in select portential new target genes.*

| Source | Target |
|---|---|
| **SNP Correlations between Potri.007G115100 (AGL22) and Potri.016G107900 (laccase 7)** | |
| SNP 07:13647758 | SNP 16:11083690 |
| SNP 07:13647758 | SNP 16:11083708 |
| SNP 07:13647758 | SNP 16:11083712 |
| SNP 07:13647758 | SNP 16:11083737 |
| SNP 07:13647978 | SNP 16:11083690 |
| SNP 07:13647978 | SNP 16:11083708 |
| SNP 07:13647978 | SNP 16:11083712 |
| SNP 07:13647978 | SNP 16:11083737 |
| SNP 07:13648235 | SNP 16:11083690 |
| SNP 07:13648235 | SNP 16:11083708 |
| SNP 07:13648235 | SNP 16:11083712 |
| SNP 07:13648235 | SNP 16:11083737 |
| SNP 07:13648488 | SNP 16:11083690 |
| SNP 07:13648488 | SNP 16:11083708 |
| SNP 07:13648488 | SNP 16:11083712 |
| SNP 07:13648488 | SNP 16:11083737 |
| **SNP Correlations between Potri.007G115100 (AGL22) and Potri.007G116100 (nicotinamidase 1)** | |
| SNP 07:13647645 | SNP 07:13706654 |
| SNP 07:13647645 | SNP 07:13706699 |
| SNP 07:13647645 | SNP 07:13706834 |
| SNP 07:13647758 | SNP 07:13706654 |
| SNP 07:13647758 | SNP 07:13706699 |
| SNP 07:13647978 | SNP 07:13706654 |
| SNP 07:13647978 | SNP 07:13706699 |
| SNP 07:13647978 | SNP 07:13706834 |

## References

Robert Sykes, Matthew Yung, Evandro Novaes, Matias Kirst, Gary Peter, and Mark Davis. High-Throughput Screening of Plant Cell-Wall Composition Using Pyrolysis Molecular Beam Mass Spectroscopy. *Biofuels: Methods and protocols*, pages 169–183, 2009.

**Supplementary Table S3**

| **Complete Identifications** |
|---|
| 3-O-caffeoyl-quinate |
| 4-O-caffeoyl-quinate |
| 5-hydroxyferulic acid |
| 5-hydroxyferulic acid-glucoside |
| 5-O-caffeoyl-quinate |
| benzyl-coumaroyl-glucoside |
| caffeic acid |
| cis-3-O-caffeoyl-quinate |
| cis-cinnamic acid |
| cis-p-coumaric acid |
| coniferin |
| coniferyl alcohol |
| coumaric acid-4-O-glucoside |
| coumaroyl-tremuloidin |
| ferulic acid |
| guaiacylglycerol |
| p-hydroxybenzoic acid |
| phenylalanine |
| quinic acid |
| salicyl-coumaroyl-glucoside |
| salicyloyl-coumaroyl-glucoside |
| secoisolariciresinol |
| shikimic acid |
| syringaresinol |
| syringin |
| syringin |
| trans-cinnamic acid |
| trans-p-coumaric acid |
| vanillic acid-4-O-glucoside |

**Partial Identifications**

| RT | mz | ID |
|---|---|---|
| 13.26 | 249 | 13.26 249 feruloyl glycoside |
| 15.62 | 279 | 15.62 279 297 217 guaiacyl lignan glycoside |
| 15.6 | 297 | 15.60 297 279 217 guaiacyl lignan glycoside |
| 15.71 | 267 | 15.71 267 204 hydroxyphenyl lignan glycoside |
| 15.98 | 297 | 15.98 297 361 209 guaiacyl lignan glycoside |
| 16.02 | 350 | 16.02 350 361 219 glycoside |
| 16.05 | 219 | 16.05 219 coumaroyl conjugate |
| 16.08 | 297 | 16.08 297 583 361 glycoside |
| 16.12 | 297 | 16.12 297 225 guaiacyl lignan |
| 16.16 | 327 | 16.16 327 307 syringyl lignan |
| 16.2 | 219 | 16.20 219 468 453 coumaroyl conjugate |
| 16.27 | 105 | 16.27 105 396 179 benzoyl-salicylate caffeic acid conjugate |
| 16.3 | 297 | 16.30 297 204 583 glycoside |
| 16.5 | 297 | 16.50 297 guaiacyl lignan |
| 16.51 | 327 | 16.51 327 syringyl lignan |
| 16.54 | 327 | 16.54 327 297 369 syringyl-guaiacyl lignan |
| 16.61 | 219 | 16.61 219 283 204 glycoside |
| 16.69 | 297 | 16.69 297 354 171 209 phenolic |
| 17.14 | 297 | 17.14 297 282 phenolic |
| 17.43 | 91 | 17.43 91 476 benzyl conjugate |
| 17.44 | 171 | 17.44 171 219 331 coumaroyl conjugate |
| 17.5 | 171 | 17.50 171 219 331 coumaroyl conjugate |
| 17.65 | 538 | 17.65 538 644 452 320 293 219 |
| 17.69 | 219 | 17.69 219 204 coumaroyl glycoside |
| 17.96 | 171 | 17.96 171 381 219 204 coumaroyl caffeoyl glycoside |
| 18.04 | 171 | 18.04 171 381 219 204 coumaroyl caffeoyl glycoside |
| 18.07 | 255 | 18.07 255 219 171 119 coumaroyl conjugate |

| | | |
|---|---|---|
| 18.2 | 219 | 18.20 219 307 caffeoyl conjugate |
| 18.32 | 219 | 18.32 219 331 171 coumaroyl conjugate |
| 18.53 | 219 | 18.53 219 331 171 coumaroyl conjugate |
| 19.1 | 171 | 19.10 171 381 219 caffeoyl conjugate |
| 19.16 | 219 | 19.16 219 307 caffeoyl conjugate |
| 20.17 | 171 | 20.17 171 204 307 469 caffeoyl-quercetin glycoside |
| 20.26 | 171 | 20.26 171 469 204 307 caffeoyl-quercetin-glycoside |

Supplementary Table S4

| *Populus. trichocarpa* gene id | ***Arabidopsis thaliana* best hit** | **MapMan Annotation** | **LOE Depth Score** | **LOE Breadth Score** | **LOE Co-expression Score** | **LOE Co-methylation Score** | **LOE Metabolite GWAS Score** | **LOE pyMBMS GWAS Score** | **LOE SNP Correlation Score** | **LOE Combined GWAS Score** |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.001G006600 | AT1G55320 | lipid metabolism.FA synthesis and FA elongation.acyl coa ligase | 15 | 3 | 2 | 12 | 1 | 0 | 0 | 1 |
| Potri.001G021000 | AT5G13460 | signalling.calcium | 8 | 3 | 2 | 5 | 1 | 0 | 0 | 1 |
| Potri.001G032500 | AT5G07050 | development.unspecified | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G064100 | AT1G74170,<br>AT1G54470 | signalling.receptor kinases.misc : stress.biotic | 11 | 3 | 0 | 2 | 1 | 0 | 8 | 1 |
| Potri.001G064600 | AT1G74170,<br>AT1G54470 | stress.biotic : signalling.receptor kinases.misc | 30 | 3 | 0 | 1 | 2 | 0 | 27 | 2 |
| Potri.001G073400 | AT4G20300 | not assigned.unknown | 6 | 3 | 2 | 3 | 1 | 0 | 0 | 1 |
| Potri.001G089700 | AT4G23950,<br>AT1G71360 | not assigned.unknown | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.001G092300 | AT1G64385 | not assigned.unknown | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.001G107000 | AT5G23750 | not assigned.unknown | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.001G111000 | AT1G12600,<br>AT4G23010 | transport.NDP-sugars at the ER | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.001G111400 | AT4G22990 | not assigned.no ontology | 5 | 4 | 1 | 2 | 1 | 0 | 1 | 1 |
| Potri.001G119100 | AT1G62780 | not assigned.unknown | 15 | 4 | 2 | 10 | 2 | 0 | 1 | 2 |
| Potri.001G125000 | AT5G62410 | DNA.synthesis/chromatin structure | 13 | 3 | 2 | 10 | 1 | 0 | 0 | 1 |
| Potri.001G132800 | AT1G32120 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G148200 | AT1G73320 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G180900 |  | not assigned.unknown | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.001G183400 | AT2G25737 | not assigned.unknown | 5 | 3 | 1 | 2 | 2 | 0 | 0 | 2 |
| Potri.001G197900 | AT3G13920 | protein.synthesis.initiation | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G201200 |  | not assigned.unknown | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.001G203600 | AT3G15890 | development.unspecified : signalling.receptor kinases.misc | 8 | 3 | 0 | 6 | 1 | 0 | 1 | 1 |
| Potri.001G209600 | AT3G26070 | cell.organisation | 12 | 3 | 4 | 7 | 1 | 0 | 0 | 1 |
| Potri.001G214700 | AT1G08465 | RNA.regulation of transcription.C2C2(Zn) YABBY family | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G243700 |  | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G253300 | AT2G13360 | amino acid metabolism.synthesis.central amino acid metabolism.alanine.alanine-glyoxylate aminotransferase : PS.photorespiration.aminotransferases peroxisomal : amino acid metabolism.synthesis.serine-glycine-cysteine group.glycine.serine glyoxylate aminotransferase | 5 | 3 | 2 | 2 | 1 | 0 | 0 | 1 |
| Potri.001G280100 | AT3G51770 | hormone metabolism.ethylene.synthesis-degradation | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G286900 | AT5G09650 | nucleotide metabolism.phosphotransfer and pyrophosphatases.misc | 9 | 3 | 4 | 3 | 2 | 0 | 0 | 2 |
| Potri.001G289100 | AT2G16270 | not assigned.unknown | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.001G317600 | AT1G04920 | major CHO metabolism.synthesis.sucrose.SPS | 7 | 3 | 1 | 5 | 1 | 0 | 0 | 1 |
| Potri.001G357000 | AT4G26910 | TCA / org transformation.TCA.2-oxoglutarate dehydrogenase | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.001G358100 | AT4G21070 | protein.degradation.ubiquitin.E3.RING | 7 | 3 | 1 | 5 | 1 | 0 | 0 | 1 |
| Potri.001G358600 | AT1G30320 | RNA.regulation of transcription.unclassified | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.001G377100 | AT1G12790 | not assigned.unknown | 5 | 3 | 1 | 2 | 2 | 0 | 0 | 2 |
| Potri.001G404700 | AT3G15520 | cell.cycle.peptidylprolyl isomerase | 9 | 3 | 5 | 3 | 1 | 0 | 0 | 1 |
| Potri.001G464700 | AT5G44440 | misc.nitrilases, *nitrile lyases, berberine bridge enzymes, reticuline oxidases, troponine reductases | 6 | 3 | 3 | 2 | 1 | 0 | 0 | 1 |
| Potri.001G468100 | AT4G26530 | glycolysis.cytosolic branch.aldolase : PS.calvin cycle.aldolase | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 1 |
| Potri.002G013400 | AT5G42250 | misc.alcohol dehydrogenases | 8 | 3 | 2 | 2 | 4 | 0 | 0 | 4 |
| Potri.002G072400 | AT1G77090 | PS.lightreaction.photosystem II.PSII polypeptide subunits | 9 | 3 | 1 | 7 | 1 | 0 | 0 | 1 |
| Potri.002G077600 | AT4G37925 | PS.lightreaction.NADH DH | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.002G083400 | AT1G77580 | not assigned.no ontology | 4 | 3 | 2 | 0 | 1 | 0 | 1 | 1 |
| Potri.002G113600 | AT1G10070 | Co-factor and vitamine metabolism.pantothenate.branched-chain amino acid aminotransferase : amino acid metabolism.synthesis.branched chain group.common.branched-chain amino acid aminotransferase | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.002G114900 | AT1G44920 | not assigned.unknown | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.002G125200 | AT1G45207 | RNA.regulation of transcription.unclassified | 6 | 3 | 3 | 2 | 1 | 0 | 0 | 1 |
| Potri.002G155500 | AT3G61200 | not assigned.no ontology | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.002G168200 | AT2G46370 | hormone metabolism.auxin.induced-regulated-responsive-activated | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.002G204900 | AT5G61770 | protein.synthesis.ribosome biogenesis.BRIX | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.002G210000 | AT1G74680 | misc.UDP glucosyl and glucoronyl transferases | 6 | 3 | 1 | 3 | 2 | 0 | 0 | 2 |
| Potri.002G212300 | | not assigned.unknown | 4 | 3 | 0 | 2 | 1 | 0 | 1 | 1 |
| Potri.002G215100 | AT2G30950 | protein.degradation.metalloprotease | 11 | 3 | 7 | 3 | 1 | 0 | 0 | 1 |
| Potri.002G218300 | AT2G44310 | signalling.calcium | 9 | 3 | 0 | 2 | 1 | 0 | 6 | 1 |
| Potri.002G251600 | AT5G48460 | cell.organisation | 6 | 3 | 3 | 2 | 1 | 0 | 0 | 1 |
| Potri.002G253400 | AT4G28780 | misc.GDSL-motif lipase | 6 | 3 | 1 | 2 | 3 | 0 | 0 | 3 |
| Potri.002G253600 | AT4G28760 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.003G099600 | AT1G32080 | not assigned.no ontology | 10 | 3 | 6 | 0 | 3 | 0 | 1 | 3 |
| Potri.003G100000 | AT4G19020 | RNA.regulation of transcription.DNA methyltransferases | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.003G112800 | AT2G45190 | RNA.regulation of transcription.C2C2(Zn) YABBY family | 10 | 3 | 2 | 5 | 3 | 0 | 0 | 3 |
| Potri.003G134300 | AT1G64150 | not assigned.unknown | 16 | 3 | 5 | 10 | 1 | 0 | 0 | 1 |
| Potri.003G145900 | AT5G41970 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.003G152400 | | not assigned.unknown | 5 | 3 | 1 | 0 | 2 | 0 | 2 | 2 |
| Potri.003G162000 | AT1G27440 | misc.UDP glucosyl and glucoronyl transferases | 8 | 3 | 3 | 4 | 1 | 0 | 0 | 1 |
| Potri.003G163200 | AT5G13300 | signalling.G-proteins | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.003G168800 | AT1G27980 | lipid metabolism.exotics (steroids, squalene etc).sphingolipids | 9 | 3 | 4 | 0 | 2 | 0 | 3 | 2 |
| Potri.003G173200 | AT1G67370 | RNA.regulation of transcription.putative transcription regulator | 10 | 3 | 1 | 7 | 2 | 0 | 0 | 2 |
| Potri.003G199100 | AT3G14470 | stress.biotic.PR-proteins | 20 | 3 | 1 | 0 | 2 | 0 | 17 | 2 |
| Potri.003G200200 | AT3G14470 | stress.biotic.PR-proteins | 12 | 3 | 0 | 6 | 4 | 0 | 2 | 4 |
| Potri.003G215800 | AT1G69770 | RNA.regulation of transcription.DNA methyltransferases | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.004G008200 | AT4G22030 | protein.degradation.ubiquitin.E3.SCF.FBOX | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.004G031600 | AT4G21270,<br>AT4G05190 | cell.organisation | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.004G065200 | AT4G18360 | PS.photorespiration.glycolate oxydase | 13 | 3 | 3 | 9 | 1 | 0 | 0 | 1 |
| Potri.004G068600 | | not assigned.unknown | 8 | 3 | 2 | 5 | 1 | 0 | 0 | 1 |
| Potri.004G069400 | AT1G48380 | RNA.regulation of transcription.Orphan family : development.unspecified | 7 | 3 | 1 | 2 | 4 | 0 | 0 | 4 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.004G070500 | AT4G00690,<br>AT1G14920,<br>AT2G01570 | RNA.regulation of transcription.GRAS transcription factor family | 8 | 3 | 6 | 0 | 1 | 0 | 1 | 1 |
| Potri.004G162500 | AT4G08350 | RNA.regulation of transcription.Global transcription factor group | 5 | 3 | 0 | 2 | 2 | 0 | 1 | 2 |
| Potri.004G162600 | AT4G38960 | RNA.regulation of transcription.C2C2(Zn) CO-like, Constans-like zinc finger family | 12 | 3 | 0 | 10 | 1 | 0 | 1 | 1 |
| Potri.004G173300 | AT4G38650 | not assigned.no ontology | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.004G177400 | AT4G14210 | secondary metabolism.isoprenoids.carotenoids.phytoene dehydrogenase | 9 | 3 | 4 | 4 | 1 | 0 | 0 | 1 |
| Potri.004G199400 | AT1G08380 | PS.lightreaction.photosystem I.PSI polypeptide subunits | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 1 |
| Potri.005G000500 | AT5G26742 | RNA.processing.RNA helicase | 11 | 3 | 3 | 7 | 1 | 0 | 0 | 1 |
| Potri.005G010700 | AT4G15510 | PS.lightreaction.photosystem II.PSII polypeptide subunits | 5 | 3 | 2 | 1 | 2 | 0 | 0 | 2 |
| Potri.005G014600 | AT2G24100 | not assigned.unknown | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.005G020800 | AT3G05710 | cell.vesicle transport | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.005G021100 | AT5G27000,<br>AT1G09170 | cell.organisation | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G039100 | AT5G27410 | misc.aminotransferases.aminotransferase class IV family protein | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.005G039300 | AT2G29150 | secondary metabolism.N misc.alkaloid-like : misc.nitrilases, *nitrile lyases, berberine bridge enzymes, reticuline oxidases, troponine reductases | 8 | 3 | 4 | 0 | 1 | 0 | 3 | 1 |
| Potri.005G040700 | AT1G54690 | DNA.synthesis/chromatin structure.histone.core.H2A | 5 | 4 | 2 | 1 | 1 | 0 | 1 | 1 |
| Potri.005G041000 | AT3G04970 | RNA.regulation of transcription.unclassified | 4 | 3 | 1 | 0 | 2 | 0 | 1 | 2 |
| Potri.005G044900 | AT1G08820 | cell.vesicle transport | 5 | 3 | 2 | 2 | 1 | 0 | 0 | 1 |
| Potri.005G058400 | AT4G03520 | redox.thioredoxin | 9 | 3 | 6 | 1 | 2 | 0 | 0 | 2 |
| Potri.005G059500 | AT5G24550 | secondary metabolism.sulfur-containing.glucosinolates.degradation.myrosinase : misc.gluco-, galacto- and mannosidases : stress.biotic | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G067000 | AT1G77280 | protein.postranslational modification.kinase.receptor like cytoplasmatic kinase VI | 11 | 3 | 1 | 9 | 1 | 0 | 0 | 1 |
| Potri.005G077200 | AT5G65140 | minor CHO metabolism.trehalose.TPP | 8 | 3 | 0 | 6 | 1 | 0 | 1 | 1 |
| Potri.005G092700 | AT5G23150,<br>AT5G08230 | RNA.regulation of transcription.PWWP domain protein | 4 | 3 | 0 | 1 | 1 | 0 | 2 | 1 |
| Potri.005G094700 | AT3G19800 | not assigned.unknown | 9 | 3 | 4 | 4 | 1 | 0 | 0 | 1 |
| Potri.005G096700 | AT1G31340 | protein.degradation.ubiquitin : protein.degradation.ubiquitin.ubiquitin | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G108900 | AT5G42180 | misc.peroxidases | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G112400 | AT4G34950 | development.unspecified | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G117600 | AT5G66320 | RNA.regulation of transcription.C2C2(Zn) GATA transcription factor family | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G136400 | AT4G37080 | not assigned.unknown | 9 | 3 | 5 | 3 | 1 | 0 | 0 | 1 |
| Potri.005G171400 | AT1G77810 | misc.UDP glucosyl and glucoronyl transferases : protein.glycosylation | 10 | 3 | 1 | 8 | 1 | 0 | 0 | 1 |
| Potri.005G178200 | AT1G21790 | not assigned.unknown | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |
| Potri.005G185600 | AT1G76990 | amino acid metabolism.misc | 9 | 3 | 2 | 6 | 1 | 0 | 0 | 1 |
| Potri.005G192500 | | not assigned.unknown | 8 | 3 | 0 | 5 | 1 | 0 | 2 | 1 |
| Potri.005G212500 | AT3G63230 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.005G220400 | AT1G06730 | minor CHO metabolism.others | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G227600 | AT4G28560 | protein.postranslational modification | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G231600 | AT1G35180,<br>AT1G45010 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.005G254100 | AT1G42970 | PS.calvin cycle.GAP | 9 | 3 | 6 | 2 | 1 | 0 | 0 | 1 |
| Potri.005G255800 | AT2G04865 | protein.postranslational modification | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.006G014300 | AT1G53430 | signalling.receptor kinases.leucine rich repeat VIII.VIII-2 | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |
| Potri.006G064300 | AT5G20280 | major CHO metabolism.synthesis.sucrose.SPS | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.006G067600 | AT4G02530 | not assigned.no ontology | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.006G107300 | AT3G51730 | not assigned.no ontology | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 1 |
| Potri.006G112800 | AT3G54200 | not assigned.no ontology | 8 | 3 | 3 | 4 | 1 | 0 | 0 | 1 |
| Potri.006G126700 | AT5G03290 | TCA / org transformation.other organic acid transformatons.IDH | 4 | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Potri.006G141500 | AT2G05940 | protein.postranslational modification.kinase.receptor like cytoplasmatic kinase VII | 6 | 3 | 2 | 3 | 1 | 0 | 0 | 1 |
| Potri.006G151100 | AT5G56850 | not assigned.unknown | 9 | 3 | 3 | 3 | 3 | 0 | 0 | 3 |
| Potri.006G153300 | AT5G19740 | protein.degradation | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 1 |
| Potri.006G158200 | AT4G21900 | transport.misc | 4 | 3 | 1 | 0 | 1 | 0 | 2 | 1 |
| Potri.006G161400 | AT4G29080 | RNA.regulation of transcription.Aux/IAA family | 5 | 3 | 1 | 0 | 1 | 0 | 3 | 1 |
| Potri.006G164600 | AT4G26080 | hormone metabolism.abscisic acid.signal transduction : protein.postranslational modification | 6 | 3 | 0 | 1 | 1 | 0 | 4 | 1 |
| Potri.006G165500 | AT5G57340 | not assigned.unknown | 14 | 3 | 0 | 12 | 1 | 0 | 1 | 1 |
| Potri.006G170800 | AT5G57620 | RNA.regulation of transcription.MYB domain transcription factor family | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.006G175300 | AT2G18710 | protein.targeting.chloroplast | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.006G190000 | AT5G25220 | RNA.regulation of transcription.HB,Homeobox transcription factor family | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.006G221400 | AT1G33240 | RNA.regulation of transcription.Trihelix, Triple-Helix transcription factor family | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.006G249100 | AT5G60990 | protein.synthesis.ribosome biogenesis.Pre-rRNA processing and modifications.DExD-box helicases | 8 | 3 | 1 | 5 | 2 | 0 | 0 | 2 |
| Potri.006G258700 | AT2G25270 | not assigned.unknown | 10 | 3 | 3 | 6 | 1 | 0 | 0 | 1 |
| Potri.006G265400 | AT2G25080 | redox.ascorbate and glutathione.glutathione | 5 | 3 | 2 | 2 | 1 | 0 | 0 | 1 |
| Potri.006G267400 | AT2G24800 | misc.peroxidases | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |
| Potri.006G268500 | AT2G29110 | signalling.in sugar and nutrient physiology | 18 | 3 | 1 | 0 | 2 | 0 | 15 | 2 |
| Potri.006G268800 | AT5G10730 | not assigned.unknown | 8 | 3 | 0 | 4 | 3 | 0 | 1 | 3 |
| Potri.006G269500 | AT5G10720 | hormone metabolism.cytokinin.signal transduction | 11 | 4 | 1 | 5 | 1 | 0 | 4 | 1 |
| Potri.007G030100 | AT4G36760 | protein.degradation | 13 | 3 | 0 | 1 | 1 | 0 | 11 | 1 |
| Potri.007G038200 | AT4G37160 | misc.oxidases - copper, flavone etc | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.007G039500 | AT5G66870 | RNA.regulation of transcription.AS2,Lateral Organ Boundaries Gene Family | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.007G056400 | AT2G17820 | hormone metabolism.cytokinin.signal transduction | 3 | 3 | 1 | 0 | 1 | 1 | 0 | 2 |
| Potri.007G063100 | AT4G35500 | protein.postranslational modification | 10 | 3 | 5 | 4 | 0 | 1 | 0 | 1 |
| Potri.007G063600 | AT4G35890 | not assigned.no ontology | 6 | 3 | 0 | 4 | 1 | 0 | 1 | 1 |
| Potri.007G082500 | AT5G64940 | stress.abiotic | 15 | 3 | 3 | 0 | 2 | 0 | 10 | 2 |
| Potri.007G105900 | AT5G64040 | PS.lightreaction.photosystem I.PSI polypeptide subunits | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 1 |
| Potri.007G107900 | AT4G05410 | RNA.processing : signalling.G-proteins | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.007G109100 | AT5G09330 | development.unspecified | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.007G115100 | AT4G24540,<br>AT2G22540 | RNA.regulation of transcription.MADS box transcription factor family | 4 | 3 | 0 | 0 | 1 | 1 | 2 | 2 |
| Potri.007G133000 | AT2G01570 | RNA.regulation of transcription.GRAS transcription factor family | 58 | 4 | 2 | 1 | 1 | 0 | 54 | 1 |
| Potri.007G142900 | | not assigned.unknown | 9 | 3 | 0 | 2 | 1 | 0 | 6 | 1 |
| Potri.008G011400 | AT3G10660 | signalling.calcium | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.008G021200 | AT2G36290 | not assigned.no ontology | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.008G032700 | AT5G04310 | cell wall.degradation.pectate lyases and polygalacturonases | 9 | 3 | 2 | 6 | 1 | 0 | 0 | 1 |
| Potri.008G041000 | AT3G54890 | PS.lightreaction.photosystem I.LHC-I | 8 | 3 | 5 | 2 | 1 | 0 | 0 | 1 |
| Potri.008G058800 | AT2G44920 | not assigned.no ontology | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.008G058900 | AT3G11420 | not assigned.no ontology | 11 | 3 | 0 | 8 | 1 | 0 | 2 | 1 |
| Potri.008G069900 | AT3G55990 | stress.abiotic.cold | 9 | 3 | 3 | 5 | 1 | 0 | 0 | 1 |
| Potri.008G090500 | AT1G14030 | PS.calvin cycle.rubisco interacting | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.008G092700 | AT2G03360 | not assigned.unknown | 13 | 3 | 11 | 1 | 1 | 0 | 0 | 1 |
| Potri.008G093900 | AT3G29320 | major CHO metabolism.degradation.starch.starch phosphorylase | 9 | 3 | 1 | 7 | 1 | 0 | 0 | 1 |
| Potri.008G117600 | AT1G29700 | not assigned.unknown | 7 | 3 | 1 | 5 | 1 | 0 | 0 | 1 |
| Potri.008G119600 | AT4G39670 | not assigned.unknown | 10 | 3 | 1 | 7 | 2 | 0 | 0 | 2 |
| Potri.008G128600 | AT1G23030 | protein.degradation.ubiquitin : protein.degradation.ubiquitin.E3.RING | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.008G134200 | AT1G24610 | not assigned.no ontology.SET domain-containing protein | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.008G137600 | AT2G25430 | not assigned.no ontology.epsin N-terminal homology (ENTH) domain-containing protein | 12 | 3 | 2 | 6 | 4 | 0 | 0 | 4 |
| Potri.008G146000 | AT1G24040 | misc.GCN5-related N-acetyltransferase | 7 | 3 | 3 | 2 | 2 | 0 | 0 | 2 |
| Potri.008G169000 | AT4G14330 | cell.organisation | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.008G172900 | AT5G05800 | not assigned.unknown | 7 | 3 | 0 | 3 | 2 | 0 | 2 | 2 |
| Potri.008G181700 | AT1G67730 | secondary metabolism.wax | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.008G181900 | AT1G67740 | PS.lightreaction.photosystem II.PSII polypeptide subunits | 10 | 3 | 6 | 3 | 1 | 0 | 0 | 1 |
| Potri.008G182800 | AT1G13170 | cell.vesicle transport | 13 | 3 | 1 | 11 | 1 | 0 | 0 | 1 |
| Potri.008G190600 | AT1G10417 | not assigned.unknown | 9 | 3 | 0 | 5 | 1 | 0 | 3 | 1 |
| Potri.008G197500 | AT5G18860 | nucleotide metabolism.degradation | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.008G220600 | AT4G15920 | development.unspecified | 7 | 3 | 2 | 4 | 1 | 0 | 0 | 1 |
| Potri.009G005900 | AT1G08130 | DNA.synthesis/chromatin structure | 5 | 3 | 1 | 1 | 3 | 0 | 0 | 3 |
| Potri.009G006100 | AT5G04320 | not assigned.unknown | 7 | 3 | 1 | 5 | 1 | 0 | 0 | 1 |
| Potri.009G007400 | AT5G28780 | DNA.unspecified | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.009G016200 | AT2G35930 | RNA.regulation of transcription.PHOR1 | 10 | 3 | 1 | 8 | 1 | 0 | 0 | 1 |
| Potri.009G037000 | AT3G46780 | RNA.transcription | 14 | 3 | 6 | 7 | 1 | 0 | 0 | 1 |
| Potri.009G051300 | AT3G21090 | transport.ABC transporters and multidrug resistance systems | 14 | 3 | 2 | 11 | 1 | 0 | 0 | 1 |
| Potri.009G053900 | AT5G12870 | RNA.regulation of transcription.MYB domain transcription factor family | 7 | 3 | 2 | 4 | 1 | 0 | 0 | 1 |
| Potri.009G054200 | AT3G12080 | signalling.G-proteins | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.009G055900 | AT3G63470 | protein.degradation.serine protease | 12 | 3 | 2 | 9 | 1 | 0 | 0 | 1 |
| Potri.009G060400 | AT3G07970 | cell wall.degradation.pectate lyases and polygalacturonases | 9 | 3 | 1 | 7 | 1 | 0 | 0 | 1 |
| Potri.009G077800 | AT2G30200 | lipid metabolism.FA synthesis and FA elongation.Acetyl CoA Transacylase | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.009G077900 | AT1G07010 | misc.calcineurin-like phosphoesterase family protein | 9 | 3 | 5 | 3 | 1 | 0 | 0 | 1 |
| Potri.009G087500 | AT4G35270 | RNA.regulation of transcription.NIN-like bZIP-related family | 7 | 3 | 2 | 4 | 1 | 0 | 0 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.009G113700 | AT1G49340 | signalling.phosphinositides | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |
| Potri.009G117000 | AT2G18360 | not assigned.no ontology | 8 | 3 | 4 | 2 | 2 | 0 | 0 | 2 |
| Potri.009G149700 | AT5G22740 | cell wall.cellulose synthesis | 12 | 3 | 5 | 6 | 1 | 0 | 0 | 1 |
| Potri.010G009200 | AT3G05640 | protein.postranslational modification | 6 | 3 | 1 | 2 | 3 | 0 | 0 | 3 |
| Potri.010G014500 | AT5G49030 | protein.aa activation.isoleucine-tRNA ligase | 11 | 3 | 2 | 7 | 2 | 0 | 0 | 2 |
| Potri.010G015500 | AT4G21380 | signalling.receptor kinases.S-locus glycoprotein like | 5 | 3 | 0 | 1 | 1 | 0 | 3 | 1 |
| Potri.010G021200 | AT5G41980 | not assigned.unknown | 8 | 3 | 0 | 6 | 1 | 0 | 1 | 1 |
| Potri.010G027300 | AT5G18910 | protein.postranslational modification.kinase.receptor like cytoplasmatic kinase VI | 12 | 3 | 2 | 9 | 1 | 0 | 0 | 1 |
| Potri.010G034300 | AT1G69850 | transport.nitrate | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.010G042000 | AT1G60470 | minor CHO metabolism.raffinose family.galactinol synthases.putative : minor CHO metabolism.raffinose family.galactinol synthases.known | 9 | 3 | 3 | 5 | 1 | 0 | 0 | 1 |
| Potri.010G048300 | AT3G26380 | minor CHO metabolism.galactose.alpha-galactosidases | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.010G106900 | AT2G34930 | stress.biotic.PR-proteins | 8 | 3 | 0 | 5 | 1 | 0 | 2 | 1 |
| Potri.010G112800 | AT1G70940 | hormone metabolism.auxin.signal transduction | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |
| Potri.010G113000 | AT1G70950 | not assigned.unknown | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.010G119100 | AT2G01170 | transport.amino acids | 16 | 3 | 2 | 13 | 1 | 0 | 0 | 1 |
| Potri.010G120000 | | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.010G125800 | AT1G68560 | misc.gluco-, galacto- and mannosidases.alpha-galactosidase | 12 | 3 | 0 | 10 | 1 | 0 | 1 | 1 |
| Potri.010G130800 | AT3G25690 | not assigned.no ontology.hydroxyproline rich proteins | 8 | 3 | 4 | 3 | 1 | 0 | 0 | 1 |
| Potri.010G141000 | AT5G49330 | RNA.regulation of transcription.MYB domain transcription factor family | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |
| Potri.010G153700 | AT3G53480 | transport.ABC transporters and multidrug resistance systems | 4 | 3 | 0 | 2 | 1 | 0 | 1 | 1 |
| Potri.010G153800 | AT4G15230 | transport.ABC transporters and multidrug resistance systems | 17 | 3 | 2 | 14 | 1 | 0 | 0 | 1 |
| Potri.010G155600 | AT1G53440 | signalling.receptor kinases.leucine rich repeat VIII.VIII-2 | 24 | 3 | 0 | 1 | 1 | 0 | 22 | 1 |
| Potri.010G161400 | AT2G03420 | not assigned.unknown | 9 | 3 | 6 | 2 | 1 | 0 | 0 | 1 |
| Potri.010G169000 | AT3G28960 | transport.amino acids | 8 | 3 | 2 | 5 | 1 | 0 | 0 | 1 |
| Potri.010G170300 | AT3G19830 | not assigned.no ontology.C2 domain-containing protein | 6 | 3 | 3 | 0 | 2 | 0 | 1 | 2 |
| Potri.010G171300 | AT1G79560 | protein.degradation.metalloprotease | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.010G199900 | AT3G11420 | not assigned.no ontology | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.010G210000 | AT2G39470 | PS.lightreaction.photosystem II.PSII polypeptide subunits | 7 | 3 | 5 | 1 | 1 | 0 | 0 | 1 |
| Potri.010G230900 | AT5G04440 | not assigned.unknown | 7 | 3 | 5 | 1 | 1 | 0 | 0 | 1 |
| Potri.010G254700 | AT4G37930, AT5G26780 | PS.photorespiration.serine hydroxymethyltransferase | 9 | 3 | 6 | 0 | 1 | 0 | 2 | 1 |
| Potri.011G025300 | AT2G33460 | signalling.G-proteins : stress.biotic.PR-proteins | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |
| Potri.011G043500 | AT4G21540 | lipid metabolism.Phospholipid synthesis.diacylglycerol kinase : lipid metabolism.exotics (steroids, squalene etc).sphingolipids | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.011G057400 | | not assigned.unknown | 11 | 3 | 1 | 8 | 2 | 0 | 0 | 2 |
| Potri.011G061500 | AT5G45970 | signalling.G-proteins | 7 | 3 | 5 | 1 | 1 | 0 | 0 | 1 |
| Potri.011G066200 | AT1G78830 | misc.myrosinases-lectin-jacalin | 7 | 3 | 3 | 3 | 1 | 0 | 0 | 1 |
| Potri.011G066900 | AT1G09850 | protein.degradation.cysteine protease | 8 | 3 | 2 | 0 | 1 | 0 | 5 | 1 |
| Potri.011G069300 | AT4G18740 | not assigned.unknown | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.011G076700 | AT5G45650 | protein.degradation.subtilases | 10 | 3 | 3 | 6 | 1 | 0 | 0 | 1 |
| Potri.011G086400 | AT1G55930 | not assigned.no ontology | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.011G106300 | AT3G14850 | not assigned.unknown | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.011G116900 | AT5G53890 | signalling.receptor kinases.leucine rich repeat X | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.011G124400 | AT4G27220 | stress.biotic | 10 | 3 | 2 | 7 | 1 | 0 | 0 | 1 |
| Potri.011G129300 | AT4G27290 | signalling.receptor kinases.S-locus glycoprotein like | 6 | 3 | 0 | 2 | 1 | 0 | 3 | 1 |
| Potri.011G136300 | AT5G54840 | signalling.G-proteins | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.011G142200 | AT1G79040 | PS.lightreaction.photosystem II.PSII polypeptide subunits | 16 | 3 | 7 | 8 | 1 | 0 | 0 | 1 |
| Potri.011G156200 | AT1G06620, AT1G06650 | secondary metabolism.sulfur-containing.glucosinolates.synthesis.aliphatic.2-oxoglutarate-dependent dioxygenase : redox.ascorbate and glutathione : hormone metabolism.ethylene.synthesis-degradation | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.011G158800 | AT2G34790 | misc.nitrilases, *nitrile lyases, berberine bridge enzymes, reticuline oxidases, troponine reductases | 11 | 3 | 0 | 9 | 1 | 0 | 1 | 1 |
| Potri.011G162700 | AT4G20820 | misc.nitrilases, *nitrile lyases, berberine bridge enzymes, reticuline oxidases, troponine reductases | 12 | 3 | 3 | 7 | 2 | 0 | 0 | 2 |
| Potri.012G020600 | AT1G71400 | stress.biotic.PR-proteins : signalling.receptor kinases.misc : stress.biotic : stress.biotic.kinases | 11 | 3 | 0 | 7 | 2 | 0 | 2 | 2 |
| Potri.012G025700 | AT4G13810 | signalling.receptor kinases.misc : stress.biotic.PR-proteins : stress.biotic | 10 | 3 | 0 | 5 | 1 | 0 | 4 | 1 |
| Potri.012G026300 | | not assigned.unknown | 6 | 3 | 0 | 4 | 1 | 0 | 1 | 1 |
| Potri.012G032700 | AT1G74100 | secondary metabolism.sulfur-containing.glucosinolates.synthesis.aliphatic.sulfotransferase : misc.sulfotransferase : secondary metabolism.sulfur-containing.glucosinolates.synthesis.indole.indole-3-methyl-desulfoglucosinolate sulfotransferase | 5 | 3 | 1 | 2 | 2 | 0 | 0 | 2 |
| Potri.012G060300 | AT1G49010 | RNA.regulation of transcription.MYB-related transcription factor family | 6 | 3 | 4 | 1 | 0 | 1 | 0 | 1 |
| Potri.012G062300 | AT5G60910 | RNA.regulation of transcription.MADS box transcription factor family : development.unspecified | 5 | 3 | 0 | 3 | 1 | 1 | 0 | 2 |
| Potri.012G062700 | AT3G25800 | protein.postranslational modification | 7 | 3 | 0 | 3 | 3 | 1 | 0 | 4 |
| Potri.012G067300 | | not assigned.unknown | 9 | 4 | 1 | 6 | 1 | 1 | 0 | 2 |
| Potri.012G070700 | AT1G69850 | transport.nitrate : transport.peptides and oligopeptides | 13 | 3 | 3 | 9 | 1 | 0 | 0 | 1 |
| Potri.012G103500 | AT5G13180 | RNA.regulation of transcription.NAC domain transcription factor family : development.unspecified | 9 | 3 | 3 | 5 | 1 | 0 | 0 | 1 |
| Potri.012G107000 | AT1G33420 | not assigned.unknown | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.012G129500 | AT5G62260 | RNA.regulation of transcription.putative transcription regulator | 8 | 3 | 6 | 1 | 1 | 0 | 0 | 1 |
| Potri.012G129800 | AT4G25370 | protein.targeting.unknown | 4 | 3 | 2 | 1 | 0 | 1 | 0 | 1 |
| Potri.013G011700 | AT5G02070 | signalling.receptor kinases.wall associated kinase | 6 | 3 | 2 | 3 | 1 | 0 | 0 | 1 |
| Potri.013G026600 | AT3G05030 | transport.unspecified cations | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.013G055500 | AT5G17980 | not assigned.no ontology.C2 domain-containing protein | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |
| Potri.013G060200 | AT3G04030 | RNA.regulation of transcription.G2-like transcription factor family, GARP : RNA.regulation of transcription.MYB domain transcription factor family | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.013G073900 | AT5G35970 | DNA.unspecified | 10 | 3 | 1 | 8 | 1 | 0 | 0 | 1 |
| Potri.013G075400 | AT1G64860 | RNA.regulation of transcription.sigma like plant | 10 | 3 | 4 | 5 | 1 | 0 | 0 | 1 |
| Potri.013G084300 | AT2G43945 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.013G091200 | AT1G08580 | not assigned.unknown | 10 | 3 | 0 | 2 | 1 | 0 | 7 | 1 |
| Potri.013G092400 | AT4G10350 | development.unspecified | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.013G097200 | AT5G17680 | stress.biotic.PR-proteins : stress.biotic | 13 | 3 | 0 | 7 | 1 | 0 | 5 | 1 |
| Potri.013G100500 | AT4G32480 | not assigned.unknown | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.013G102600 | AT1G71692 | RNA.regulation of transcription.MADS box transcription factor family | 11 | 3 | 4 | 4 | 3 | 0 | 0 | 3 |
| Potri.013G103900 | AT1G75280 | secondary metabolism.flavonoids.isoflavones.isoflavone reductase | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.013G111900 | AT1G71960 | transport.ABC transporters and multidrug resistance systems | 10 | 3 | 3 | 5 | 2 | 0 | 0 | 2 |
| Potri.013G113500 | AT4G09620 | not assigned.unknown | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.013G118000 | AT2G31130 | not assigned.unknown | 7 | 3 | 1 | 4 | 2 | 0 | 0 | 2 |
| Potri.013G124900 | AT1G73280, AT5G09640 | protein.degradation.serine protease | 7 | 3 | 1 | 4 | 2 | 0 | 0 | 2 |
| Potri.013G145700 | AT2G20830 | not assigned.no ontology | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G000800 | AT3G13050 | transport.misc | 11 | 3 | 0 | 7 | 1 | 0 | 3 | 1 |
| Potri.014G002300 | AT3G14470 | stress.biotic.PR-proteins | 5 | 3 | 3 | 0 | 1 | 0 | 1 | 1 |
| Potri.014G009600 | AT3G14470 | stress.biotic.PR-proteins | 4 | 3 | 0 | 2 | 1 | 0 | 1 | 1 |
| Potri.014G010700 | AT3G14470 | stress.biotic.PR-proteins | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.014G036200 | AT1G27040 | transport.peptides and oligopeptides : transport.nitrate | 9 | 3 | 0 | 1 | 1 | 0 | 7 | 1 |
| Potri.014G086100 | AT2G45990 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G088300 | AT2G22590 | secondary metabolism.flavonoids.anthocyanins : secondary metabolism.flavonoids.anthocyanins.anthocyanidin 3-O-glucosyltransferase | 8 | 3 | 5 | 0 | 1 | 0 | 2 | 1 |
| Potri.014G120700 | AT3G62410 | PS.calvin cycle | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G121000 | AT4G02060 | DNA.synthesis/chromatin structure | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |
| Potri.014G122000 | AT3G62550 | hormone metabolism.ethylene.induced-regulated-responsive-activated | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G132500 | AT4G13420 | transport.potassium | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.014G138500 | AT2G48120 | development.unspecified | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G143400 | AT2G42750 | stress.abiotic.heat | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G145200 | AT5G46860 | cell.vesicle transport | 6 | 3 | 0 | 3 | 1 | 0 | 2 | 1 |
| Potri.014G148500 | AT3G23700 | not assigned.no ontology.S RNA-binding domain-containing protein | 6 | 3 | 4 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G155300 | AT2G32540 | cell wall.cellulose synthesis.cellulose synthase | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.014G170300 | AT2G04270 | RNA.processing.plastidial RNA.RNE Complex.RNE | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.014G190900 | AT3G07330 | cell wall.cellulose synthesis | 11 | 3 | 4 | 6 | 1 | 0 | 0 | 1 |
| Potri.014G191100 | AT5G48520 | not assigned.unknown | 5 | 3 | 1 | 2 | 2 | 0 | 0 | 2 |
| Potri.015G016900 | AT5G53450 | protein.postranslational modification | 10 | 3 | 0 | 8 | 1 | 0 | 1 | 1 |
| Potri.015G033900 | AT3G17700 | transport.cyclic nucleotide or calcium regulated channels | 5 | 3 | 1 | 0 | 3 | 0 | 1 | 3 |
| Potri.015G034200 | AT2G40540 | transport.potassium | 6 | 3 | 3 | 1 | 2 | 0 | 0 | 2 |
| Potri.015G034300 | | not assigned.unknown | 4 | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Potri.015G034700 | AT5G37600 | N-metabolism.ammonia metabolism.glutamine synthetase | 8 | 3 | 3 | 3 | 2 | 0 | 0 | 2 |
| Potri.015G050200 | AT1G75290 | secondary metabolism.flavonoids.isoflavones.isoflavone reductase | 12 | 3 | 5 | 3 | 4 | 0 | 0 | 4 |
| Potri.015G055600 | AT1G18490 | amino acid metabolism.synthesis.serine-glycine-cysteine group.cysteine | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |
| Potri.015G063300 | AT3G47710 | RNA.regulation of transcription.bHLH,Basic Helix-Loop-Helix family | 4 | 3 | 1 | 2 | 1 | 0 | 0 | 1 |
| Potri.015G063400 | AT3G47740 | transport.ABC transporters and multidrug resistance systems | 7 | 3 | 1 | 5 | 1 | 0 | 0 | 1 |
| Potri.015G065900 | AT5G55220 | protein.folding | 10 | 3 | 3 | 5 | 2 | 0 | 0 | 2 |
| Potri.015G089700 | AT3G48460 | misc.GDSL-motif lipase | 5 | 3 | 1 | 2 | 2 | 0 | 0 | 2 |
| Potri.015G108800 | AT5G61820 | not assigned.unknown | 10 | 3 | 4 | 4 | 2 | 0 | 0 | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.015G127100 | AT1G62660 | major CHO metabolism.degradation.sucrose.invertases.vacuolar | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.015G127200 | AT4G25240 | misc.oxidases - copper, flavone etc | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.015G127400 | AT1G12260 | development.unspecified | 7 | 3 | 0 | 5 | 1 | 0 | 1 | 1 |
| Potri.015G132200 | AT5G62230,<br>AT5G07180 | signalling.receptor kinases.leucine rich repeat XIII | 12 | 3 | 1 | 10 | 1 | 0 | 0 | 1 |
| Potri.016G017200 | AT3G21750 | hormone metabolism.abscisic acid.synthesis-degradation : misc.UDP glucosyl and glucoronyl transferases | 4 | 3 | 2 | 0 | 1 | 0 | 1 | 1 |
| Potri.016G018700 | AT2G32240 | not assigned.unknown | 9 | 4 | 2 | 5 | 1 | 0 | 1 | 1 |
| Potri.016G020200 | AT1G22360 | hormone metabolism.cytokinin.synthesis-degradation : misc.UDP glucosyl and glucoronyl transferases | 8 | 4 | 1 | 4 | 1 | 0 | 2 | 1 |
| Potri.016G020500 | AT1G22380 | hormone metabolism.cytokinin.synthesis-degradation : misc.UDP glucosyl and glucoronyl transferases | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.016G020700 | AT1G22380 | hormone metabolism.cytokinin.synthesis-degradation : misc.UDP glucosyl and glucoronyl transferases | 5 | 3 | 0 | 2 | 1 | 0 | 2 | 1 |
| Potri.016G020900 | AT1G22380 | misc.UDP glucosyl and glucoronyl transferases : hormone metabolism.cytokinin.synthesis-degradation | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.016G051100 | AT3G57800 | RNA.regulation of transcription.bHLH,Basic Helix-Loop-Helix family | 8 | 3 | 1 | 5 | 2 | 0 | 0 | 2 |
| Potri.016G061500 | AT1G53440 | signalling.receptor kinases.leucine rich repeat VIII.VIII-2 | 30 | 3 | 0 | 1 | 1 | 0 | 28 | 1 |
| Potri.016G075800 | AT1G12910 | development.unspecified | 6 | 3 | 2 | 2 | 2 | 0 | 0 | 2 |
| Potri.016G078400 | AT3G11470 | not assigned.no ontology | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.016G078600 | AT5G03940 | protein.targeting.chloroplast | 7 | 4 | 2 | 1 | 3 | 0 | 1 | 3 |
| Potri.016G080200 | AT3G53540 | not assigned.unknown | 7 | 3 | 1 | 5 | 1 | 0 | 0 | 1 |
| Potri.016G088500 | AT5G03150 | RNA.regulation of transcription.C2H2 zinc finger family | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.016G095100 | AT3G53710 | RNA.regulation of transcription.unclassified : protein.postranslational modification | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.016G125500 | AT2G38320 | not assigned.unknown | 11 | 3 | 3 | 7 | 1 | 0 | 0 | 1 |
| Potri.016G140100 | AT1G06840 | signalling.receptor kinases.leucine rich repeat VIII.VIII-1 | 6 | 3 | 1 | 0 | 1 | 0 | 4 | 1 |
| Potri.017G007900 | AT5G38260 | lipid metabolism.lipid degradation.lysophospholipases.glycerophosphodiester phosphodiesterase : signalling.receptor kinases.Catharanthus roseus-like RLK1 : signalling.receptor kinases.wheat LRK10 like : signalling.receptor kinases.thaumatin like | 21 | 3 | 0 | 1 | 1 | 0 | 19 | 1 |
| Potri.017G043100 | AT4G33630 | not assigned.unknown | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.017G053900 | AT1G15800 | not assigned.unknown | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.017G056400 | AT3G07550 | protein.degradation.ubiquitin.E3.SCF.FBOX | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.017G090700 | AT5G15460 | secondary metabolism.isoprenoids.non-mevalonate pathway | 8 | 3 | 1 | 6 | 1 | 0 | 0 | 1 |
| Potri.017G098000 | | not assigned.unknown | 12 | 3 | 4 | 0 | 1 | 0 | 7 | 1 |
| Potri.017G113200 | AT5G16040 | cell.division | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.017G130000 | AT5G37710 | signalling.calcium | 5 | 3 | 3 | 1 | 1 | 0 | 0 | 1 |
| Potri.017G137600 | AT5G16560 | RNA.regulation of transcription.G2-like transcription factor family, GARP | 12 | 3 | 0 | 3 | 1 | 0 | 8 | 1 |
| Potri.018G005300 | AT1G22640,<br>AT5G14750 | RNA.regulation of transcription.MYB domain transcription factor family : secondary metabolism.sulfur-containing.glucosinolates.regulation.indole : signalling.receptor kinases.misc | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 1 |
| Potri.018G010700 | AT2G29120 | signalling.in sugar and nutrient physiology | 4 | 3 | 1 | 0 | 1 | 0 | 2 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.018G013100 | AT2G29120 | signalling.in sugar and nutrient physiology | 4 | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Potri.018G015100 | AT5G10770 | RNA.regulation of transcription.unclassified | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.018G017200 | AT5G25060 | RNA.RNA binding | 8 | 3 | 0 | 6 | 1 | 0 | 1 | 1 |
| Potri.018G017800 | AT4G31860 | protein.postranslational modification | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.018G062900 | AT1G74960 | lipid metabolism.FA synthesis and FA elongation.ketoacyl ACP synthase | 4 | 3 | 1 | 1 | 2 | 0 | 0 | 2 |
| Potri.018G083400 | AT4G28950 | signalling.G-proteins | 9 | 3 | 5 | 2 | 2 | 0 | 0 | 2 |
| Potri.018G083600 | AT3G06880 | not assigned.no ontology | 11 | 3 | 1 | 9 | 1 | 0 | 0 | 1 |
| Potri.018G086200 | AT5G19680 | protein.postranslational modification | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.018G086300 | AT2G20000 | cell.division | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.018G090300 | AT2G18960 | transport.p- and v-ATPases : transport.p- and v-ATPases.H+-exporting ATPase | 5 | 3 | 2 | 1 | 2 | 0 | 0 | 2 |
| Potri.018G097500 | AT2G18710 | protein.targeting.chloroplast | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.018G102500 | AT4G28080 | not assigned.unknown | 5 | 3 | 2 | 2 | 1 | 0 | 0 | 1 |
| Potri.018G105600 | AT2G24020 | not assigned.unknown | 9 | 3 | 4 | 0 | 3 | 0 | 2 | 3 |
| Potri.018G105700 | AT4G30610 | protein.degradation.serine protease | 13 | 3 | 0 | 10 | 2 | 0 | 1 | 2 |
| Potri.018G113200 | AT3G18760 | protein.synthesis.ribosomal protein.prokaryotic.unknown organellar.30S subunit.S6 | 121 | 3 | 0 | 1 | 1 | 0 | 119 | 1 |
| Potri.018G113400 | AT5G35170 | nucleotide metabolism.phosphotransfer and pyrophosphatases.adenylate kinase | 12 | 3 | 5 | 5 | 2 | 0 | 0 | 2 |
| Potri.018G122100 | | not assigned.unknown | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.018G144700 | AT1G74190 | signalling.receptor kinases.misc : stress.biotic | 9 | 3 | 0 | 1 | 1 | 0 | 7 | 1 |
| Potri.018G148100 | AT4G04940 | protein.synthesis.ribosome biogenesis.Pre-rRNA processing and modifications.WD-repeat proteins | 4 | 3 | 2 | 0 | 1 | 0 | 1 | 1 |
| Potri.018G148300 | AT1G16120 | signalling.receptor kinases.wall associated kinase | 15 | 3 | 0 | 8 | 1 | 0 | 6 | 1 |
| Potri.019G012200 | AT1G08800 | not assigned.no ontology | 8 | 3 | 2 | 5 | 1 | 0 | 0 | 1 |
| Potri.019G013400 | ATCG00820 | protein.synthesis.ribosomal protein.prokaryotic.chloroplast.30S subunit.S19 | 4 | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Potri.019G024800 | AT5G18430,<br>AT5G33370 | misc.GDSL-motif lipase | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.019G033200 | AT2G23840 | DNA.synthesis/chromatin structure | 9 | 3 | 2 | 6 | 1 | 0 | 0 | 1 |
| Potri.019G045800 | AT3G03630 | amino acid metabolism.synthesis.serine-glycine-cysteine group.cysteine.OASTL | 11 | 3 | 0 | 9 | 1 | 0 | 1 | 1 |
| Potri.019G050300 | | not assigned.unknown | 5 | 3 | 0 | 1 | 2 | 0 | 2 | 2 |
| Potri.019G054400 | AT5G16390 | lipid metabolism.FA synthesis and FA elongation.Acetyl CoA Carboxylation.heteromeric Complex.Biotin Carboxyl Carrier Protein | 6 | 3 | 2 | 3 | 1 | 0 | 0 | 1 |
| Potri.019G057200 | AT1G09795 | amino acid metabolism.synthesis.histidine.ATP phosphoribosyl transferase | 7 | 3 | 3 | 2 | 2 | 0 | 0 | 2 |
| Potri.019G069300 | AT1G71380 | misc.gluco-, galacto- and mannosidases.endoglucanase : cell wall.degradation.cellulases and beta -1,4-glucanases | 6 | 3 | 1 | 4 | 1 | 0 | 0 | 1 |
| Potri.019G070100 | AT4G12010 | stress.biotic.PR-proteins | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| Potri.019G076600 | AT1G71691 | misc.GDSL-motif lipase | 5 | 3 | 1 | 3 | 1 | 0 | 0 | 1 |
| Potri.019G097800 | AT5G17680 | stress.biotic.receptors : stress.biotic : stress.biotic.PR-proteins | 14 | 3 | 0 | 3 | 1 | 0 | 10 | 1 |
| Potri.019G098000 | AT3G59780 | not assigned.unknown | 12 | 3 | 0 | 1 | 2 | 0 | 9 | 2 |
| Potri.019G103000 | AT4G03420 | not assigned.unknown | 7 | 3 | 0 | 4 | 1 | 0 | 2 | 1 |
| Potri.019G107900 | | not assigned.unknown | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Potri.019G109900 | AT4G29990 | signalling.receptor kinases.leucine rich repeat I : signalling.receptor kinases.misc : protein.postranslational modification | 16 | 3 | 0 | 2 | 1 | 0 | 13 | 1 |
| Potri.T002600 | AT5G36930 | stress.biotic.PR-proteins | 5 | 3 | 0 | 2 | 1 | 0 | 2 | 1 |
| Potri.T003000 | AT5G36930 | stress.biotic.PR-proteins | 3 | 3 | 1 | 0 | 1 | 0 | 1 | 1 |
| Potri.T125200 | AT2G16130, AT3G18830 | transport.sugars | 6 | 3 | 0 | 2 | 2 | 0 | 2 | 2 |
| Potri.T010100 | AT1G29740 | signalling.receptor kinases.leucine rich repeat VIII.VIII-2 | 4 | 3 | 0 | 1 | 2 | 0 | 1 | 2 |
| Potri.T015200 | AT4G27220 | stress.biotic | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.T023400 | AT4G27290 | signalling.receptor kinases.S-locus glycoprotein like | 14 | 3 | 4 | 9 | 1 | 0 | 0 | 1 |
| Potri.T024600 | AT5G23530 | Biodegradation of Xenobiotics | 5 | 3 | 3 | 0 | 1 | 0 | 1 | 1 |
| Potri.T032600 | AT1G29740 | signalling.receptor kinases.leucine rich repeat VIII.VIII-2 | 7 | 3 | 1 | 0 | 1 | 0 | 5 | 1 |
| Potri.T053100 | | not assigned.unknown | 7 | 3 | 2 | 0 | 1 | 0 | 4 | 1 |
| Potri.T055300 | AT4G13440 | signalling.calcium | 7 | 3 | 3 | 0 | 1 | 0 | 3 | 1 |
| Potri.T059900 | AT5G37478 | not assigned.unknown | 13 | 3 | 1 | 11 | 1 | 0 | 0 | 1 |
| Potri.T061600 | | not assigned.unknown | 5 | 3 | 0 | 3 | 1 | 0 | 1 | 1 |
| Potri.T064000 | AT5G38210 | signalling.receptor kinases.wheat LRK10 like | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.T070100 | AT5G17920 | amino acid metabolism.synthesis.aspartate family.methionine | 8 | 3 | 5 | 0 | 2 | 0 | 1 | 2 |
| Potri.T077100 | AT5G17680 | stress.biotic.PR-proteins | 11 | 3 | 0 | 5 | 1 | 0 | 5 | 1 |
| Potri.T078500 | AT3G42170 | DNA.unspecified | 3 | 3 | 0 | 1 | 1 | 0 | 1 | 1 |
| Potri.T167300 | AT1G64940 | misc.cytochrome P450 | 6 | 3 | 0 | 3 | 1 | 0 | 2 | 1 |