Gene Regulatory Networks for Lignin Biosynthesis in Switchgrass (*Panicumvirgatum*)

Xiaolan Rao[1,2], Xin Chen[2,7], Hui Shen[1,2,8], Qin Ma[2,9], Guifen Li[4], Yuhong Tang[2,4], Maria Pena[2,5], William York[2,5], Taylor Frazier[6], Scott Lenaghan[6], Xirong Xiao[1], Fang Chen[1,2,3] and Richard A. Dixon[1,2,3*]

[1]BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, Texas 76203

[2]BioEnergy Science Center (BESC), Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

[3]Center for Bioenergy Innovation (CBI), Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

[4]Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, OK, 73401

[5]Complex Carbohydrate Research Center and Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, 30602.

[6]Department of Plant Sciences, University of Tennessee, Knoxville, TN, 37996.

[7]Present address: Center for Applied Mathematics, Tianjin University, Tianjin, 300072, China

[8]Present address: Marker-assisted Breeding and Traits, Chromatin Inc, Lubbock, TX, 79404, USA

[9]Present address: Department of Plant Science, South Dakota State University, Brookings, 57006, USA

Received…….

*Correspondence (Tel +1 940 565 2308; email Richard.Dixon@unt.edu)

Word count

Summary: 235
Introduction: 797
Results: 3,317
Discussion: 973
Experimental procedures: 965
Acknowledgements: 71
Figure legends: 797

Summary

Cell wall recalcitrance is the major challenge to improving saccharification efficiency in converting lignocellulose into biofuels. However, information regarding the transcriptional regulation of secondary cell wall biogenesis remains poor in switchgrass (*Panicumvirgatum*), which has been selected as a biofuel crop in the United States. In this study, we present a combination of computational and experimental approaches to develop gene regulatory networks for lignin formation in switchgrass. To screen transcription factors (TFs) involved in lignin biosynthesis, we developed a new method to perform co-expression network analysis using 14 lignin biosynthesis genes as bait genes. The switchgrass lignin co-expression network was further extended by adding 14 TFs identified in this study, and seven TFs identified in previous studies, as bait genes. Six TFs (PvMYB58/63, PvMYB42/85, PvMYB4, PvWRKY12, PvSND2 and PvSWN2) were targeted to generate overexpressing and/or downregulated transgenic switchgrass lines. The alteration of lignin content, cell wall composition and/or plant growth in the transgenic plants supported the role of the TFs in controlling secondary wall formation. RNA-seq analysis of four of the transgenic switchgrass linesrevealed downstream target genes of the secondary wall-related TFs and crosstalk with other biological pathways. In vitro transactivation assays further confirmed the regulation of specific lignin pathway genes by four of the TFs. Our meta-analysis provides a hierarchical networkof TFs and their potential target genes for future manipulation of secondary cell wall formationfor lignin reduction or valorization in switchgrass.

Introduction

The cell wall is deposited outside the plant cell membrane as a cellular exoskeleton that can be classified as primary or secondary depending on function and composition (Vogel, 2008). The secondarycell wall polymers cellulose, hemicellulose and lignin constitutethe most abundant source of plant biomass thatprovides raw materials for generating renewable biofuels (Bouton, 2007; Pauly and Keegstra, 2010). However, plant cell walls possess chemical and structuralproperties (called “biomass recalcitrance”) to help prevent microbial and enzymatic deconstruction(Himmel *et al.*, 2007). Overcoming biomass recalcitrance is a major challenge in the lignocellulosic bioenergyindustry. Lignin, covalently incorporated into the cross-linked matrixof polysaccharides, prevents access of hydrolytic enzymes for degradingthe lignocellulosic components to monosaccharidesfor subsequent biofuel production (Pauly and Keegstra, 2010).Conversely, lignin has potential to be developed as a high value co-product of bioprocessing (Ragauskas et al., 2014). For both applications, knowledge of the gene regulatory networks underlying lignin and secondary cell wall biosynthesis is necessary to facilitate targeted genetic interventions.

Switchgrass (*Panicumvirgatum L.*) has been selected as a major cellulosic feedstock for bioconversion toethanol in the United States(Bouton, 2007). Manipulating cell walls in switchgrass through down-regulation of GAUT4 (*galacturonosyltransferase4,* involved in pectin biosynthesis), COMT (*caffeic acid/5-hydroxyferulic acid 3-O-methyltransferase,* involved in lignin biosynthesis) and FPGS (*folylpolyglutamate synthase 1,* involved in C1 metabolism) or overexpression of MYB4 (a repressor of lignin biosynthesis) in all cases leads to a reduction of cell wall recalcitrance and an increased efficiency of sugar release (Dumitrache *et al.*, 2017). However, as a non-model organism, the lack of genetic information on secondary cell wall formation in switchgrass limits biotechnological approaches to feedstock development. Previously we have identified genes involved in monolignol biosynthesis and provided candidate structural genes associated withsecondary wall development in brassinosteroid-induced switchgrass suspension cultures(Shen *et al.*, 2013; Rao *et al.*, 2017).Besides characterizing metabolic genes in secondary wall biosynthesis, understanding the transcription factors (TFs) that regulate secondary wall formation is required to enable rational biodesign of switchgrass as a bioenergy crop.

In recent decades, studies on *Arabidopsis thaliana* have contributed to a thorough analysis of TFs involved in secondary wall regulation, including sub-group members of the NAC, MYB, WRKY and other TF families (Zhong and Ye, 2015; Wang and Dixon, 2012; Nakano *et al.*, 2015; Taylor-Teeples *et al.*, 2015). A hierarchical organization is observed in the transcriptional regulatory system in Arabidopsis that regulates secondary wall biosynthesis in concert with other metabolic pathways. However, few TFs have been reported that control secondary wall developmentin grasses, especially in switchgrass (Zhong *et al.*, 2015; Shen *et al.*, 2012; Rao and Dixon, 2018). Switchgrass (Pv) SWNs and PvMYB46 have been shown to be activators of secondary wall biosynthesis based on ectopic expression in Arabidopsis (Zhong *et al.*, 2015). PvSWNs and PvMYB46 are capable of rescuing the secondary wall defects in the Arabidopsis *snd1*/*nst1* and *myb46*/*myb83* double mutants, respectively, and overexpression of PvSWNs and PvMYB46 in Arabidopsis leads to activation of the secondary wall biosynthesis program (Zhong *et al.*, 2015). We have characterized PvMYB4 as a lignin repressor; overexpression of PvMYB4 in transgenic switchgrass and tobacco results in the downregulation of lignin biosynthesis genes and a reduced lignin content (Shen *et al.*, 2012). Other TFs involved in switchgrass secondary wall regulation remain largely unexplored.

With the increased number of public microarray datasets, a co-expression approach has been widely used to investigate TF candidates and their potential target genes because the expression of transcriptional regulators and their targets tends to be coordinated (Ruprecht and Persson, 2012;

Serin*et al.*, 2016). Several groups have applied a combination of large-scale co-expression analysis and experimental evaluation to discover novel TFs involved in cell wall formation in Arabidopsis, rice, maize and sugarcane (Hirano *et al.*, 2013a; Ruprecht *et al.*, 2011; Ruprecht and Persson, 2012; Cassan-Wang *et al.*, 2013; Hansen *et al.*, 2014; Ferreira *et al.*, 2016).

To understand the transcriptional network regulating lignin biosynthesis in switchgrass, here we present a computational approach with co-expression and RNA-seq analyses to explore genes that are strongly associated with validated TFs, complemented by in planta validation by transgenesis. First, we developed a new method (QUBIC) to detect TFs co-expressed with previously validated lignin biosynthesis genes in switchgrass, and extended the switchgrass co-expression network by adding 21 switchgrass TF genes identified as secondary wall regulators as seed genes. Second, we targeted six TFs (PvMYB58/63, PvMYB42/85, PvMYB4, PvWRKY12, PvSND2 and PvSWN2)for overexpression and/or downregulation in transgenic switchgrass, and analyzed the consequences by RNA-seq analysis. Finally, interactions between TFs and their candidate target promoters were interrogated by in vitro transactivation assays. Together these analyses revealed secondary wall-related TFs, their downstream target genes, and the crosstalk between secondary wall-related TFs and other biological pathways, to inform targeted modification of cell wall composition for enhancing processing of switchgrass biomass.

Results

Co-expression screening of switchgrass TFs involved in lignin biosynthesis

To detect genes with similar expression patterns in large datasets, we developed a methodfor co-expression analysis based on QUBIC, a previously reported bi-clustering algorithm (Li *et al.*, 2009). Microarray datasets are visualized as a matrix with rows defined as genes and columns defined as conditions. First, we normalized the expression values of probes under all conditions into three groups: low expression, normal expression and high expression, which are represented by -1, 0 and 1, respectively (Supplementary Figure 1).Then, we defined the BF score between two genes g_i and $g_j$in an expression matrix dataset with N samples as defined in the following formula.

$$\mathrm{BF}(g_i, g_j) = \left(\sum_{k=1}^{N} \mathrm{I}(g_i^k, g_j^k)\ /\mathrm{N}\right) \times \mathrm{PCC}(g_i, g_j)^{\text{where } \mathrm{I}(g_i^k, g_j^k)=1}$$

A detailed description of the approach and derivation of the BF score is provided in Supplementary Methods.

To identify candidate TFs involved in secondary cell wall biosynthesis, we performed a comparative co-expression analysis with a large scale dataset of Arabidopsis and switchgrass transcriptomes using lignin biosynthesis genes as guide genes. Specifically, 16 and 14 lignin biosynthesis genes were used as baits to search genes with correlated expression pattern against Arabidopsis (GSE34188, 99 data points) (Hanada *et al.*, 2013) and switchgrass (93 data points) (Zhang *et al.*, 2013) public microarray datasets, respectively (Supplementary Table 1). The bait genes represented all eleven families of enzymes responsible for monolignol biosynthesis. The selected microarray datasets in Arabidopsis and switchgrass represented similar conditions including major tissue and organ types during developmental stages of the entire plant life-cycle, and mature plants exposed to a series of stress treatments. The co-expressed genes were determined by a cut-off value of top 5% for the significance of correlation rank between two genes. Using the approach, we identified 645 and 1,274 transcription factors coordinated with 16 and 14 lignin biosynthesis genes in Arabidopsis and switchgrass, respectively (Supplementary Dataset 1).

In both species, TFs in the co-expression network with lignin biosynthesis genes displayed a similar distribution of TF families (Figure 1A, Supplementary Table 2). They shared the same categories of the four most abundant families sorted in descending order; MYB (81 in Arabidopsis and 124 in switchgrass), bHLH (70 in Arabidopsis and 111 in switchgrass), NAC (64 in Arabidopsis and 109 in switchgrass) and ERF (51 in Arabidopsis and 102 in switchgrass). The C2H2-, WRKY- and bZIP-type TFs were subsequently listed in the fifth to seventh places in both Arabidopsis and switchgrass. To obtain phylogenetic relationships of the secondary wall-associated TFs, we generated phylogenetic trees of the MYB, NAC, bHLH, ERF and WRKY-type TFs that appeared in the co-expression networks (Supplementary Figure 2). The analysis showed that many of the Arabidopsis and switchgrass genes were grouped into the same clades according to their protein sequences (Supplementary Figure 2). We next used the blastP program to search switchgrass TF proteins against Arabidopsis TF proteins in the network. This revealed that 522 switchgrass TFs have homologs that were present in the Arabidopsis network, corresponding to 202 Arabidopsis TFs (Supplementary Dataset 1). Many of the homologs tended to be co-expressed with similar lignin biosynthesis genes in the two species (Figure 1B). Some have previously been shown to be involved in secondary cell wall regulation in Arabidopsis and/or switchgrass. For example, the expression of AtMYB46 (AT5G12870) and its homolog PvMYB46 (KanlowCTG44101_s_at) is tightly correlated with PAL, 4CL, COMT and CCoAoMT, but not with C3´H and HCT, in both species (Supplementary Dataset 1). PvMYB46 is a functional ortholog of Arabidopsis MYB46, with ability to rescue cell wall defects in the Arabidopsis *myb46*/*myb83* double mutant (Zhong *et al.*, 2015).

The above observations suggest that Arabidopsis and switchgrass have evolved a conserved mechanism of secondary cell wall biosynthesis by recruiting homologous members of certain TF families. However, a close observation revealed different features in the TF co-expression networks between the two species (Supplementary Figure 3). For example, Arabidopsis F5H shared a smaller proportion of co-expressed TFs with other lignin biosynthesis genes (38 out of 78 TFs) than did switchgrass F5H (238 of 259 TFs). AtMYB58 and AtMYB63 are co-expressed with most lignin biosynthesis genes except F5H in Arabidopsis (Supplementary Figure 3).This is consistent with their role in activating the expression of lignin biosynthesis genes via binding to AC elements in the target gene's promoter region with exclusion of F5H, the promoter of which does not contain AC-rich elements (Zhou *et al.*, 2009; Zhao and Dixon, 2011). However, the switchgrass homologs of MYB58/63 (AP13ITG56055_at and AP13ITG57154RC_at) showed a correlated expression with F5H and most other lignin biosynthesis genes (Supplementary Figure 3), suggesting that PvMYB58/63 may be involved in the regulation of F5H expression in switchgrass, but not in Arabidopsis.

Functional analysis of TF candidates bytransgenesis

To evaluate the TF candidates identified in the co-expression network, we selected six genes belonging to the MYB and NAC families for overexpression and/or down-regulation in transgenic switchgrass. The phenotype of the PvMYB4-OX line has been described in our previous report (Shen *et al.*, 2012).

PvMYB58/63. Four genes in switchgrass clustered in the MYB58/63 clade (Supplementary Figure 4). Compared with rice, sorghum and *Brachypodiumdistachyon*, the expanded number of MYB58/63 members in switchgrass is due to its increased ploidy level. We named the genes PvMYB58/63A, PvMYB58/63B, PvMYB58/63C and PvMYB58/63D. Among them, PvMYB58/63A and PvMYB58/63B share 95% similarity at the amino acid level. All are highly co-expressed with lignin biosynthesis genes (Supplementary Dataset 1). Quantitative PCR analysis showed that PvMYB58/63A is more highly expressed in stem than in leaf blade and leaf sheaths, and is overall more highly expressed than PvMYB58/63C (Figure 2A). In situ hybridization further showed that PvMYB58/63A and PvMYB58/63C are expressed in all cell types in switchgrass stem cross sections(Figure 4Band Supplementary Figure 5). These results suggest that, like their homologs in Arabidopsis and rice, PvMYB58/63s, especially PvMYB58/63A, are probably involved in secondary wall formation.

To investigate the function of PvMYB58/63 in vivo, we generated overexpression- and downregulation- lines of PvMYB58/63A in transgenic switchgrass. The RNA interference (RNAi) approach was used to achieve downregulation of the target gene. Transcript levels of the target gene

inthetransgenic lines were measured using qPCR (Figure 2C and Supplementary Figure 6). The selected overexpression and knockdown lines exhibited at least a 50-fold increase and a 70% to 85% reduction in PvMYB58/63 expression compared with that of the vector control plant, respectively (Figure 2C and Supplementary Figure 6).

Compared with the controls, the PvMYB58/63A-overexpressing plants displayed on average a 40% reduction in height(Figure 2D, E). Cross-sections of mature stems were stained with phloroglucinol-HCL to assess total lignin. All MYB58/63A-overexpressors displayed increased staining in mature stem (Figure 2F). Increased total lignin content was confirmed by the AcBr method (1.1 to 1.3-fold increase, Figure 2G) and by thioacidolysis (1.2 to 1.6- fold increase) in the whole tillers at the E4 stage (Figure 2H). In addition, an increased S:G ratio was observed in the MYB58/63A overexpressors due to a disproportionate increase in S units (Figure 2I).qRT-PCR analysis revealed, consistent with the co-expression analysis, increased expression of F5H (Figure 2J) and all other lignin biosynthesis genes (except C3´H) in PvMYB58/63A-overexpression lines (Supplementary Figure 7). In contrast, and similar to a previous report on rice OsMYB58/63-RNAi lines (Hirano *et al.*, 2013b), no obvious effects of knockdown of PvMYB58/63were observed on plant growth, stem structure or lignin content, although several lignin biosynthesis genes showed reduced expression (Supplementary Figure 8). This possibly reflects redundancy by homologous genes and/or other TFs.Taken together, our results show that PvMYB58/63A is an activator of secondary wall formation via upregulation of most lignin biosynthesis genes including F5H.

PvMYB42/85. Four genes were found to be homologs of AtMYB42/85 in the switchgrass genome (Supplementary Figure 3), and named *PvMYB42/85A, B, C* and *D*. Three of them were found to be co-expressed with lignin biosynthesis genes. Among them, *PvMYB42/85A*, which shared 94% amino acid similarity with *PvMYB42/58B*, showed highest expression in stems (Figure 3A).In situ hybridization indicated that both *PvMYB42/85A* and *PvMYB42/85C* were expressed throughout the stem, including vascular bundles and epidermal cells, with *PvMYB42/85A* being preferentially expressed in xylem vessels (Figure 3Band Supplementary Figure 5). As with PvMYB58/63**,** we generated overexpression and RNAi lines for PvMYB42/85A in transgenic switchgrass. PvMYB42/85A transcript levels exhibited at least a 50-fold increase in selected PvMYB42/85-overexpressing lines, and at least 70% reduction in RNAi lines based on qPCR (Figure 3C and Supplementary Figure 6). A similar phenotype was observed in the PvMYB42/85A overexpressors as in the PvMYB58/63A overexpressors. This comprised reduced plant height (36% reduction on average, Figure 3D, E), increased phloroglucinol staining of stem cross sections (Figure 3F), elevated total lignin content (1.2-fold on average by the AcBr method, Figure 3G; 1.8-fold on

average by the thioacidolysis method, Figure 3H), increased S:G ratio (Figure 3I) and increased transcript levels of *F5H* and the other lignin biosynthesis genes (Figure 3J and Supplementary Figure 9). No comparable opposite changes were detected in the PvMYB42/85-RNAi line although several lignin biosynthesis genes were down-regulated (Supplementary Figure 10). PvMYB42/85A, like PvMYB58/63A, therefore activates lignin biosynthesis via expression of lignin biosynthesis genes including F5H.

PvWRKY12. Neither AtWRKY12 nor its homolog PvWRKY12 appeared in the co-expression network of lignin biosynthesis genes in Arabidopsis and switchgrass. This is consistent with WRKY12 not being a direct regulator of lignin formation (Wang *et al.*, 2010). Quantitative PCR analysis revealed that PvWRKY12 is expressed in leaf blades, leaf sheaths and stems, with the highest transcript abundance in stems (Figure 4A), and in situ hybridization of mature stem tissue sections suggested highest expression in pith parenchyma cells (Figure 4B).

We have previously reported the generation of WRKY12-downregulated transgenic lines of switchgrass using the dominant repressor (DR) strategy (Gallego-Giraldo *et al.*, 2016). These showed a severe to mild phenotype of reduced plant growth and disruption of xylem vessels and tracheids in the stems (Gallego-Giraldo *et al.*, 2016). We here further extend this analysis. Compared with the controls, the relative thickness of the stems relative to the central cavity was increased in PvWRKY12-DR lines, but with decreased overall diameter of the stem(Figure 4Cand Supplementary Figure 11).Phloroglucinol stainingof cross sections of mature stems showed an enhanced lignification in epidermal cells and vascular bundles in the PvWRK12-DR lines (Figure 4D). Furthermore, the thickness of the pith cell wallswas significantly increased in PvWRKY12-DR lines (Figure 4D-F), consistent with previous reports of Arabidopsis and Medicago *wrky12* mutant lines (Wang *et al.*, 2010).

PvSWN2 and PvSND2. PvSWN2A (AP13ITG56117RC_at) and PvSWN2B (KanlowCTG11938_s_at), close homologs of the Arabidopsis secondary cell wall master regulator SND1/NST1/2, showed co-expression with six and eight lignin biosynthesis genes, respectively, in switchgrass (Supplementary Dataset 1). We generated more than ten independent PvSWN2-RNAi lines in switchgrass of which six exhibited strongly reduced expression of the target genes (Supplementary Figure 12). Compared with the vector control, no obvious changes in plant growth, lignin content or stem structure were observed in any of the PvSWN2-RNAi lines, except slightly increased S:G ratio in two lines (Supplementary Figure 12).Quantitative PCR analysis showed that the decreased expression of PvSWN2 in the transgenic lines resulted in no obvious effects on the expression of either secondary cell wall-related transcription factors (PvSWN1, PvMYB4, PvMYB46, PvMYB58/63 and PvMYN42/85) or the lignin biosynthesis gene

F5H (Supplementary Figure 12). This suggests that PvSWN2 may be functionally redundantwith other TFs (eg. PvSWN1) and not exclusively regulate the expression of F5H.

The expression of genes grouped in the SND2 clades (Supplementary Figure 13) show correlation with multiple lignin biosynthesis genes in both Arabidopsis (AT4G28500) and switchgrass (KanlowCTG43583_s_at, AP13CTG15049_s_at, AP13ITG71892_at, AP13ITG74807_s_at and KanlowCTG37619_s_at) (Supplementary Dataset 1). To explore the potential role of PvSND2 in switchgrass secondary wall formation, more than ten PvSND2 RNAi lines were generated and a selection confirmed by qPCR analysis (Supplementary Figure 14). Reduced expression of MYB103, CESA4 and CESA9 (two cellulose biosynthesis genes involved in secondary wall) were detected in the transgenic lines compared with the vector control. Cell wall polysaccharide analysis indicated a reduced content of hemicellulose in one transgenic line (Supplementary Figure 14). In addition, the decreased level of two glycosyl residues, galactose and glucose, were observed in all transgenic lines and in one transgenic line, respectively (Supplementary Figure 14). No significant changes were observed in plant growth, lignin content or stem structure (Supplementary Figure 15). The results suggest that PvSND2 may be involved in secondary wall-related cellulose and hemi-cellulose biosynthesis.

An extended co-expression analysis for selected TFs

To further evaluate the potential role of TFs in secondary wall formation and other biological pathways in switchgrass, we extended the co-expression network by adding 19 and 21 TFs as bait genes in Arabidopsis and switchgrass, respectively. These TFs included WRKY12, SWNs, MYB46/83, MYB4, MYB58/63 and MYB42/85 (Supplementary Table 3). A hierarchical structure of correlated-relationships among the genes was observed in both the Arabidopsis and switchgrass networks (Supplementary Dataset 2).

There were clear similarities in the association between TFs in Arabidopsis and switchgrass (Figure 5). The coordinated expression of AtWRKY12 with AtNST1, AtNST2 and AtSND1 is consistent with the observation that AtWRKY12 can directly regulate the expression of downstream NAC master switches (Wang *et al.*, 2010; Wang *et al.*, 2011). Similarly, PvWRKY12 shows co-expression with PvSWN1 and PvSWN2B (the homologs of AtNST1, AtNST2 and AtSND1). MYB46 is tightly co-expressed with SWN and MYB42/85 and MYB58/63 in both Arabidopsis and switchgrass. This suggests a central role of MYB46, controlled by upstream SWNs and modulating downstream TFs in the transcriptional regulatory program. In contrast, MYB4, the secondary wall repressor, is less co-expressed with other TFs in

switchgrass compared with its homolog in Arabidopsis. We suggest that expression of MYB4 in switchgrass is partially independent of the NAC-MYB46-MYB transcriptional regulatory cascade.

To understand the functional differentiation of genes co-expressed with the above TFs in Arabidopsis and switchgrass, we inferred the biological pathways in which these genes are significantly enriched, using the Fisher exact test (Supplementary Figure 16). Consistent with the involvement of the TFs in secondary wall formation, many co-expressed genes in Arabidopsis and switchgrass are enriched in the functional pathways of cell wall degradation, cell wall modification, cellulose/hemicellulose synthesis and phenylpropanoid synthesis. Besides cell wall-related genes, groups of genes assigned to flavonoid biosynthesis, hormone and photosynthesis pathways were tightly co-expressed with these TFs both species (Figure 5). The appearance of conserved co-expression modules in the Arabidopsis and switchgrass co-expression networks suggests that these TFs may be involved in the coordination and cross-talk between cell wall development and other biological pathways in both species.

RNA-seq analysis of selected TF transgenic lines

To investigate the global changes in gene expression modulated by secondary wall TFs in switchgrass, we performed RNA sequencing on four of the previously generated transgenic switchgrass lines (MYB4-OX, MYB58/63A-OX, MYB42/85A-OX and WRKY12-DR). For each set of transgenic plants, we selected one strongly expressing line, one or two weakly expressing lines, and one control line (Supplementary Table 4), and generated transcriptomes for internode tissues and whole tillers (Supplementary Datasets 3 and 4). Pearson correlation and cluster (PCC) analysis of transcript profiles were used to assess the level of correlation between biological replicates (Supplementary Figure 17). Overall, PCC values ranged from 0.70 to 0.99 within replicates.

Thousands of genes were identified to be differentially expressed in the TF strongly-expressing lines compared with the controls (Figure 6A and Supplementary Dataset 5). The MYB4-OX and MYB58/63A-OX lines displayed the highest abundance of differentially expressed genes in tillers (11,112 genes in MYB4-OX and 7,484 genes in MYB58/63A-OX) and internodes (12,509 genes in MYB4-OX and 11,393 genes in MYB58/63A-OX)(Figure 6A and Supplementary Table 5), consistent with their severe phenotypes compared with the control. Accompanying the opposing lignification phenotypes between MYB4-OX and MYB58/63A-OX, 900 and 1,162 genes, respectively, exhibited opposite expression patterns in tillers and internodes between the two transgenic lines (Supplementary Dataset 3 and Supplementary Dataset 4).

The enrichment analysis of differentially expressed genes on biological pathways showed that the significantly enriched functional categories included phenylpropanoid, cellulose, and cell wall precursor synthesis (Figure 6B).A further survey of the differentially expressed genes grouped in the cell wall-related categories indicated that the four TFs display both commonalities and differences in target genes, and differential levels of regulation on target genes in internode and stem (Supplementary Table 6). PvMYB4 and PvMYB58/63A altered the expression of all lignin biosynthesis genes (except C3´H) in opposite directions in both internodes and tillers, whereas PvMYB42/85A significantly activated lignin biosynthesis genes with the exception of PAL, C3´H and CCoAOMT in tillers. Furthermore, the weakly-overexpressing PvMYB4 and PvMYB58/63 lines exhibited altered expression of 4CL and HCT, and CCR and F5H, respectively (Supplementary Table 5). Thus, PvMYB4, PvMYB58/63A and PvMYB42/85 can be considered as regulators of lignin biosynthesis genes with different preferential targets.

An overall decrease of gene expression involved in flavonoid and both primary and secondary cell wall cellulose biosynthesis was detected in PvMYB4-OX lines in internodes, but this effect was much weaker or did not occur at all in tillers (Supplementary Table 6). In contrast, PvMYB58/63A overexpression strongly upregulated genes involved in flavonoid and secondary cell wall cellulose biosynthesis in both internodes and tillers (Supplementary Table 6).

We next assessed the effects of the four TFs on expression of other TFs(Supplementary Tables 5 and 6). The RNAseq data confirmed that expression of the MYB4, MYB58/63, MYB42/85 and WRKY12 target genes was strongly altered in the corresponding transgenic lines, consistent with qPCR analysis. The transcript level of SWN1, SWN2, PvMYB46, MYB42/85, MYB32, MYB103 and SND2 was significantly increased in internodes and tillers of the PvMYB58/63-OX line, whereas the expression of these genes was decreased in internodes but not (or less) affected in tillersof the PvMYB4-OX and PvMYB42/85-OX line. This suggests an opposite direction of regulation by MYB58/63, and MYB4 and MYB42/85 in the internode, but not in the tillers. Furthermore, MYB58/63, MYB4 and MYB42/85 are considered as downstream targets of SWN1, SWN2 and MYB46 (Zhong *et al.*, 2011; Zhong *et al.*, 2015; McCarthy *et al.*, 2009; Zhong *et al.*, 2007).Our data suggest a “feed-back” regulation in the NAC-MYB-based transcriptional network of secondary wall development in switchgrass. Interestingly, no significant changes in expression of SWNs were observed in the WRKY12-DR lines compared with the control (Supplementary Tables 5 and 6), a result that was further confirmed by q-PCR (Supplementary Figure 18). To explore the potential target genes of PvWRKY12, we combined the RNA-seq and co-expression analysis to discover 125 genes that were differentially expressed in both internodes and tillers of PvWRKY12-DR lines and showed correlated expression with PvWRKY12 in large scale- microarray

datasets (Supplementary Dataset 6). Among them, we found an increased expression of a homolog of the bZIP63 transcription factor (Pavir.Cb00258) and decreased expression of a homolog of the bZIP44 transcription factor (Pavir.Fb01406) and two WRKY41 homologs (Pavir.Bb02088 and Pavir.J16737) in the PvWRKY12DR line. Arabidopsis bZIP63, bZIP44 and WRKY41 are involved in glucose signaling, cell wall modification and multiple developmental processes, respectively (Iglesias-Fernandez *et al.*, 2013; Matiolli *et al.*, 2011; Duan *et al.*, 2018). We suggest these genes could be considered as candidate targets of PvWRKY12.

Promoter transactivation assays

To directly confirm the transcriptional regulation of lignin biosynthesis genes by TFs, we performed transient promoter transactivation analysis using a promoter-luciferase reporter system. The reporter vectors, which contained the promoters of switchgrass *F5H* and *COMT* driving the firefly luciferase gene, were co-transfected into the protoplasts with four separate TF effectors (MYB4, MYB58/63A, MYB42/85A and MYB46) under the control of the constitutive cauliflower mosaic virus 35S promoter (Figure 6C and Supplementary Figure 19). Expression from the PvCOMT promoter was significantly activated by PvMYB58/63A, PvMYB42/85A and PvMYB46, and weakly down-regulated by PvMYB4. In contrast, the PvF5H promoter was significantly but only weakly activated by PvMYB42/85A and PvMYB46, but was not activated byPvMYB58/63A or repressed by PvMYB4 (Figure 6C and Supplementary Figure 19). These results directly confirm the ability of PvMYB4, PvMYB58/63 and PvMYB42/85 to regulate the expression of lignin biosynthesis genes, but the observed in planta modulation of F5H expression on overexpressing PvMYB4 and PvMYB58/63A may not be through direct recognition of its promoter region by these TFs.

Discussion

An improved strategy to identify regulators and their target candidates using co-expression analysis

Co-expression analysis reflects the correlated expression of pairs of genes. It is a powerful tool to decipher transcriptional regulatory relationships because the expression of TFs tends to be transcriptionally coordinated with that of their functional targets (Hansen *et al.*, 2014). The first step in co-expression network construction is to measure the similarity of expression values for genes in a pairwise fashion (Serin *et al.*, 2016). Each method for determining the similarity score displays its own specific features. For example, Pearson correlation, the most popular method in co-expression analysis, exclusively represents linear relationships in gene pairs (Serin *et al.*, 2016). However, the expression

levels of a gene and its regulator could be associated by non-linear correlation (Usadel *et al.*, 2009). Here we utilize a bi-clustering algorithm termed QUBIC to assess the strength of the correlated relationships. QUBIC is effective and efficient in detecting monotonic relationships between pairs of variables under both all and partial conditions (Li *et al.*, 2009).

Using this method, we constructed large- scale co-expression networks of switchgrass and Arabidopsis using lignin biosynthesis genes as bait genes. The networks contained 645 and 1274 TFs in Arabidopsis and switchgrass, respectively, which could be associated with secondary wall development. The distribution of TF families was consistent with a previous survey of co-expressed TFs in secondary wall formation in Arabidopsis, rice, maize and sugarcane identified by Pearson correlation, Spearman correlation and mutual rank methods(Hansen *et al.*, 2014; Ruprecht *et al.*, 2011; Hirano *et al.*, 2013a; Cassan-Wang *et al.*, 2013; Ferreira *et al.*, 2016), supporting the validity of our co-expression method to identify co-expression modules in large-scale datasets. Of these conserved TF families, multiple TF subclades such as SWNs, SNDs, MYB46, MYB4/32, MYB42/85 and MYB58/63 have been identified previously as playing a role in secondary cell wall regulation in Arabidopsis (Nakano *et al.*, 2015; Zhong and Ye, 2015; Zhong *et al.*, 2010). The experimental validation of these TFs using transgenic technology in the present work confirmed their role in switchgrass secondary wall development. To explore the potential targets of the TFs, we extended the co-expression networks by adding 21 TFs as bait genes and conducted transcriptome sequencing on TF transgenic lines. We found that many genes co-expressed with certain TFs that appeared to be differentially expressed in the corresponding TF transgenic line. We suggest that comparative co-expression analysis across plant species is a feasible strategy for initial investigation of transcriptional regulators using structural genes, and for target genes using transcriptional regulators.

Commonalities and differences in transcriptional regulation of secondary wall formation in switchgrass and Arabidopsis

Grasses and dicots, divergent after the establishment of vascular plants, may share conserved functionalities in the transcriptional networks that regulate their secondary wall formation (Zhong *et al.*, 2010; Rao and Dixon, 2018). The appearance of orthologous TFs in co-expression networks of lignin biosynthesis, and the similarity of network structure in the TF-extending co-expression networks in Arabidopsis and switchgrass suggests that these TFs may play similar roles in secondary wall formation in both species. Here we investigated the roles of MYB4, MYB58/63, MYB42/85, WRKY12, SWN2 and SND2 in the secondary cell wall biosynthesis program in switchgrass. The phenotypes of the transgenic

lines with altered expression of these TFs displayed many similarities to those of the corresponding lines in Arabidopsis. For example, overexpressing either AtMYB58/63 in Arabidopsis (Zhou *et al.*, 2009) or PvMYB58/63 in switchgrass caused reduced plant height, increased lignin content, and elevated expression of monolignol biosynthesis, and downregulating PvWRKY12 in switchgrass resulted in enhancement of pith cell walls, similar to the phenotype of the *atwrky12* mutant (Wang *et al.*, 2010). However, we also detected several differences in switchgrass secondary wall development compared to the situation in Arabidopsis.

Although both PvMYB58/63 and AtMYB58/63 function as activators of lignin biosynthesis, an increased expression of secondary wall-associated cellulose synthase and xylan synthase genes, as well as flavonoid biosynthesis genes, was observed here in PvMYB58/63-OX plants but not in Arabidopsis MYB58/63 overexpressors(Zhou *et al.*, 2009). Considering that OsMYB58/63 could also upregulate secondary wall-related cellulose synthase genes in transactivation assay (Noda *et al.*, 2015) and the overexpression of SbMYB60 (the ortholog of AtMYB58/63) in sorghum altered cell wall composition of cellulose and xylan(Scully *et al.*, 2016), we suggest that the MYB58/63 clade in grasses may function as a broader activator regulating both lignin and other two secondary wall-component (cellulose and hemicellulose)biosynthesis, rather than possessing the “lignin- specific regulator” function ascribed to AtMYB58/63 in Arabidopsis.

Differences in regulation of lignin biosynthesis genes were also observed between orthologous Arabidopsis and switchgrass TF genes. Differences in co-expression modules suggested that Arabidopsis and switchgrass may operate differently to regulate F5H, the entry point to S lignin biosynthesis; Arabidopsis may assign an additional group of TFs to regulate the expression of F5H, whereas switchgrass may utilize the common TFs that regulate the expression of other lignin biosynthesis genes. In Arabidopsis, the transcript abundance of F5H was influenced by SND1 and MYB103 (Ohman *et al.*, 2013; Zhao *et al.*, 2010). In contrast, while decreased expression of F5H was not observed in PvSWN2-DR lines, we found that PvMYB42/85 and PvMYB46 directly activated the expression of PvF5H, and thatsignificantly altered expression of PvF5H occurredin PvMYB4 and PvMYB58/63 overexpressor lines. All these four MYB TFs had the capability to modulate other lignin biosynthesis genes, which may contribute to the correlated expression of F5H with other lignin biosynthesis genes in switchgrass.

In conclusion, switchgrass displays a complex transcriptional network for regulation of its secondary wall biosynthetic program. This shares many common features with the secondary wall regulatory program in Arabidopsis, but also exhibits several differences. The network, supported by trransgenesis, RNA-seq analysis, transactivation and co-expression analyses, is summarized in Figure

7.Our analyses identify gene targets for the potential modification of multiple cell wall components in swicthgrass.

Experimental Procedures

Co-expression analysis

For co-expression analysis, the Arabidopsis microarray database was downloaded from NCBI GEO website (GSE34188). The switchgrass microarray database was accessed from the Noble Foundation Switchgrass Functional Genomics Server at http://switchgrassgenomics.noble.org/ (Zhang *et al.*, 2013). The RMA method is applied to normalize these Affymetrix microarray data. A new approach, QUBIC, based on our previous Bi-cluster method (Li *et al.*, 2009) was used to detect the co-expression modules in Arabidopsis and switchgrass. The details of this co-expression method are provided in Supplementary Methods.

Plant materials, transformation and growth conditions

Switchgrass was grown in the greenhouse under 28°C with 16 h light. For plant transformation, the pANIC vector (Mann *et al.*, 2012) was used for plasmid construction. Briefly, for overexpression construct, the full length of coding sequence was cloned into pANIC vector under the control of the ZmUbi1 promoter (Mann *et al.*, 2012); for RNAi construct, the partial length of coding sequence was cloned into pANIC-RNAi binary vectorunder the control of the ZmUbi1 promoter(Mann *et al.*, 2012); for dominant repressor construct, the PvWRKY12-DR sequence was cloned and inserted into the destination vector pANIC10A as described in (Gallego-Giraldo *et al.*, 2016). All primers used for cloning are listed in Supplementary Table 7. The switchgrass ST2 line was used for stable transformation. Agrobacterium-mediated transformation in switchgrass was performed as described previously(Shen *et al.*, 2012; Xi *et al.*, 2009).

Quantitative reverse transcription

RNA isolation, cDNA synthesis, and real-time RT-PCR were performed as previously described (Rao *et al.*, 2016). For determination of tissue-specific expression, total RNA was isolated from leaf, leaf sheath and stem of switchgrass at the reproduction (R1) developmental stage. qRT-PCR was performed with three

biological replicates for each gene tested. Data were collected using an optical 384 well plate with QuantStudio™ 6 Flex (Applied Biosystems). The *Ubi* gene was used as a reference for relative quantification of transcript levels. All primers designed for qRT-PCT are listed in Supplemental Table 6.

In situ hybridization

Samples of at least ten switchgrass plants were harvested for in situ hybridization. The tissue preparation including fixation, dehydration, and paraffin embedding was performed according to Long's protocol (http://www.its.caltech.edu/~plantlab/protocols/insitu.pdf). Pre-hybridization, hybridization, washing and imaging were performed as described previously (Rao *et al.*, 2016).

Determination of lignin content

Switchgrass whole tillers were collected to prepare cell wall residues (CWR) by sequential exaction with chloroform/methanol (1:1), 100% methanol, 50% methanol, and water (three times each). Twenty-five milligrams of CWR was used for lignin analysis. The acetyl bromide method and thioacidolysis method followed by GC-MS were used to determine total lignin content and quantify lignin-derived monomers, respectively. All analytical methods were performed as previously described (Shen *et al.*, 2009).

Cell wall polysaccharide analysis

Switchgrass whole tillers were ground and washed in organic solvent to prepare the alcohol insoluble residue (AIR) as described previously (ref). Cell walls were obtained after removing starch from the AIR using Novazyme specific enzymes. Cell walls were treated with 2N trifluoroacetic acid (TFA) to solubilize and hydrolyze non-cellulosic polysaccharides. The amount of sugar solubilized was determined using the phenol-sulfuric acid colorimetric assay (ref). The identity of the solubilized sugars was determined by GC-FID analysis after derivatization of the sugars to their corresponding alditol acetates (ref). The insoluble material after the TFA hydrolysis was washed extensively and the pellets were used to determine the cellulose content. The cellulose was hydrolyzed using the Shaeman hydrolysis method (ref) and the amount of glucose solubilized was determined by phenol-sulfuric acid colorimetric assay.

Histochemical staining

Switchgrass internode samples were collected and cut with a vibratome (Thermo Fisher Scientific, Germany). Internode 2 of E4 tillers were collected for Phloroglucinol-HCL staining as previouslydescribed (Shen *et al.*, 2009).

Next generation sequencing and RNA-seq analysis

The whole tillers at R1 stage (Shen *et al.*, 2009) were harvested as "tiller" samples and a middle section from internode 4 stem were cut from the tiller as "internode" samples. All samples were taken at same time of day (between 1-3 pm) and stored at -80°C for later RNA extraction. Each sample was submitted for 150 bp paired-end sequencing to generate 40-50 million reads. Each sample group has at least two biological replicates. Paired-end Illumina reads after filtering and trimming treatment were mapped to the Switchgrass genome *Panicumvirgatum*v1.1 (http://phytozome.jgi.doe.gov/) using Bowtie 2 (v2.3.2.0) (Langmead and Salzberg, 2012)and TopHat (v2.2.1) (Kim *et al.*, 2013)with default parameters. Differentially expressed genes between the controls and transgenic lines were determined using Cufflinks (v2.2.1) (Trapnell *et al.*, 2013)with default setting of adjusted P-value < 0.05.As previously described (Rao *et al.*, 2016), classification for differentially expressed genes was based on MapManmappings of their Arabidopsis homologs (Thimm *et al.*, 2004). Significant functional enrichment was subjected to Fisher's exact test with Benjamini–Hochberg multiple testing correction [false discovery rate (FDR) $\leqslant$0.1].

Transactivation assay

The effector constructs were generated by inserting coding sequences of MYB TFs after the 35S promoter of the Gateway overexpression vector P2GW7 (Karimi *et al.*, 2002). The reporter constructs were generated by inserting promoters of lignin biosynthesis genes into the vector P2GWL7 (Wang *et al.*, 2010). Primers used for plasmid construction are listed in Supplementary Table 7. The effector and reporter plasmids were co-transfected into Arabidopsis leaf protoplasts isolated according to a previously report(Wang *et al.*, 2010). Promoter activities are presented as Firefly LUC/Renilla LUC activities, and normalized to the value obtained from protoplasts transformed with empty effector vector. The data were the average of three biological replicates.

Statistical analysis

Experimental data were subjected to statistical treatment by the Student's *t*-test (Microsoft office Excel 2013). Significant difference between two groups was determined by $p < 0.05$ and indicated by asterisks above bars.

Accession numbers

The data sets supporting the results of this article are available in the NCBI Sequence Read Archive (SRA) repository, NCBI SRA accession no. XXXXXXX.

Acknowledgments

The *Panicumvirgatum* genome v1.1 sequence data were produced by the US Department of Energy Joint Genome Institute. We thank Dr Rajeev Azadfor critical reading of the manuscript. This work was supported by the BioEnergy Science Center and the Center for Bioenergy Innovation (Oak Ridge National Laboratory), US Department of Energy (DOE) Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Conflict of interest

The authors declare they have no conflict of interest

References

Bouton JH (2007) Molecular breeding of switchgrass for use as a biofuel crop. Current Opinion in Genetics and Development **17:** 553-558

Cassan-Wang H, Goué N, Saidi M, Legay S, Sivadon P, Goffner D, Grima-Pettenati J (2013) Identification of novel transcription factors regulating secondary cell wall formation in Arabidopsis. Frontiers in Plant Science **4:** 189

Duan S, Wang J, Gao C, Jin C, Li D, Peng D, Du G, Li Y, Chen M (2018) Functional characterization of a heterologously expressed *Brassica napus* WRKY41-1 transcription factor in regulating anthocyanin biosynthesis in *Arabidopsis thaliana*. Plant Science **268:** 47-53

Dumitrache A, Natzke J, Rodriguez M, Yee KL, Thompson OA, Poovaiah CR, Shen H, Mazarei M, Baxter HL, Fu C et al. (2017) Transgenic switchgrass (*Panicum virgatum* L.) targeted for reduced recalcitrance to bioconversion: a 2-year comparative analysis of field-grown lines modified for target gene or genetic element expression. Plant Biotechnology Journal **15:** 688-697

Ferreira SS, Hotta CT, Poelking VGdC, Leite DCC, Buckeridge MS, Loureiro ME, Barbosa MHP, Carneiro MS, Souza GM (2016) Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. Plant Molecular Biology **91:** 15-35

Gallego-Giraldo L, Shadle G, Shen H, Barros-Rios J, Fresquet Corrales S, Wang H, Dixon RA (2016) Combining enhanced biomass density with reduced lignin level for improved forage quality. Plant Biotechnology Journal **14:** 895-904

Hanada K, Higuchi-Takeuchi M, Okamoto M, Yoshizumi T, Shimizu M, Nakaminami K, Nishi R, Ohashi C, Iida K, Tanaka M et al. (2013) Small open reading frames associated with morphogenesis are hidden in plant genomes. Proceedings of the National Academy of Sciences USA **110:** 2395-2400

Hansen BO, Vaid N, Musialak-Lange M, Janowski M, Mutwil M (2014) Elucidating gene function and function evolution through comparison of co-expression networks of plants. Frontiers in Plant Science **5:** 394

Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science **315:** 804-807

Hirano K, Aya K, Morinaka Y, Nagamatsu S, Sato Y, Antonio BA, Namiki N, Nagamura Y, Matsuoka M (2013a) Survey of genes involved in rice secondary cell wall formation through a co-expression network. Plant and Cell Physiology **54:** 1803-1821

Hirano K, Kondo M, Aya K, Miyao A, Sato Y, Antonio BA, Namiki N, Nagamura Y, Matsuoka M (2013b) Identification of transcription factors involved in rice secondary cell wall formation. Plant and Cell Physiology **54:** 1791-1802

Iglesias-Fernandez R, Barrero-Sicilia C, Carrillo-Barral N, Onate-Sanchez L, Carbonero P (2013) *Arabidopsis thaliana* bZIP44: a transcription factor affecting seed germination and expression of the mannanase-encoding gene AtMAN7. Plant Journal **74:** 767-780

Karimi M, Inze D, Depicker A (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. Trends in Plant Science **7:** 193-195

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology **14:** R36

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods **9:** 357-359

Li G, Ma Q, Tang H, Paterson AH, Xu Y (2009) QUBIC: a qualitative biclustering algorithm for analyses of gene expression data. Nucleic Acids Research **37:** e101-e101

Mann DG, Lafayette PR, Abercrombie LL, King ZR, Mazarei M, Halter MC, Poovaiah CR, Baxter H, Shen H, Dixon RA et al. (2012) Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. Plant Biotechnology Journal **10:** 226-236

Matiolli CC, Tomaz JP, Duarte GT, Prado FM, Del Bem LE, Silveira AB, Gauer L, Correa LG, Drumond RD, Viana AJ et al. (2011) The Arabidopsis bZIP gene AtbZIP63 is a sensitive integrator of transient abscisic acid and glucose signals. Plant Physiology **157:** 692-705

McCarthy RL, Zhong R, Ye ZH (2009) MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant and Cell Physiology **50:** 1950-1964

Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M (2015) NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. Frontiers in Plant Science **6:** 288

Noda S, Koshiba T, Hattori T, Yamaguchi M, Suzuki S, Umezawa T (2015) The expression of a rice secondary wall-specific cellulose synthase gene, OsCesA7, is directly regulated by a rice transcription factor, OsMYB58/63. Planta **242:** 589-600

Ohman D, Demedts B, Kumar M, Gerber L, Gorzsas A, Goeminne G, Hedenstrom M, Ellis B, Boerjan W, Sundberg B (2013) MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in Arabidopsis stems. Plant Journal **73:** 63-76

Pauly M, Keegstra K (2010) Plant cell wall polymers as precursors for biofuels. Current Opinion in Plant Biology **13:** 305-312

Ragauskas AJ, Beckham GT, Biddy MJ, ChandraR, Chen F, Davis MF, DavisonBH, Dixon RA, GilnaP, KellerM, Langan P, NaskarAK, SaddlerJN, TschaplinskiTJ, Tuskan GA, Wyman CE (2014). Lignin valorization: improving lignin processing in the biorefinery. Science344: 1246843. DOI: 10.1126/science.1246843

Rao X, Dixon RA (2018) Current models for transcriptional regulation of secondary cell wall biosynthesis in grasses. Frontiers in Plant Science:**9**doi: 10.3389/fpls.2018.00399

Rao X, Lu N, Li G, Nakashima J, Tang Y, Dixon RA (2016) Comparative cell-specific transcriptomics reveals differentiation of C4 photosynthesis pathways in switchgrass and other C4 lineages. Journal of Experimental Botany **67:** 1649-1662

Rao X, Shen H, Pattathil S, Hahn MG, Gelineo-Albersheim I, Mohnen D, Pu Y, Ragauskas AJ, Chen X, Chen F et al. (2017) Dynamic changes in transcriptome and cell wall composition underlying brassinosteroid-mediated lignification of switchgrass suspension cells. Biotechnology for Biofuels **10:** 266

Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, Persson S (2011) Large-scale co-expression approach to dissect secondary cell wall formation across plant species. Frontiers in Plant Science **2:** 23

Ruprecht C, Persson S (2012) Co-expression of cell-wall related genes: new tools and insights. Frontiers in Plant Science **3:** 83

Scully ED, Gries T, Sarath G, Palmer NA, Baird L, Serapiglia MJ, Dien BS, Boateng AA, Ge Z, Funnell-Harris DL et al. (2016) Overexpression of SbMyb60 impacts phenylpropanoid biosynthesis and alters secondary cell wall composition in Sorghum bicolor. Plant Journal **85:** 378-395

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W (2016) Learning from co-expression networks: possibilities and challenges. Frontiers in Plant Science **7:** 444

Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang Z, Chen F (2009) Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. BioEnergy Research **2:** 233-245

Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DG, Wang H, Jackson L, Tang Y, Stewart CN, Jr. et al. (2012) Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. New Phytologist **193:** 121-136

Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu CX, Pu YQ, Rudis MR, Tang YH, Xiao XR, Jackson L et al. (2013) A genomics approach to deciphering lignin biosynthesis in switchgrass. Plant Cell **25:** 4342-4361

Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R et al. (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature **517:** 571-575

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant Journal **37:** 914-939

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology **31:** 46-53

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant, Cell and Environment **32:** 1633-1651

Vogel J (2008) Unique aspects of the grass cell wall. Current Opinion Plant Biology**11:** 301-307

Wang H-Z, Dixon RA (2012) On–off switches for secondary cell wall biosynthesis. Molecular Plant **5:** 297-303

Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA (2010) Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. Proceedings of the National Academy of Sciences USA **107:** 22338-22343

Wang H, Zhao Q, Chen F, Wang M, Dixon RA (2011) NAC domain function and transcriptional control of a secondary cell wall master switch. Plant Journal **68:** 1104-1114

Xi Y, Ge Y, Wang ZY (2009) Genetic transformation of switchgrass. Methods in Molecular Biology **581:** 53-59

Zhang JY, Lee YC, Torres-Jerez I, Wang M, Yin Y, Chou WC, He J, Shen H, Srivastava AC, Pennacchio C et al. (2013) Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (*Panicum virgatum* L.). Plant Journal **74:** 160-173

Zhao Q, Dixon RA (2011) Transcriptional networks for lignin biosynthesis: more complex than we thought? Trends in Plant Ccience **16:** 227-233

Zhao Q, Wang H, Yin Y, Xu Y, Chen F, Dixon RA (2010) Syringyl lignin biosynthesis is directly regulated by a secondary cell wall master switch. Proceedings of the National Academy of Sciences USA **107:** 14496-14501

Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye ZH (2011) Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. Plant and Cell Physiology **52:** 1856-1871

Zhong R, Lee C, Ye ZH (2010) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. Trends in Plant Science **15:** 625-632

Zhong R, Richardson EA, Ye ZH (2007) The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. Plant Cell **19:** 2776-2792

Zhong R, Ye ZH (2015) Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant and Cell Physiology **56:** 195-214

Zhong R, Yuan Y, Spiekerman JJ, Guley JT, Egbosiuba JC, Ye ZH (2015) Functional characterization of NAC and MYB transcription factors involved in regulation of biomass production in switchgrass (*Panicum virgatum*). PloS one **10:** e0134611

Zhou J, Lee C, Zhong R, Ye ZH (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. Plant Cell **21:** 248-266

Figure legends

Figure 1. Co-expression networks for lignin biosynthesis in Arabidopsis and switchgrass. (A) Distribution of TF families that are co-expressed with lignin biosynthesis genes. (B) Selected TF families in the co-expression network of lignin biosynthesis. Colors represent TF families. Circles and triangles represent genes with and without homologs, respectively, appearing in both the Arabidopsis and switchgrass networks.

Figure 2. Phenotype of transgenic switchgrass overexpressing PvMYB58/63A. (A) qRT-PCR analysis of PvMYB58/63A and PvMYB58/63C transcript levels in leaf blade, leaf sheath and stem. The transcript level in leaf blade was set to 1. (B) In situ hybridization of PvMYB58/63A in switchgrass stem. VT, vascular tissue; E, epidermis. Bars, 100 µm. (C) qRT-PCR analysis of PvMYB58/63A transcripts in control and PvMYB58/63A-OX transgenic switchgrass. UBQ, ubiquitin. (D) Representative photograph of control and PvMYB58/63A-OX lines. (E) Plant height for control and PvMYB58/63A-OX lines. (F) Phloroglucinol-HCl staining of stem cross-sections of PvMYB58/63A-OX transgenic switchgrass. Bars, 100 µm. (G) Total lignin content of whole tillers determined by the AcBr method. (H) Total lignin content and (I) S/G ratio determined by thioacidolysis. (J) qRT-PCR analysis of ferulate 5-hydroxylase (F5H) transcripts in control and PvMYB58/63A-OX transgenic switchgrass. All data are means ± SE (n = 3). Significant differences from the equivalent control were determined by the Student's t test and are represented by a single asterisk ($p < 0.05$) or double asterisk ($p < 0.01$).

Figure 3. Phenotype of transgenic switchgrass overexpressing PvMYB42/85A. (A) qRT-PCR analysis of PvMYB42/85A transcript levels in leaf blade, leaf sheath and stem. The transcript level in leaf blade was set to 1. (B) In situ hybridization of PvMYB42/85A in switchgrass stem. Bars, 100 µm. (C) qRT-PCR analysis of PvMYB42/85A transcripts in control and PvMYB42/85A-OX transgenic switchgrass. UBQ, ubiquitin. (D) Representative photograph of control and PvMYB42/85A-OX lines. (E) Plant height for control and PvMYB42/85A-OX lines. (F) Phloroglucinol-HCl staining of stem cross-sections of PvMYB42/85A-OX transgenic switchgrass. Bars, 100 µm. (G) Total lignin content of whole tillers determined by the AcBr method. (H) Total lignin content and (I) S/G ratio determined by thioacidolysis. (J) qRT-PCR analysis of F5H transcripts in control and PvMYB42/85A-OX transgenic switchgrass. All data are means ± SE (n = 3). Significant differences from the equivalent control were determined by the Student's t test and are represented by a single asterisk ($p < 0.05$) or double asterisk ($p < 0.01$).

Figure 4. Phenotype of transgenic switchgrass with PvWRKY12 downregulated by the dominant repressor strategy (Gallego-Giraldo *et al.*, 2016). (A) qRT-PCR analysis of PvWRKY12 transcript levels in leaf blade, leaf sheath and stem. The transcript level in leaf blade was set to 1. (B) In situ hybridization of PvWRKY12 in switchgrass stem. Bars, 100 µm. (C) Representative photograph of stem sections from control and PvWRKY12-DR lines. (D) Phloroglucinol-HCl staining of stem cross-sections of PvWRKY12-DR transgenic switchgrass. The arrows indicate pith cell wall. Bars, 100 µm. (E) Stem cross-sections of PvWRKY12-DR transgenic switchgrass. The arrows indicate pith cell wall. Bars, 10 µm. (F) Measurement of pith cell wall thickness in stems of PvWRKY12-DR transgenic switchgrass. E4I2, internode 2 of E4 tillers for (C to F). All data are means ± SE (n = 3). Significant differences from the equivalent control were determined by the Student's t test and are represented by a single asterisk ($p < 0.05$) or double asterisk ($p < 0.01$).

Figure 5. Extended co-expression network of Arabidopsis and switchgrass using selected transcription factors as baits. Genes involved in selected biological functions (cell wall, flavonoid, hormone, phenylpropanoid and photosynthesis) were present in the network. Red lines represent the correlated relationship between bait genes.

Figure 6. Overview of RNA-seq analysis of switchgrass lines with modified expression of TFs, and ability of TFs to activate lignin gene promoters. (A) Venn diagrams showed the number of genes differentially

expressed in tillers and internodes of PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR transgenic lines compared with control. (B) Functional distribution of genes differentially expressed in tillers and internodes of PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR transgenic lines. Blue represents functional groups that were significantly enriched (determined by Fisher exact test, FDR<0.1). (C) Trans-activation of the PvCOMT and PvF5H promoters by PvMYB4, PvMYB46, PvMYB58/63A and PvMYB42/85A. Firefly luciferase activities were quantified and normalized to Renilla luciferase activities. The activities in the protoplasts transfected with promoter-reporter construct and an empty effector construct was set to 1.All data are means ± SE (n = 3). Significant differences forthe activation of PvCOMT and PvF5H promoters were determined by the Student's t test and are represented by a single asterisk ($p < 0.05$) or double asterisk ($p < 0.01$).

Figure 7. Proposed transcriptional network for secondary cell wall formation in switchgrass. Yellow and green colors represent negative and positive regulators, respectively. Arrows and bars at the ends of lines represent positive and negative transcriptional regulation, respectively. Dashed line indicates co-expression relationship.

Supplementary Material

Supplementary Methods. The QUBIC method for co-expression analysis.

Supplementary Figures

Supplementary Figure 1. Diagrammatic representation of the QUBIC algorithm.

Supplementary Figure 2. Phylogenetic trees of MYB, NAC, bHLH, ERF and WRKY TFs that appear in Arabidopsis and switchgrass co-expression networks of lignin biosynthesis.

Supplementary Figure 3. Co-expression network of MYB and NAC TFs with lignin biosynthesis genes in Arabidopsis and switchgrass.

Supplementary Figure 4. Phylogenetic tree of MYB58/63 and MYB42/85 orthologs from Arabidopsis, Medicago, poplar, switchgrass, maize, rice, sorghum and *Brachypodium*.

Supplementary Figure 5. In situ hybridization of PvMYB58/63C and PvMYB42/85C in switchgrass stem.

Supplementary Figure 6. qRT-PCR analysis of target transcripts in leaf of PvMYB58/63-RNAi and PvMYB42/85-RNAi transgenic switchgrass.

Supplementary Figure 7.qRT-PCR analysis of lignin biosynthesis genes in leaf of PvMYB58/63A-OX transgenic switchgrass.

Supplementary Figure 8. Phenotype of PvMYB58/63-RNAi transgenic lines.

Supplementary Figure 9.qRT-PCR analysis of lignin biosynthesis genes in leaf of PvMYB42/85A-OX transgenic switchgrass.

Supplementary Figure 10. Phenotype of PvMYB42/85-RNAi transgenic lines.

Supplementary Figure 11. Stem outer diameters and stem radial thickness in PvWRKY12-DR transgenic lines.

Supplementary Figure 12. Phenotype of PvSWN2-RNAi transgenic lines.

Supplementary Figure 13. Phylogenetic tree of NST and SND2 orthologs from Arabidopsis, poplar, maize, rice and switchgrass.

Supplementary Figure 14. Cell wall-related gene expression and cell wall component analysis in PvSND2-RNAi transgenic switchgrass.

Supplementary Figure 15. Phenotype of PvSND2-RNAi transgenic lines.

Supplementary Figure 16. Functional distribution of genes co-expressed with TFs in Arabidopsis and **switchgrass.**

Supplementary Figure 17. Correlation matrix of switchgrass transcriptomes determined by the Pairwise Pearson correlation coefficients (PCC) method.

Supplementary Figure 18.qRT-PCR analysis of SWN1 and SWN2 in PvWRKY12DR transgenic switchgrass.

Supplementary Figure 19. Trans-activation assays of the PvCOMT and PvF5H promoters by PvMYB4.

Supplementary Tables

Supplementary Table 1. Lignin biosynthesis genes as bait genes for co-expression analysis in Arabidopsis and Switchgrass

Supplementary Table 2. Numbers of TFs of different families that are co-expressed with lignin biosynthesis genes in Arabidopsis and Switchgrass.

Supplementary Table 3. Secondary wall-related TFs as bait genes for co-expression analysis in Arabidopsis and Switchgrass.

Supplementary Table 4. Numbers of differentially expressed genes in RNA-seq analysis of PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR switchgrass lines.

Supplementary Table 5. Expression of cell wall-related genes in all switchgrass transgenic lines.

Supplementary Table 6. Differential expression of secondary cell wall-related genes in tillers and internodes of PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR transgenic switchgrass lines compared with control.

Supplementary Table 7. Sequences for the gene-specific primers used in this work.

Supplementary Datasets

Supplementary Dataset 1. Co-expression analysis of lignin biosynthesis genes in Arabidopsis and Switchgrass.

Supplementary Dataset 2. Co-expression analysis of secondary wall-related transcription factors in Arabidopsis and Switchgrass.

Supplementary Dataset 3. Gene expression profiling (FPKM) from internodes of switchgrass PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR lines.

Supplementary Dataset 4. Gene expression profiling (FPKM) from tillers of switchgrass PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR lines.

Supplementary Dataset 5. List of differentially expressed genes from internodes and tillers of switchgrass PvMYB4OX, PvMYB58/63OX, PvMYB42/85OX and PvWRKY12DR lines compared with their equivalent control ($P < 0.05$).

Supplementary Dataset 6. List of genes that are both co-expressed with PvWRKY12 and differentially expressed in PvWRKY12DR line.