Title: Inferring biosynthetic and gene regulatory networks from *Artemisia annua* RNA sequencing data on a credit card-sized ARM computer

Authors: Qiao Wen Tan[1], Marek Mutwil[1*]

[1]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

***Corresponding author:**

Marek Mutwil

School of Biological Sciences,

Nanyang Technological University,

60 Nanyang Drive,

637551, Singapore,

Singapore

Email: mutwil@ntu.edu.sg

Highlights

- Processing of large scale transcriptomic data with affordable single-board computers
- Transcription factors can be found in the same network as their targets
- Co-expression of transcription factors and genes in secondary cell wall biosynthesis
- Co-expression of transcription factors and genes involved in artemisinin biosynthesis

0. ABSTRACT

Prediction of gene function and gene regulatory networks is one of the most active topics in bioinformatics. The accumulation of publicly available gene expression data for hundreds of plant species, together with advances in bioinformatical methods and affordable computing, sets ingenuity as one of the major bottlenecks in understanding gene function and regulation. Here, we show how a credit card-sized computer retailing for less than 50 USD can be used to rapidly predict gene function and infer regulatory networks from RNA sequencing data. To achieve this, we constructed a bioinformatical pipeline that downloads and allows quality-control of RNA sequencing data; and generates a gene co-expression network that can reveal enzymes and transcription factors participating and controlling a given biosynthetic pathway. We exemplify this by first identifying genes and transcription factors involved in the biosynthesis of secondary cell wall in the plant *Artemisia annua*, the main natural source of the anti-malarial drug artemisinin. Networks were then used to dissect the artemisinin biosynthesis pathway, which suggest potential transcription factors regulating artemisinin biosynthesis. We provide the source code of our pipeline (https://github.com/mutwil/LSTrAP-Lite) and envision that the ubiquity of affordable computing, availability of biological data and increased bioinformatical training of biologists will transform the field of bioinformatics.

0.KEYWORDS

Co-expression, artemisinin, Artemisia, RNA sequencing, single-board computer

1. INTRODUCTION

To extract useful knowledge from rapidly accumulating genomic information, we are dependent on our capacity to correctly assign biological functions to gene products. Since gene products can have multiple functions, genetic redundancy caused by large gene families can obscure knock-out phenotypes [1] and genetic transformation of plants takes months [2], functional characterization of a plant gene can take years. This explains why, despite decades of intensive research, only 12 % of genes in Arabidopsis thaliana (3382 out of 28392 genes) are functionally characterised as of 2018 [3].

To facilitate functional characterisation, bioinformatical gene function prediction can help experimentalists by (i) identifying novel genes relevant for the biological process studied and (ii) by suggesting relevant experimental approaches to dissect the function of an unknown gene. For example, several studies used *in silico* predictions to identify genes involved in seed germination [4], plant viability [5], cyclic electron flow [6], cell division [7], shade avoidance [8], drought sensitivity and lateral root development [9]. Consequently, gene function prediction is one of the most intensive bioinformatics research efforts [3,10–12].

Apart from predicting gene function, a substantial effort in bioinformatics is given to identifying transcription regulatory networks, especially transcription factor (TF) - target associations. TFs can regulate their target genes by binding to short DNA sequences called TF binding sites (TFBSs), or through other interaction partners such as TF complexes and other proteins that interact with the TFBS [13]. Prediction of TFBS has been made possible with the massive amounts of TFBS data available from chromatin immunoprecipitation, protein-binding microarrays and others (reviewed in [14]). Typically, binding motifs can be predicted using positional weight matrices of nucleotide sequences. The prediction can be further improved when other factors, such as DNA shape and dinucleotide dependencies are taken into consideration [15,16]. The full set of regulatory interactions between a TF and its target genes forms a gene regulatory network (GRN). Since gene expression regulation is fundamental to all life, unravelling the GRN is pivotal to understand how different biological processes such as development, growth and stress responses are controlled.

GRNs are typically elucidated by yeast one-hybrid (Y1H), which tests for binding of a transcription factor to a promoter. Y1H was used to unravel the root-specific GRN in Arabidopsis [17], secondary cell wall synthesis [18], regulators for SHORTROOT-SCARECROW [19] and others. Further experimental methods include ChIP-chip or ChIP-Seq methods that determine TF binding to cis-elements *in vivo* [20], open chromatin profiling by DNAse I hypersensitivity and ATAC-seq [21], protein binding microarrays [21], DNA affinity purification assays [22] and others. Gene expression data capturing global gene expression changes upon transcription factor knockout or overexpression [23] is invaluable in predicting GRNs but largely unavailable for plants. Experimental inference of GRNs is further compounded by the fact that TFs can form heterodimers and dynamically expand their targets [24]. While invaluable, the experimental methods to uncover GRNs are labor intensive and might not reveal all regulatory interactions [25]. In these cases, computational predictions have proven highly efficient in supporting the elucidation of GRNs. For example, by integrating regulatory interactions from publicly available data, AtRegNet (https://agris-knowledgebase.org/, [26]) allows the visualisation of complex networks formed by TFs and their target genes. Other approaches take advantage of cross-species or cross-ecotype conservation of non-coding regulatory sequences in promoters [27–30]. These approaches can be further augmented by integrating gene expression data to identify transcriptionally co-regulated (co-expressed) genes, that are often functionally related (http://bioinformatics.psb.ugent.be/webtools/TF2Network/, [31–33]).

In general, the gene function and GRN prediction methods can be classified into three “generations”: (i) single, (ii) integrative and (iii) ensemble [32]. Regardless of the method, these approaches are based on the guilt-by-association principle, where genes with similar features are assumed to have the same function [12,34,35]. First generation methods use a single data type to predict gene function or GRN. For example, protein-protein interaction data can reveal

genes that work together in the same protein complex [36], while analysis of gene expression can reveal co-expressed genes [12,31,32]. Second generation methods integrate multiple data types, thus increasing the coverage of the functional associations [12] and decreasing false positives [37–39]. Third-generation ensemble (also called community) methods integrate the predictions of many first- and second-generation methods. Integration of 29 predicted GRNs in yeast and *E. coli* produced an ensemble GRN that outperformed the individual methods [23]. In plants, ensemble methods have been applied to predict subcellular localization of proteins (http://suba3.plantenergy.uwa.edu.au/, [40]), GRNs [41] and gene function [3].

Co-expression networks can be used to infer functional and regulatory associations between genes [42–48], and we were able to successfully use these methods in our studies on secondary cell wall biosynthesis pathways [31,32,46,49]. This shows that GRNs can be elucidated by co-expression analyses for biological processes and pathways if they are under strong transcriptional control, such as secondary metabolism.

In this paper, we demonstrate how GRNs can be rapidly inferred on the Rock64, a device worth less than 50 USD, by taking advantage of the recent advances in affordable single board computers (https://store.pine64.org/?product=rock64-media-board-computer) and efficient RNA-seq quantification programs [50]. The Rock64 contains a quad-core ARM-based central processing unit, 4 GB RAM, numerous input/output ports (Figure 1A) and is supported by various Linux-based operating systems.

Next-generation sequencing technologies are currently outpacing Moore's law, which reflects a trend observed in the computer industry that involves the doubling of computing power every two years [51,52]. The cost of resequencing a human genome has dropped to 1000 USD in less than two decades (https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data). Similarly, the sequencing hardware has become more affordable and compact, changing from bulky, expensive devices weighing tens of kilograms and costing >100,000 USD, to portable nanopore-based sequencers weighing <100 grams and costing <1000 USD (https://nanoporetech.com/). The improvements in software needed to process sequencing data have followed similar trends. For example, kallisto and salmon, programs used to estimate gene expression are at least 10 times faster and require less memory than the previous generation of corresponding software [50,53]. Consequently, RNA sequencing analyses that would take weeks on an expensive computer cluster can now be performed in a matter of days on a laptop. These developments prompted us to showcase how to use the Rock64 to process the available RNA sequencing data for *Artemisia annua*, to infer biosynthetic and GRN networks of secondary cell wall and artemisinin biosynthesis.

2. MATERIAL AND METHODS

RNA sequencing experiments for *Artemisia annua* was downloaded as fastq files from European Nucleotide Archive (ENA, [54]) via Aspera v3.5.6.110647. To run Aspera on the ARM CPU of Rock64, we have used x86 CPU emulator ExaGear (https://eltechs.com/product/exagear-desktop/). Using Aspera is optional, but it provides fast download speeds. For paired-end experiments, only the file containing the first read, designated with "_1" was downloaded. The mapping of reads was done using kallisto v0.44.0 [50]. Kallisto index file was generated for *A. annua*

(https://www.ncbi.nlm.nih.gov/assembly/GCA_003112345.1/, [55]) coding sequence file with default parameters. The mapping was done using kallisto quant for single-end library with default parameters, with command 'kallisto quant -i <index> -t 4 -o <output_directory> --single -l 200 -s 20 <fastq>' (fragment length of 200bp and estimated standard deviation of 20). In total, 41 experiments from *A. annua* were included in the co-expression network and annotated based on the information available from NCBI Sequence Read Archive. The Transcripts Per Kilobase Million (TPM) values from the kallisto outputs of selected experiments were represented as an expression matrix where genes were arranged in rows and experiments in columns (Table S1). To annotate the Artemisia CDS, we have used Mercator with standard settings [56]. Gene co-expression networks were calculated using Pearson Correlation Coefficient (PCC, [31]). Significance of enrichment of the cell wall and artemisinin biosynthesis genes in the co-expression neighborhoods was tested using hypergeometric test. Genes involved in secondary cell wall biosynthesis were defined as having the MapMan bin codes starting with "10" while genes involved in artemisinin biosynthesis were identified through blastn v2.6.0+ using artemisinin biosynthetic genes retrieved from NCBI (Supplementary Table S2). The scripts used to perform these analyses are available from https://github.com/mutwil/LSTrAP-Lite and from Supplemental Text 1. Resource usage was monitored using the GNU time module for experiments SRR5714147, SRR6472947, SRR5242542, SRR6808226 and SRR6898431, which are representative of files with various sizes.

3. RESULTS

3.1 Establishing the gene co-expression network pipeline on Rock64 single board computer

With the increased efficiency of transcript quantification softwares [50,53], estimating gene expression from RNA sequencing data is no longer a significant computational bottleneck. A recent blog post demonstrated that the kallisto program, used to estimate gene expression data, can successfully run on a credit card-sized single board computer Rock64 (https://liorpachter.wordpress.com/2018/01/29/bioinformatics-on-a-rock64/, Figure 1A).

We have established a computational pipeline that downloads the raw RNA-sequencing data from a public repository and constructs a gene expression matrix from the specified samples (Figure 1B). After compiling the expression matrix, the pipeline reports quality statistics of the individual samples by showing the number of reads that map to the coding sequences. This allows the user to remove samples that show poor mapping of the RNA-sequencing reads to the samples by specifying a list of selected accessions that should be used for downstream analyses. Finally, the pipeline accepts a gene of interest from the user to reveal co-expressed genes (Figure 1B). The co-expression relationships are reported as a co-expression list or a Cytoscape-compatible co-expression network [57].

To investigate the suitability of this single board computer in generating co-expression networks, we set to investigate the gene expression data of *Artemisia annua* [55], a plant that is cultivated globally as the main natural source of the potent anti-malarial compound, artemisinin. We used Rock64 to download and process all of the available RNA-seq accessions found in the European Nucleotide Archive and noted that most of the accession files are around 2 GB in size (Table S1). By using Aspera download client, we achieved average download speeds of 157

seconds per gigabyte (Figure 1C), with kallisto mapping taking 270 seconds per gigabyte. In addition, we tracked the resource usage of 5 experiments of various sizes and observed that performance of aspera (average 20% CPU and average max RAM of 91 Mb) and kallisto (average 349% CPU and average maximum RAM of 1797 Mb) jobs were similar regardless of file size (Table S3). Overall, it took 4 hours to download and 7 hours to map 42 samples, comprising 94 gigabytes of RNA-sequencing data (Table S1), showing that Rock64 is a viable platform to perform RNA-sequencing analyses. Analysis of the mapping statistics revealed one sample, SRR6472949, to map relatively poorly compared to the majority and was excluded from the analysis. The gene annotation, MapMan bins and gene expression matrix is available in Table S4.

3.2 Elucidating the gene regulatory network for secondary cell wall biosynthesis

Secondary cell walls (SCW) provide mechanical support, water and nutrient transport and stress management in the vascular plant lineage. Since they also are an abundant resource of renewable feed, fiber, and fuel, a significant effort is directed to understand how SCWs are formed [58]. SCWs consist mainly of cellulose microfibrils and other polysaccharides such as hemicelluloses and pectins. In contrast to other cell wall types, SCWs contain lignin that makes the cell walls more rigid and less permeable to water [59]. Due to the importance of SCWs as a renewable resource, the biosynthetic and regulatory networks behind have received much attention and are well understood (Figure 2A). The genes involved in biosynthesis and regulation of SCWs have been shown to be co-expressed in multiple plant species, including gymnosperms, monocots and dicots [32,49,60]. The transcription factors (TFs) controlling SCW are secondary wall NACs (SWNs), and are top-level master switches capable of inducing SCW formation ectopically [38]. The SWNs in turn regulate the expression and activity of several MYB transcription factors, that in turn activate expression of genes important for xylan, cellulose and lignin biosynthesis ([58], Figure 2A).

To demonstrate how co-expression networks can be used to reveal functionally related genes and GRN, we used our pipeline to retrieve 50 most co-expressed genes with *PWA69565.1*, a SCW-specific cellulose synthase 7 ortholog from *Artemisia* (Figure 2B, Figure S1,Table S5). The network revealed functionally related genes that are known participants of cellulose biosynthesis (CESAs [*PWA71152.1*, *PWA69565.1*, *PWA38807.1*], COBRAs [*PWA87916.1*, *PWA76896.1*] and TBL3 [*PWA59244.1*]) [61,62], deposition of cellulose microfibrils (FLA11 [*PWA48845.1*] and TPX2 [*PWA54862.1*]) [63,64], lignin biosynthesis (PAL2 [*PWA82924.1*], laccases [*PWA97538.1*], CYP450 [*PWA58824.1*]) [65,66], pectin biosynthesis (UGD2 [*PWA99088.1*]) [67], xylan biosynthesis (IRX9 [*PWA64911.1*], PGSIP1 [*PWA94002.1*], DUF579 [*PWA92498.1*, *PWA41493.1*, *PWA57753.1*], TBL29 [*PWA55037.1*], GAUT12 [*PWA83971.1*], TBL33 [*PWA77031.1*] [68–73] and cytoskeleton (MAPs [*PWA98831.1*, *PWA53461.1*]). Enrichment analysis revealed that genes involved in cell wall biosynthesis are significantly enriched (hypergeometric test $P<0.001$) [65].

Furthermore, several well studied TFs known for regulating various aspects of SCW formation were identified, such as NST1 (*PWA92922.1*), a master regulator of SCW biosynthesis; MYB83, MYB42 and MYB85 (*PWA56935.1*, *PWA93789.1* and *PWA65860.1* respectively) [58], positive regulators of SCW biosynthesis and LBD15 (*PWA65001.1*), a negative regulator of cellulose biosynthesis [74]. In addition to the conventional TFs associated

with SCW, we also observe other TFs that may be involved in SCW in our network. For example, BLH7 was observed to be involved in the biosynthesis of cellulose-rich tension wood in *Populus* [75]. Suberin biosynthesis in the exocarp of cell wall of apples was proposed to be regulated by MYB93 [76]. Lastly, we also found MYB36, a TF associated with casparian strip formation, which consists of lignified primary cell walls [77].

Taken together, these results show that co-expression networks can find the biosynthetic genes and their regulators.

3.3 Elucidating the gene regulatory network for artemisinin biosynthesis

Malaria is one of the most deadly endemic infectious diseases. The World Health Organization reports that >1 billion people are at risk of malaria, and despite the recent efforts in the last decade, approximately 435,000 deaths were caused by malaria in 2017 [78,79]. Malaria is caused by microscopic parasites of the *Plasmodium* genus, among which *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi* are capable of infecting humans [80,81]. Unfortunately, *Plasmodium has* gained resistance to most medicines, such as quinine and chloroquine [80,82,83]. However, artemisinin from *Artemisia annua* is still an effective drug to treat malaria [84,85].

Artemisinin is an unusual endoperoxide sesquiterpene lactone [86] which is biosynthesized in the cytosol from isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP, [87]). The biosynthetic network surrounding artemisinin also produces other valuable precursors, such as artemisinic acid and arteannuin B (Figure 3A, [88]). While the pathway is well understood and biosynthesis of artemisinin has been established in yeast [89], we hypothesize that other components, such as chaperones, co-factors or scaffolds could increase the biosynthesis of artemisinin [90].

To reveal the components involved in the biosynthesis of these high-value compounds, we used *PWA62882.1*, an amorpha-4,11-diene synthase (ADS) to retrieve its co-expression neighborhood (Figure 3B, Figure S2A, Table S6). The expression of *PWA62882.1* peaks in the bud and young leaf (Figure 3D). In the neighbourhood of *PWA62882.1*, we observe the presence of most genes involved in artemisinin biosynthesis [ADS (*PWA62882.1*, *PWA56512.1*, *PWA61897.1*), cytochrome P450-like (*PWA69692.1*), cytochrome P450 mono-oxygenase CYP71AV1 (*PWA40082.1*), alcohol dehydrogenase 1 (*PWA91359.1*) and aldehyde dehydrogenase 1 (*PWA96689.1*)], except for artemisinic aldehyde delta-11(13) reductase (DBR2) [91]. Enrichment analysis revealed that genes involved in artemisinin biosynthesis are significantly enriched in this neighborhood (hypergeometric test $P<0.001$). Notably, three ATP-binding cassette (ABC) transporters (*PWA91382.1*, *PWA86185.1*, *PWA34723.1*) are also found in the neighbourhood of *PWA62882.1*, making them interesting candidates for future optimisation of artemisinin biosynthesis as ABC transporters have been found to have a significant impact on artemisinin biosynthesis [92].

To date, various AP2/ERF, bHLH, WRKY and bZIP transcription factors have been reported to be involved in the positive regulation of artemisinin biosynthesis [92]. In the neighbourhood of *PWA62882.1*, we observe *PWA43096.1* (ORA) and *PWA68925.1*, a putative bZIP transcription factor, which may likely to play a role in the regulation of the artemisinin biosynthetic genes which co-expressed in the neighbourhood. Other transcription factors found

in the network [aintegumenta-like 6 (*PWA61088.1*) [93], ovate family protein (*PWA59375.1*) [94], remorin (*PWA45423.1*) [95] and growth regulating factor (*PWA85344.1*) [96]] are not likely to be directly involved in the regulation of artemisinin biosynthesis but may be associated with their potential involvement in general growth and development of the plant under stress.

For a more complete picture of the genes and regulators involved in artemisinin biosynthesis, we looked at the co-expression neighbourhood of *PWA95606.1* (DBR2), the only gene in the biosynthetic pathway that was not identified in the neighbourhood of *PWA62882.1* (Figure 3C, Figure S2B, Table S7). The expression profile of *PWA95606.1* is similar to *PWA62882.1*, peaking at the bud and young leaf (Figure 3D). As expected, some genes involved in the biosynthesis of artemisinin were found and enriched (hypergeometric test $P<0.001$) in the neighbourhood of *PWA95606.1*, such as ADS (*PWA56513.1*), cytochrome P450 reductases (*PWA74471.1*, *PWA60095.1*) and alcohol dehydrogenase 1 (*PWA39423.1*). In contrast to the *PWA62882.1* neighbourhood, the DBR2 neighbourhood did not seem to associate with any of the well-studied transcription factors involved in the regulation of artemisinin biosynthesis. The transcription factors zinc-finger domain of monoamine-oxidase A repressor R1 (*PWA54654.1*, *PWA68788.1*), zinc finger (*PWA93224.1*), PGK (*PWA55893.1*), homeodomain-containing proteins (*PWA70268.1* and *PWA70038.1*), homeobox-KN domain containing protein (*PWA76380.1*) and cycloidea-like 8 (*PWA72753.1*) found in the neighbourhood of DBR2 also did not show any probable involvement in the regulation of artemisinin biosynthesis based on our current understanding. Thus, as exemplified in the examples described in Figure 3B and 3C, the choice of gene to be used as bait is highly crucial and can sometimes give very different or incomplete perspectives on the biological process of interest.

4. DISCUSSION

Gene function prediction and inference of GRN are in the focal point of bioinformatics, and have been the subject of numerous DREAM challenges, which invite participants to compete in providing the most accurate and complete predictions [23,97]. The comparison of 35 GRN inference methods by the DREAM5 challenge showed that correlation performed comparably to GRNs constructed by regression, mutual information and Bayesian approaches [23,98]. The conclusions from these challenges and other studies is that integration of multiple data types and inference methods typically improves performance of the bioinformatical predictions [3,99–101].

Here, we show that a simple co-expression analysis is able to infer regulators, enzymes and structural proteins involved in secondary cell wall synthesis and artemisinin in *Artemisia* (Figure 2-3). These results show that biosynthetic pathways and their regulators can be readily inferred by this routine analysis [46,60,102–104]. However, it remains unclear whether there is a correlation between the predictability of biosynthetic pathways of secondary metabolism and predictability of the regulatory networks controlling them. Secondary metabolism genes tend to have higher proliferation rates by local gene duplications, are often co-localized in the genome and are co-expressed [105–107], but these observations cannot explain why these genes are often co-expressed with transcription factors controlling them. While these findings suggest that,

in contrast to other processes, secondary metabolism is perhaps under simpler, exclusively transcriptional control, further work is needed to confirm this hypothesis.

Affordable computing and availability of biological data is providing us with tools to generate high quality predictions of gene function and GRNs. The ease and affordability of generating and analyzing new data requires an overhaul of education and training of all researchers, from undergraduates to faculty [108]. These steps are already seen across the globe, with topics such as computational thinking, programming and data science being introduced as compulsory subjects for biology undergraduates. In addition, we envision that the increased power of miniature computers (e.g. cell phones), further development of portable sequencers and more efficient bioinformatics software will bring the genome sequencing technology to every biology lab in a university, research institute or high school.

DATA AVAILABILITY

The expression matrix is found in the supplemental data, while the scripts used on Rock64 are found at https://github.com/mutwil/LSTrAP-Lite

SUPPLEMENTARY DATA

Supplementary Text 1. User manual for LSTrAP-Lite pipeline.

Figure S1. Co-expression network of CESA7 from *Artemisia annua*. Nodes represent genes, while edges connect co-expressed genes. Red nodes indicate genes involved in cell wall biosynthesis, green nodes indicate transcription factors and blue nodes indicate other genes in the network.

Figure S2. Co-expression networks of ADS and DBR2 from *Artemisia annua* A) Co-expression network of ADS (*PWA62882.1*). Red, orange, green and grey nodes indicate enzymes, transcription factors, ABC transporters and other genes associated with the network respectively. Abbreviations for genes include AIL6: aintegumenta-like 6; GRF: growth regulating factor and OFP: ovate family protein. For brevity, only 50 most co-expressed genes of the query gene are shown. B) Co-expression network of DBR2 (*PWA95606.1*). Abbreviations for genes include Znf-4CXXC_R1: zinc-finger domain of monoamine-oxidase A repressor R1; CYC-like 8: cycloidea-like 8 and ZF: zinc finger. For brevity, only 50 most co-expressed genes of the query gene are shown.

Table S1. Sample annotation, download times and kallisto mapping statistics for Artemisia.

Table S2. BLASTN output from query genes involved in artemisinin biosynthesis.

Table S3. Summary of resource usage for 5 RNA-seq experiments.

Table S4. Gene annotation and expression matrix of *Artemisia annua*. The first, second and third columns contain gene identifiers, gene annotations from Mercator and MapMan identifiers, respectively. The expression profiles start from the 4th column.

Table S5 Table of top 50 co-expressed genes with the *PWA68565.1* (*CESA7*).

Table S6. Table of top 50 co-expressed genes with *PWA62882.1* (*ADS*).

Table S7. Table of top 50 co-expressed genes with *PWA95606.1* (*DBR2*).

ACKNOWLEDGEMENT

We would like to thank Dr. Daniela Mutwil-Anderwald for constructive comments on this manuscript.

FUNDING

We would like to thank Nanyang Technological University Start Up Grant for funding.

REFERENCES

[1] N. Bouché, D. Bouché, Arabidopsis gene knockout: phenotypes wanted, Curr. Opin. Plant Biol. 4 (2001) 111–117. doi:10.1016/S1369-5266(00)00145-X.

[2] X. Zhang, R. Henriques, S.-S. Lin, Q.-W. Niu, N.-H. Chua, Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method, Nat. Protoc. 1 (2006) 641–646. doi:10.1038/nprot.2006.97.

[3] B.O. Hansen, E.H. Meyer, C. Ferrari, N. Vaid, S. Movahedi, K. Vandepoele, Z. Nikoloski, M. Mutwil, Ensemble gene function prediction database reveals genes important for complex I formation in Arabidopsis thaliana., New Phytol. 217 (2018) 1521–1534. doi:10.1111/nph.14921.

[4] G.W. Bassel, H. Lan, E. Glaab, D.J. Gibbs, T. Gerjets, N. Krasnogor, A.J. Bonner, M.J. Holdsworth, N.J. Provart, Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions, Proc. Natl. Acad. Sci. 108 (2011) 9709–9714. doi:10.1073/pnas.1100958108.

[5] M. Mutwil, B. Usadel, M. Schütte, A. Loraine, O. Ebenhöh, S. Persson, M. Schutte, A. Loraine, O. Ebenhoh, S. Persson, Assembly of an interactive correlation network for the Arabidopsis genome using a novel Heuristic Clustering Algorithm, Plant Physiol. 152 (2010) 29–43. doi:10.1104/pp.109.145318.

[6] A. Takabayashi, N. Ishikawa, T. Obayashi, S. Ishida, J. Obokata, T. Endo, F. Sato, Three novel subunits of Arabidopsis chloroplastic NAD(P)H dehydrogenase identified by bioinformatic and reverse genetic approaches, Plant J. 57 (2009) 207–219. doi:10.1111/j.1365-313X.2008.03680.x.

[7] N. Takahashi, T. Lammens, V. Boudolf, S. Maes, T. Yoshizumi, G. De Jaeger, E. Witters, D. Inzé, L. De Veylder, The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1., EMBO J. 27 (2008) 1840–1851. doi:10.1038/emboj.2008.107.

[8] J.M. Jiménez-Gómez, A.D. Wallace, J.N. Maloof, Network analysis identifies ELF3 as a QTL for the shade avoidance response in arabidopsis, PLoS Genet. 6 (2010).

doi:10.1371/journal.pgen.1001100.
[9] I. Lee, B. Ambaru, P. Thakkar, E.M. Marcotte, S.Y. Rhee, Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana, Nat. Biotechnol. 28 (2010) 149–156. doi:10.1038/nbt.1603.
[10] P. Radivojac, W.T. Clark, T.R. Oron, A.M. Schnoes, T. Wittkop, A. Sokolov, K. Graim, C. Funk, K. Verspoor, A. Ben-Hur, G. Pandey, J.M. Yunes, A large-scale evaluation of computational protein function prediction, Nat. Methods. 10 (2013) 221–227. doi:10.1038/nmeth.2340.
[11] Y. Jiang, T.R. Oron, W.T. Clark, A.R. Bankapur, D. D'Andrea, R. Lepore, C.S. Funk, I. Kahanda, K.M. Verspoor, A. Ben-Hur, D.C.E. Koo, D. Penfold-Brown, D. Shasha, N. Youngs, R. Bonneau, A. Lin, S.M.E. Sahraeian, P.L. Martelli, G. Profiti, R. Casadio, R. Cao, Z. Zhong, J. Cheng, A. Altenhoff, N. Skunca, C. Dessimoz, T. Dogan, K. Hakala, S. Kaewphan, F. Mehryary, T. Salakoski, F. Ginter, H. Fang, B. Smithers, M. Oates, J. Gough, P. Törönen, P. Koskinen, L. Holm, C.-T. Chen, W.-L. Hsu, K. Bryson, D. Cozzetto, F. Minneci, D.T. Jones, S. Chapman, D. BKC, I.K. Khan, D. Kihara, D. Ofer, N. Rappoport, A. Stern, E. Cibrian-Uhalte, P. Denny, R.E. Foulger, R. Hieta, D. Legge, R.C. Lovering, M. Magrane, A.N. Melidoni, P. Mutowo-Meullenet, K. Pichler, A. Shypitsyna, B. Li, P. Zakeri, S. ElShal, L.-C. Tranchevent, S. Das, N.L. Dawson, D. Lee, J.G. Lees, I. Sillitoe, P. Bhat, T. Nepusz, A.E. Romero, R. Sasidharan, H. Yang, A. Paccanaro, J. Gillis, A.E. Sedeño-Cortés, P. Pavlidis, S. Feng, J.M. Cejuela, T. Goldberg, T. Hamp, L. Richter, A. Salamov, T. Gabaldon, M. Marcet-Houben, F. Supek, Q. Gong, W. Ning, Y. Zhou, W. Tian, M. Falda, P. Fontana, E. Lavezzo, S. Toppo, C. Ferrari, M. Giollo, D. Piovesan, S.C.E. Tosatto, A. del Pozo, J.M. Fernández, P. Maietta, A. Valencia, M.L. Tress, A. Benso, S. Di Carlo, G. Politano, A. Savino, H.U. Rehman, M. Re, M. Mesiti, G. Valentini, J.W. Bargsten, A.D.J. van Dijk, B. Gemovic, S. Glisic, V. Perovic, V. Veljkovic, N. Veljkovic, D.C. Almeida-e-Silva, R.Z.N. Vencio, M. Sharan, J. Vogel, L. Kansakar, S. Zhang, S. Vucetic, Z. Wang, M.J.E. Sternberg, M.N. Wass, R.P. Huntley, M.J. Martin, C. O'Donovan, P.N. Robinson, Y. Moreau, A. Tramontano, P.C. Babbitt, S.E. Brenner, M. Linial, C.A. Orengo, B. Rost, C.S. Greene, S.D. Mooney, I. Friedberg, P. Radivojac, An expanded evaluation of protein function prediction methods shows an improvement in accuracy, Genome Biol. 17 (2016) 184. doi:10.1186/s13059-016-1037-6.
[12] S.Y. Rhee, M. Mutwil, Towards revealing the functions of all genes in plants, Trends Plant Sci. 19 (2014) 212–221. doi:10.1016/j.tplants.2013.10.006.
[13] R. Gordan, A.J. Hartemink, M.L. Bulyk, Distinguishing direct versus indirect transcription factor-DNA interactions, Genome Res. 19 (2009) 2090–2100. doi:10.1101/gr.094144.109.
[14] Z. Xie, S. Hu, J. Qian, S. Blackshaw, H. Zhu, Systematic characterization of protein-DNA interactions, Cell. Mol. Life Sci. 68 (2011) 1657–1668. doi:10.1007/s00018-010-0617-y.
[15] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J.A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S.R. Kulkarni, G. Tan, D. Baranasic, D.J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W.W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework, Nucleic Acids Res. 46 (2018) D260–D266. doi:10.1093/nar/gkx1126.
[16] S. Inukai, K.H. Kock, M.L. Bulyk, Transcription factor–DNA binding: beyond binding site motifs, Curr. Opin. Genet. Dev. 43 (2017) 110–119. doi:10.1016/j.gde.2017.02.007.
[17] S.M. Brady, L. Zhang, M. Megraw, N.J. Martinez, E. Jiang, C.S. Yi, W. Liu, A. Zeng, M. Taylor-Teeples, D. Kim, S. Ahnert, U. Ohler, D. Ware, A.J.M. Walhout, P.N. Benfey, A stele-enriched gene regulatory network in the Arabidopsis root., Mol. Syst. Biol. 7 (2011) 459. doi:10.1038/msb.2010.114.
[18] M. Taylor-Teeples, L. Lin, M. de Lucas, G. Turco, T.W. Toal, A. Gaudinier, N.F. Young,

G.M. Trabucco, M.T. Veling, R. Lamothe, P.P. Handakumbura, G. Xiong, C. Wang, J. Corwin, A. Tsoukalas, L. Zhang, D. Ware, M. Pauly, D.J. Kliebenstein, K. Dehesh, I. Tagkopoulos, G. Breton, J.L. Pruneda-Paz, S.E. Ahnert, S.A. Kay, S.P. Hazen, S.M. Brady, An Arabidopsis gene regulatory network for secondary cell wall synthesis, Nature. 517 (2014) 571. https://doi.org/10.1038/nature14099.

[19] E.E. Sparks, C. Drapek, A. Gaudinier, S. Li, M. Ansariola, N. Shen, J.H. Hennacy, J. Zhang, G. Turco, J.J. Petricka, J. Foret, A.J. Hartemink, R. Gordan, M. Megraw, S.M. Brady, P.N. Benfey, Establishment of Expression in the SHORTROOT-SCARECROW Transcriptional Cascade through Opposing Activities of Both Activators and Repressors., Dev. Cell. 39 (2016) 585–596. doi:10.1016/j.devcel.2016.09.031.

[20] T. Ferrier, J.T. Matus, J. Jin, J.L. Riechmann, Arabidopsis paves the way: genomic and network analyses in crops., Curr. Opin. Biotechnol. 22 (2011) 260–270. doi:10.1016/j.copbio.2010.11.010.

[21] Z. Lu, B.T. Hofmeister, C. Vollmers, R.M. DuBois, R.J. Schmitz, Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes., Nucleic Acids Res. 45 (2017) e41. doi:10.1093/nar/gkw1179.

[22] R.C. O'Malley, S.-S.C. Huang, L. Song, M.G. Lewsey, A. Bartlett, J.R. Nery, M. Galli, A. Gallavotti, J.R. Ecker, Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape., Cell. 165 (2016) 1280–1292. doi:10.1016/j.cell.2016.04.038.

[23] D. Marbach, J.C. Costello, R. Küffner, N.M. Vega, R.J. Prill, D.M. Camacho, K.R. Allison, M. Kellis, J.J. Collins, G. Stolovitzky, Wisdom of crowds for robust gene network inference., Nat. Methods. 9 (2012) 796–804. doi:10.1038/nmeth.2016.

[24] C. Smaczniak, R.G.H. Immink, J.M. Muino, R. Blanvillain, M. Busscher, J. Busscher-Lange, Q.D. Dinh, S. Liu, A.H. Westphal, S. Boeren, F. Parcy, L. Xu, C.C. Carles, G.C. Angenent, K. Kaufmann, Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development, Proc. Natl. Acad. Sci. 109 (2012) 1560–1565. doi:10.1073/pnas.1112871109.

[25] J.M. Franco-Zorrilla, R. Solano, Identification of plant transcription factor target sequences, Biochim. Biophys. Acta - Gene Regul. Mech. 1860 (2017) 21–30. doi:https://doi.org/10.1016/j.bbagrm.2016.05.001.

[26] S.K. Palaniswamy, S. James, H. Sun, R.S. Lamb, R. V Davuluri, E. Grotewold, AGRIS and AtRegNet. a platform to link cis-regulatory elements and transcription factors into regulatory networks., Plant Physiol. 140 (2006) 818–829. doi:10.1104/pp.105.072280.

[27] P. Korkuc, D. Walther, The Identification of Cis-Regulatory Sequence Motifs in Gene Promoters Based on SNP Information., Methods Mol. Biol. 1482 (2016) 31–47. doi:10.1007/978-1-4939-6396-6_3.

[28] J. Van de Velde, K.S. Heyndrickx, K. Vandepoele, Inference of Transcriptional Networks in Arabidopsis through Conserved Noncoding Sequence Analysis, Plant Cell. 26 (2014) 2729 LP – 2745. doi:10.1105/tpc.114.127001.

[29] L. Baxter, A. Jironkin, R. Hickman, J. Moore, C. Barrington, P. Krusche, N.P. Dyer, V. Buchanan-Wollaston, A. Tiskin, J. Beynon, K. Denby, S. Ott, Conserved Noncoding Sequences Highlight Shared Components of Regulatory Networks in Dicotyledonous Plants, Plant Cell. 24 (2012) 3949–3965. doi:10.1105/tpc.112.103010.

[30] A. Haudry, A.E. Platts, E. Vello, D.R. Hoen, M. Leclercq, R.J. Williamson, E. Forczek, Z. Joly-Lopez, J.G. Steffen, K.M. Hazzouri, K. Dewar, J.R. Stinchcombe, D.J. Schoen, X. Wang, J. Schmutz, C.D. Town, P.P. Edger, J.C. Pires, K.S. Schumaker, D.E. Jarvis, T. Mandakova, M.A. Lysak, E. van den Bergh, M.E. Schranz, P.M. Harrison, A.M. Moses, T.E. Bureau, S.I. Wright, M. Blanchette, An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions., Nat. Genet. 45 (2013) 891–898. doi:10.1038/ng.2684.

[31] B. Usadel, T. Obayashi, M. Mutwil, F.M. Giorgi, G.W. Bassel, M. Tanimoto, A. Chow, D.

Steinhauser, S. Persson, N.J. Provart, Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats, Plant, Cell Environ. 32 (2009) 1633–1651. doi:10.1111/j.1365-3040.2009.02040.x.
[32] S. Proost, M. Mutwil, Tools of the trade: Studying molecular networks in plants, Curr. Opin. Plant Biol. 30 (2016) 130–140. doi:10.1016/j.pbi.2016.02.010.
[33] S.R. Kulkarni, D. Vaneechoutte, J. Van de Velde, K. Vandepoele, TF2Network: predicting transcription factor regulators and gene regulatory networks in Arabidopsis using publicly available binding site information., Nucleic Acids Res. 46 (2018) e31. doi:10.1093/nar/gkx1279.
[34] P. Pavlidis, J. Gillis, Progress and challenges in the computational prediction of gene function using networks, F1000Research. (2012). doi:10.12688/f1000research.1-14.v1.
[35] S. Oliver, Guilt-by-association goes global., Nature. 403 (2000) 601–603. doi:10.1038/35001165.
[36] R. Sharan, I. Ulitsky, R. Shamir, Network-based prediction of protein function., Mol. Syst. Biol. 3 (2007) 88. doi:10.1038/msb4100129.
[37] J. Kudla, R. Bock, Lighting the Way to Protein-Protein Interactions: Recommendations on Best Practices for Bimolecular Fluorescence Complementation Analyses, Plant Cell. 28 (2016) 1002–1008. doi:10.1105/tpc.16.00043.
[38] R. Zhong, C. Lee, Z.-H. Ye, Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis., Trends Plant Sci. 15 (2010) 625–632. doi:10.1016/j.tplants.2010.08.007.
[39] K.S. Heyndrickx, K. Vandepoele, Systematic Identification of Functional Plant Modules through the Integration of Complementary Data Sources, PLANT Physiol. 159 (2012) 884–901. doi:10.1104/pp.112.196725.
[40] S.K. Tanz, I. Castleden, C.M. Hooper, M. Vacher, I. Small, H. a. Millar, SUBA3: A database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis, Nucleic Acids Res. 41 (2013) 1185–1191. doi:10.1093/nar/gks1151.
[41] V. Vermeirssen, I. De Clercq, T. Van Parys, F. Van Breusegem, Y. Van de Peer, Arabidopsis ensemble reverse-engineered gene regulatory network discloses interconnected transcription factors in oxidative stress., Plant Cell. 26 (2014) 4656–4679. doi:10.1105/tpc.114.131417.
[42] M.A.A. Castro, I. de Santiago, T.M. Campbell, C. Vaughn, T.E. Hickey, E. Ross, W.D. Tilley, F. Markowetz, B.A.J. Ponder, K.B. Meyer, Regulators of genetic risk of breast cancer identified by integrative network analysis, Nat. Genet. 48 (2016) 12–21. doi:10.1038/ng.3458.
[43] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera, A. Califano, ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context, BMC Bioinformatics. 7 (2006) S7. doi:10.1186/1471-2105-7-S1-S7.
[44] A.N. Holding, F.M. Giorgi, A. Donnelly, A.E. Cullen, S. Nagarajan, L.A. Selth, F. Markowetz, VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERa at enhancers in breast cancer, Genome Biol. 20 (2019) 91. doi:10.1186/s13059-019-1698-z.
[45] A. Lachmann, F.M. Giorgi, G. Lopez, A. Califano, ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information, Bioinformatics. 32 (2016) 2233–2235. doi:10.1093/bioinformatics/btw216.
[46] M. Mutwil, S. Klie, T. Tohge, F.M. Giorgi, O. Wilkins, M.M. Campbell, A.R. Fernie, B. Usadel, Z. Nikoloski, S. Persson, PlaNet: Combined Sequence and Expression Comparisons across Plant Networks Derived from Seven Species, Plant Cell. 23 (2011) 895–910. doi:10.1105/tpc.111.083667.

[47] J.M. Stuart, E. Segal, D. Koller, S.K. Kim, A gene-coexpression network for global discovery of conserved genetic modules., Science. 302 (2003) 249–255. doi:10.1126/science.1087447.
[48] F. Markowetz, R. Spang, Inferring cellular networks – a review, BMC Bioinformatics. 8 (2007) S5. doi:10.1186/1471-2105-8-S6-S5.
[49] R. Sibout, S. Proost, B.O. Hansen, N. Vaid, F.M. Giorgi, S. Ho-Yue-Kuang, F. Legée, L. Cézart, O. Bouchabké-Coussa, C. Soulhat, N. Provart, A. Pasha, P. Le Bris, D. Roujol, H. Hofte, E. Jamet, C. Lapierre, S. Persson, M. Mutwil, Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon, New Phytol. 215 (2017) 1009–1025. doi:10.1111/nph.14635.
[50] N.L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification, Nat. Biotechnol. 34 (2016) 525–527. doi:10.1038/nbt.3519.
[51] E.R. Mardis, A decade's perspective on DNA sequencing technology., Nature. 470 (2011) 198–203. doi:10.1038/nature09796.
[52] Human Genome At Ten News Feature, The sequence explosion, Nature. (2010). doi:10.1056/NeJMoa0908094.
[53] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, Nat. Methods. 14 (2017) 417–419. doi:10.1038/nmeth.4197.
[54] R. Leinonen, R. Akhtar, E. Birney, L. Bower, A. Cerdeno-Tárraga, Y. Cheng, I. Cleland, N. Faruque, N. Goodgame, R. Gibson, G. Hoad, M. Jang, N. Pakseresht, S. Plaister, R. Radhakrishnan, K. Reddy, S. Sobhany, P. Ten Hoopen, R. Vaughan, V. Zalunin, G. Cochrane, The European Nucleotide Archive, Nucleic Acids Res. 39 (2011) D28–D31. doi:10.1093/nar/gkq967.
[55] Q. Shen, L. Zhang, Z. Liao, S. Wang, T. Yan, P. Shi, M. Liu, X. Fu, Q. Pan, Y. Wang, Z. Lv, X. Lu, F. Zhang, W. Jiang, Y. Ma, M. Chen, X. Hao, L. Li, Y. Tang, G. Lv, Y. Zhou, X. Sun, P.E. Brodelius, J.K.C. Rose, K. Tang, The Genome of Artemisia annua Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis., Mol. Plant. 11 (2018) 776–788. doi:10.1016/j.molp.2018.03.015.
[56] M. Lohse, A. Nagel, T. Herter, P. May, M. Schroda, R. Zrenner, T. Tohge, A.R. Fernie, M. Stitt, B. Usadel, Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data, Plant, Cell Environ. 37 (2014) 1250–1258. doi:10.1111/pce.12231.
[57] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks., Genome Res. 13 (2003) 2498–2504. doi:10.1101/gr.1239303.
[58] X. Rao, R.A. Dixon, Current Models for Transcriptional Regulation of Secondary Cell Wall Biosynthesis in Grasses, Front. Plant Sci. 9 (2018) 399. doi:10.3389/fpls.2018.00399.
[59] H.E. McFarlane, A. Döring, S. Persson, The cell biology of cellulose synthesis., Annu. Rev. Plant Biol. 65 (2014) 69–94. doi:10.1146/annurev-arplant-050213-040240.
[60] C. Ruprecht, M. Mutwil, F. Saxe, M. Eder, Z. Nikoloski, S. Persson, Large-Scale Co-Expression Approach to Dissect Secondary Cell Wall Formation Across Plant Species, Front. Plant Sci. 2 (2011) 1–13. doi:10.3389/fpls.2011.00023.
[61] E.R. Lampugnani, E. Flores-Sandoval, Q.W. Tan, M. Mutwil, J.L. Bowman, S. Persson, Cellulose Synthesis – Central Components and Their Evolutionary Relationships, Trends Plant Sci. (2019) 1–11. doi:https://doi.org/10.1016/j.tplants.2019.02.011.
[62] V. Bischoff, S. Nita, L. Neumetzler, D. Schindelasch, A. Urbain, R. Eshed, S. Persson, D. Delmer, W.-R. Scheible, TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Biosynthesis in

Arabidopsis, Plant Physiol. 153 (2010) 590–602. doi:10.1104/pp.110.153320.
[63] A.S. Rajangam, M. Kumar, H. Aspeborg, G. Guerriero, L. Arvestad, P. Pansri, C.J.-L. Brown, S. Hober, K. Blomqvist, C. Divne, I. Ezcurra, E. Mellerowicz, B. Sundberg, V. Bulone, T.T. Teeri, MAP20, a Microtubule-Associated Protein in the Secondary Cell Walls of Hybrid Aspen, Is a Target of the Cellulose Synthesis Inhibitor 2,6-Dichlorobenzonitrile, Plant Physiol. 148 (2008) 1283–1294. doi:10.1104/pp.108.121913.
[64] C.P. MacMillan, S.D. Mansfield, Z.H. Stachurski, R. Evans, S.G. Southerton, Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus, Plant J. 62 (2010) 689–703. doi:10.1111/j.1365-313X.2010.04181.x.
[65] R. Zhong, Z.-H. Ye, Secondary Cell Walls: Biosynthesis, Patterned Deposition and Transcriptional Regulation, Plant Cell Physiol. 56 (2014) 195–214. doi:10.1093/pcp/pcu140.
[66] N. Abdulrazzak, B. Pollet, J. Ehlting, K. Larsen, C. Asnaghi, S. Ronseau, C. Proux, M. Erhardt, V. Seltzer, J.-P. Renou, P. Ullmann, M. Pauly, C. Lapierre, D. Werck-Reichhart, A coumaroyl-ester-3-hydroxylase Insertion Mutant Reveals the Existence of Nonredundant meta-Hydroxylation Pathways and Essential Roles for Phenolic Precursors in Cell Expansion and Plant Growth, Plant Physiol. 140 (2006) 30–48. doi:10.1104/pp.105.069690.
[67] R. Reboul, C. Geserick, M. Pabst, B. Frey, D. Wittmann, U. Lütz-Meindl, R. Léonard, R. Tenhaken, Down-regulation of UDP-glucuronic Acid Biosynthesis Leads to Swollen Plant Cell Walls and Severe Developmental Defects Associated with Changes in Pectic Polysaccharides, J. Biol. Chem. . 286 (2011) 39982–39992. doi:10.1074/jbc.M111.255695.
[68] J.-B. He, X.-H. Zhao, P.-Z. Du, W. Zeng, C.T. Beahan, Y.-Q. Wang, H.-L. Li, A. Bacic, A.-M. Wu, KNAT7 positively regulates xylan biosynthesis by directly activating IRX9 expression in Arabidopsis, J. Integr. Plant Biol. 60 (2018) 514–528. doi:10.1111/jipb.12638.
[69] A. Oikawa, H.J. Joshi, E.A. Rennie, B. Ebert, C. Manisseri, J.L. Heazlewood, H.V. Scheller, An Integrative Approach to the Identification of Arabidopsis and Rice Genes Involved in Xylan and Secondary Wall Development, PLoS One. 5 (2010) e15481. https://doi.org/10.1371/journal.pone.0015481.
[70] H. Temple, J.C. Mortimer, T. Tryfona, X. Yu, F. Lopez-Hernandez, M. Sorieul, N. Anders, P. Dupree, Two members of the DUF579 family are responsible for arabinogalactan methylation in Arabidopsis, Plant Direct. 3 (2019) e00117. doi:10.1002/pld3.117.
[71] G. Xiong, K. Cheng, M. Pauly, Xylan O-Acetylation Impacts Xylem Development and Enzymatic Recalcitrance as Indicated by the Arabidopsis Mutant tbl29, Mol. Plant. 6 (2013) 1373–1375. doi:https://doi.org/10.1093/mp/sst014.
[72] S. Persson, K.H. Caffall, G. Freshour, M.T. Hilley, S. Bauer, P. Poindexter, M.G. Hahn, D. Mohnen, C. Somerville, The Arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity., Plant Cell. 19 (2007) 237–255. doi:10.1105/tpc.106.047720.
[73] Y. Yuan, Q. Teng, R. Zhong, M. Haghighat, E.A. Richardson, Z.-H. Ye, Mutations of Arabidopsis TBL32 and TBL33 Affect Xylan Acetylation and Secondary Wall Deposition, PLoS One. 11 (2016) e0146460. https://doi.org/10.1371/journal.pone.0146460.
[74] L. Zhu, J. Guo, C. Zhou, J. Zhu, Ectopic expression of LBD15 affects lateral branch development and secondary cell wall synthesis in Arabidopsis thaliana, Plant Growth Regul. 73 (2014) 111–120. doi:10.1007/s10725-013-9873-9.
[75] S. Andersson-Gunnerås, E.J. Mellerowicz, J. Love, B. Segerman, Y. Ohmiya, P.M. Coutinho, P. Nilsson, B. Henrissat, T. Moritz, B. Sundberg, Biosynthesis of cellulose-enriched tension wood in Populus: global analysis of transcripts and metabolites identifies

biochemical and developmental regulators in secondary wall biosynthesis, Plant J. 45 (2006) 144–165. doi:10.1111/j.1365-313X.2005.02584.x.
[76] S. Legay, G. Guerriero, A. Deleruelle, M. Lateur, D. Evers, C.M. André, J.-F. Hausman, Apple russeting as seen through the RNA-seq lens: strong alterations in the exocarp cell wall, Plant Mol. Biol. 88 (2015) 21–40. doi:10.1007/s11103-015-0303-4.
[77] C. Halpin, Cell Biology: Up Against the Wall, Curr. Biol. 23 (2013) R1048–R1050. doi:https://doi.org/10.1016/j.cub.2013.10.033.
[78] P.M. O'Neill, Medicinal chemistry: a worthy adversary for malaria., Nature. 430 (2004) 838–839. doi:10.1038/430838a.
[79] World Health Organisation, World Malaria Report 2018, Geneva, 2018. https://apps.who.int/iris/bitstream/handle/10665/275867/9789241565653-eng.pdf?ua=1 (accessed March 12, 2019).
[80] T. Mita, K. Tanabe, Evolution of Plasmodium falciparum drug resistance: implications for the development and containment of artemisinin resistance., Jpn. J. Infect. Dis. 65 (2012) 465–475.
[81] A. Salvador, R.M. Hernandez, J.L. Pedraz, M. Igartua, Plasmodium falciparum malaria vaccines: current status, pitfalls and future directions., Expert Rev. Vaccines. 11 (2012) 1071–1086. doi:10.1586/erv.12.87.
[82] H. Gelband, A. Seiter, A global subsidy for antimalarial drugs., Am. J. Trop. Med. Hyg. 77 (2007) 219–221.
[83] A.B.S. Sidhu, D. Verdier-Pinard, D.A. Fidock, Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations., Science. 298 (2002) 210–213. doi:10.1126/science.1074045.
[84] R.G. Ridley, Malaria: to kill a parasite., Nature. 424 (2003) 887–889. doi:10.1038/424887a.
[85] Y. Tu, The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine, Nat. Med. 17 (2011) 1217. https://doi.org/10.1038/nm.2471.
[86] J.-M. LIU, et al., Structure and reaction of arteannuin., Repr. from Acta Chim. Sin. 37 (1979) 129–143.
[87] D.-M. Ma, Z. Wang, L. Wang, F. Alejos-Gonzales, M.-A. Sun, D.-Y. Xie, A Genome-Wide Scenario of Terpene Pathways in Self-pollinated Artemisia annua., Mol. Plant. 8 (2015) 1580–1598. doi:10.1016/j.molp.2015.07.004.
[88] D. Ma, G. Li, F. Alejos-Gonzalez, Y. Zhu, Z. Xue, A. Wang, H. Zhang, X. Li, H. Ye, H. Wang, B. Liu, D.-Y. Xie, Overexpression of a type-I isopentenyl pyrophosphate isomerase of Artemisia annua in the cytosol leads to high arteannuin B production and artemisinin increase., Plant J. 91 (2017) 466–479. doi:10.1111/tpj.13583.
[89] C.J. Paddon, P.J. Westfall, D.J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M.D. Leavell, A. Tai, A. Main, D. Eng, D.R. Polichuk, K.H. Teoh, D.W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K.W. Ellens, S. Fickes, J. Galazzo, S.P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L.F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P.S. Covello, J.D. Keasling, K.K. Reiling, N.S. Renninger, J.D. Newman, High-level semi-synthetic production of the potent antimalarial artemisinin, Nature. 496 (2013) 528. https://doi.org/10.1038/nature12051.
[90] V.G. Yadav, M. De Mey, C. Giaw Lim, P. Kumaran Ajikumar, G. Stephanopoulos, The future of metabolic engineering and synthetic biology: Towards a systematic practice, Metab. Eng. 14 (2012) 233–241. doi:https://doi.org/10.1016/j.ymben.2012.02.001.
[91] M. Chen, T. Yan, Q. Shen, X. Lu, Q. Pan, Y. Huang, Y. Tang, X. Fu, M. Liu, W. Jiang, Z. Lv, P. Shi, Y.-N. Ma, X. Hao, L. Zhang, L. Li, K. Tang, GLANDULAR TRICHOME-SPECIFIC WRKY 1 promotes artemisinin biosynthesis in Artemisia annua., New Phytol.

214 (2017) 304–316. doi:10.1111/nph.14373.
[92] Z. Lv, L. Zhang, K. Tang, New insights into artemisinin regulation, Plant Signal. Behav. 12 (2017) e1366398–e1366398. doi:10.1080/15592324.2017.1366398.
[93] B.A. Krizek, C.J. Bequette, K. Xu, I.C. Blakley, Z.Q. Fu, J.W. Stratmann, A.E. Loraine, RNA-Seq Links the Transcription Factors AINTEGUMENTA and AINTEGUMENTA-LIKE6 to Cell Wall Remodeling and Plant Defense Pathways, Plant Physiol. 171 (2016) 2069–2084. doi:10.1104/pp.15.01625.
[94] S. Wang, Y. Chang, B. Ellis, Overview of OVATE FAMILY PROTEINS, A Novel Class of Plant-Specific Growth Regulators, Front. Plant Sci. 7 (2016) 417. doi:10.3389/fpls.2016.00417.
[95] S. Raffaele, S. Mongrand, P. Gamas, A. Niebel, T. Ott, Genome-wide annotation of remorins, a plant-specific protein family: evolutionary and functional perspectives, Plant Physiol. 145 (2007) 593–600. doi:10.1104/pp.107.108639.
[96] J. Liu, J.H. Rice, N. Chen, T.J. Baum, T. Hewezi, Synchronization of Developmental Processes and Defense Signaling by Growth Regulating Transcription Factors, PLoS One. 9 (2014) e98477. https://doi.org/10.1371/journal.pone.0098477.
[97] M. Gonen, B.A. Weir, G.S. Cowley, F. Vazquez, Y. Guan, A. Jaiswal, M. Karasuyama, V. Uzunangelov, T. Wang, A. Tsherniak, S. Howell, D. Marbach, B. Hoff, T.C. Norman, A. Airola, A. Bivol, K. Bunte, D. Carlin, S. Chopra, A. Deran, K. Ellrott, P. Gopalacharyulu, K. Graim, S. Kaski, S.A. Khan, Y. Newton, S. Ng, T. Pahikkala, E. Paull, A. Sokolov, H. Tang, J. Tang, K. Wennerberg, Y. Xie, X. Zhan, F. Zhu, T. Aittokallio, H. Mamitsuka, J.M. Stuart, J.S. Boehm, D.E. Root, G. Xiao, G. Stolovitzky, W.C. Hahn, A.A. Margolin, A Community Challenge for Inferring Genetic Predictors of Gene Essentialities through Analysis of a Functional Screen of Cancer Cell Lines., Cell Syst. 5 (2017) 485-497.e3. doi:10.1016/j.cels.2017.09.004.
[98] D. Marbach, R.J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, G. Stolovitzky, Revealing strengths and weaknesses of methods for gene network inference, Proc. Natl. Acad. Sci. 107 (2010) 6286–6291. doi:10.1073/pnas.0913357107.
[99] M. Franz, H. Rodriguez, C. Lopes, K. Zuberi, J. Montojo, G.D. Bader, Q. Morris, GeneMANIA update 2018, Nucleic Acids Res. 46 (2018) W60–W64. doi:10.1093/nar/gky311.
[100] T. Lee, I. Lee, AraNet: A Network Biology Server for Arabidopsis thaliana and Other Non-Model Plant Species BT - Plant Gene Regulatory Networks: Methods and Protocols, in: K. Kaufmann, B. Mueller-Roeber (Eds.), Springer New York, New York, NY, 2017: pp. 225–238. doi:10.1007/978-1-4939-7125-1_15.
[101] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life., Nucleic Acids Res. 43 (2015) D447--52. doi:10.1093/nar/gku1003.
[102] C. Ruprecht, A. Mendrinna, T. Tohge, A. Sampathkumar, S. Klie, A.R. Fernie, Z. Nikoloski, S. Persson, M. Mutwil, FamNet: A framework to identify multiplied modules driving pathway diversification in plants., Plant Physiol. 170 (2016) 1878–1894. doi:10.1104/pp.15.01281.
[103] R. Sibout, P. Le Bris, F. Legée, L. Cézard, H. Renault, C. Lapierre, Structural Redesigning Arabidopsis Lignins into Alkali-Soluble Lignins through the Expression of p-Coumaroyl-CoA:Monolignol Transferase PMT., Plant Physiol. 170 (2016) 1358–1366. doi:10.1104/pp.15.01877.
[104] T. Tohge, M.S. Ramos, A. Nunes-Nesi, M. Mutwil, P. Giavalisco, D. Steinhauser, M. Schellenberg, L. Willmitzer, S. Persson, E. Martinoia, A.R. Fernie, Toward the storage metabolome: Profiling the barley vacuole, Plant Physiol. 157 (2011). doi:10.1104/pp.111.185710.

[105] L. Chae, T. Kim, R. Nilo-Poyanco, S.Y. Rhee, Genomic signatures of specialized metabolism in plants, Science (80-.). 344 (2014) 510–513. doi:10.1126/science.1252076.
[106] I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, Natural history and evolutionary principles of gene duplication in fungi., Nature. 449 (2007) 54–61. doi:10.1038/nature06107.
[107] N. Ziemert, M. Alanjary, T. Weber, The evolution of genome mining in microbes – a review, Nat. Prod. Rep. 33 (2016) 988–1005. doi:10.1039/C6NP00025H.
[108] J. Friesner, S.M. Assmann, R. Bastow, J. Bailey-Serres, J. Beynon, V. Brendel, C.R. Buell, A. Bucksch, W. Busch, T. Demura, J.R. Dinneny, C.J. Doherty, A.L. Eveland, P. Falter-Braun, M.A. Gehan, M. Gonzales, E. Grotewold, R. Gutierrez, U. Kramer, G. Krouk, S. Ma, R.J.C. Markelz, M. Megraw, B.C. Meyers, J.A.H. Murray, N.J. Provart, S. Rhee, R. Smith, E.P. Spalding, C. Taylor, T.K. Teal, K.U. Torii, C. Town, M. Vaughn, R. Vierstra, D. Ware, O. Wilkins, C. Williams, S.M. Brady, The Next Generation of Training for Arabidopsis Researchers: Bioinformatics and Quantitative Biology, Plant Physiol. 175 (2017) 1499–1509. doi:10.1104/pp.17.01490.

TABLE AND FIGURE LEGENDS

Figure 1. Establishing RNA sequencing and co-expression network construction pipeline on Rock64. A) Anatomy of Rock64 single board computer. The major components are 1. central processing unit (CPU), 2. 4GB RAM, 3. HDR digital output, 4. Ethernet port, 5. USB ports. B) Flowchart schematic for the pipeline used to download and process the RNA sequencing data and construct a co-expression network. C) Download and kallisto mapping times for the *Artemisia annua* data.

Figure 2. Regulatory and co-expression network of secondary cell wall biosynthesis in Artemisia. A) Schematic regulatory network of CESAs. Transcription factors are indicated in green, while their targets are shown in red. B) Co-expression network of CESA7 from *Artemisia*. Nodes represent genes, while edges connect co-expressed genes. Red nodes indicate genes involved in cell wall biosynthesis and green nodes indicate transcription factors.

Figure 3. Expression profile and co-expression network of artemisinin biosynthesis pathway. A) A scheme showing biosynthesis of artemisinin. Abbreviations for enzymes include ADS, armorpha-4, 11-diene synthase, ADH1: alcohol dehydrogenase 1, ALDH1: aldehyde dehydrogenase, CPR: cytochrome P450 reductase, CYP71AV1: cytochrome P450 mono-oxygenase; DBR2: artemisinic aldehyde delta-11(13)-double bond reductase. B) Co-expression network of ADS (*PWA62882.1*). Red, orange and green nodes indicate enzymes, transcription factors and ABC transporters respectively. Abbreviations for genes include AIL6: aintegumenta-like 6; GRF: growth regulating factor and OFP: ovate family protein. For brevity, only 50 most co-expressed genes of the query gene are shown. C) Co-expression network of DBR2 (*PWA95606.1*). Abbreviations for genes include Znf-4CXXC_R1: zinc-finger domain of monoamine-oxidase A repressor R1; CYC-like 8: cycloidea-like 8 and ZF: zinc finger. For brevity, only 50 most co-expressed genes of the query gene are shown. D) Expression profile of the two query genes from the shown co-expression networks (Figure 3B and C). Transcript per Million (TPM) values for each sample were normalised by the highest TPM value for the gene across all SRA samples, as represented on the y-axis and SRA run IDs are represented on the

x-axis. *PWA62882.1* (ADS) and *PWA95606.1* (DBR2) are indicated by brown and orange lines, respectively.

Suppplementary figure 1. Co-expression network of secondary cell wall biosynthesis in *Artemisia*. Co-expression network of CESA7 from *Artemisia*. Nodes represent genes, while edges connect co-expressed genes. Red nodes indicate genes involved in cell wall biosynthesis, green nodes indicate transcription factors, while blue nodes show other genes.

Supplementary figure 2. Co-expression network of artemisinin biosynthesis pathway. A) Co-expression network of ADS (*PWA62882.1*). Red, orange and green nodes indicate enzymes, transcription factors and ABC transporters respectively, while blue nodes show other genes. Abbreviations for genes include AIL6: aintegumenta-like 6; GRF: growth regulating factor and OFP: ovate family protein. For brevity, only 50 most co-expressed genes of the query gene are shown. B) Co-expression network of DBR2 (*PWA95606.1*). Abbreviations for genes include Znf-4CXXC_R1: zinc-finger domain of monoamine-oxidase A repressor R1; CYC-like 8: cycloidea-like 8 and ZF: zinc finger. For brevity, only 50 most co-expressed genes of the query gene are shown.