New Phytologist

The *Populus* ARBORKNOX1 homeodomain transcription factor regulates woody growth through binding to evolutionarily conserved target genes of diverse function

Lijun Liu[1], Matthew Zinkgraf[1], H. Earl Petzold[2], Eric P. Beers[2], Vladimir Filkov[3] and Andrew Groover[1,4]

[1]Pacific Southwest Research Station, USDA Forest Service, Davis, CA 95618, USA; [2]Department of Horticulture, Virginia Tech, Blacksburg, VA 24061, USA; [3]Department of Computer Science, University of California, Davis, CA 95618, USA; [4]Department of Plant Biology, University of California, Davis, CA 95618, USA

Author for correspondence:
Andrew Groover
Tel: +1 530 759 1738
Email: agroover@fs.fed.us

Received: *23 August 2014*
Accepted: *1 October 2014*

New Phytologist (2015) **205:** 682–694
doi: 10.1111/nph.13151

Key words: chromatin immunoprecipitation sequencing (ChIP-seq), forest tree, *Populus*, transcriptional regulation, wood formation.

Summary

- The class I KNOX homeodomain transcription factor ARBORKNOX1 (ARK1) is a key regulator of vascular cambium maintenance and cell differentiation in *Populus*. Currently, basic information is lacking concerning the distribution, functional characteristics, and evolution of ARK1 binding in the *Populus* genome.
- Here, we used chromatin immunoprecipitation sequencing (ChIP-seq) technology to identify ARK1 binding loci genome-wide in *Populus*. Computational analyses evaluated the distribution of ARK1 binding loci, the function of genes associated with bound loci, the effect of ARK1 binding on transcript levels, and evolutionary conservation of ARK1 binding loci.
- ARK1 binds to thousands of loci which are highly enriched proximal to the transcriptional start sites of genes of diverse functions. ARK1 target genes are significantly enriched in paralogs derived from the whole-genome salicoid duplication event. Both ARK1 and a maize (*Zea mays*) homolog, KNOTTED1, preferentially target evolutionarily conserved genes. However, only a small portion of ARK1 target genes are significantly differentially expressed in an *ARK1* over-expression mutant.
- This study describes the functional characteristics and evolution of DNA binding by a transcription factor in an undomesticated tree, revealing complexities similar to those shown for transcription factors in model animal species.

Introduction

Secondary (radial) growth in woody stems is the result of coordinated cell division within the meristematic cambial zone, and differentiation of cambial daughter cells within inner bark and wood tissues (Larson, 1994). Several lines of evidence support the notion that transcriptional regulation is a principal mechanism for regulating secondary growth. For example, microarray transcript profiles across the radial developmental gradient in *Populus* stems showed strong correlations between transcript levels of genes of specific functions and developmental events (Schrader *et al.*, 2004). Functional studies have also demonstrated that specific aspects of secondary growth can be altered through misexpression of transcription factors. For example, the *Populus* class I KNOX transcription factor *ARBORKNOX1* (*ARK1*) is orthologous to Arabidopsis *SHOOTMERISTEMLESS* (*STM*), and is expressed in the cambial zone during secondary growth (Groover *et al.*, 2006). Over-expression of *ARK1* conditions pleiotropic phenotypes that include inhibition of differentiation of cambium daughter cells and alteration of transcript levels associated with secondary growth processes including cell differentiation and hormonal regulation (Groover *et al.*, 2006). Interestingly, class III homeodomain leucine zipper (class III HD ZIP) transcription factors have also been shown to play important roles in the initiation of the cambium and cambial daughter cell differentiation (Ko *et al.*, 2006; Du *et al.*, 2011; Robischon *et al.*, 2011; Zhu *et al.*, 2013), but the relationships among ARK1 and class III HD ZIPs have not been explored in *Populus*. As is true for other transcription factors associated with woody growth, basic features such as the number and identity of genes whose promoters are bound by ARK1 are unknown. As a result, more comprehensive and complex questions concerning transcriptional networks in trees are largely unexplored.

New genomic and sequencing technologies can potentially be extended to forest trees with sequenced genomes such as *Populus* spp., to reveal fundamental properties of transcriptional regulation during growth and development. This potential is illustrated by comprehensive modeling of transcriptional regulation in a handful of model animal species, aided in part by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) technologies that can identify transcription factor binding loci genome-wide (Schmidt *et al.*, 2009). Notably, the ENCODE project (Consortium, 2012) has used multiple data types including transcript levels, location of epigenetic marks,

and location of transcription factor binding to develop genome-wide descriptions of transcriptional regulation (Gerstein *et al.*, 2012). Major findings include the realization that many transcription factors bind to thousands of loci in the genome, and the action of transcription factors can be characterized in part by their preference in where they bind relative to genes (e.g. proximal or distal). The binding of an individual transcription factor to a gene is typically not sufficient to result in changes of expression (Cheng *et al.*, 2012; Karczewski *et al.*, 2014). Instead, changes in transcription reflect the combined effects of chromatin modification, sequence composition, chromatin accessibility, and the combinatorial binding of multiple transcription factors. Combinatorial binding of multiple transcription factors facilitates integration of developmental and environmental cues, and is often requisite for changes in transcription of a given gene (Cha & Zhou, 2014; Teng *et al.*, 2014).

Fundamental features of transcriptional regulation are also being addressed in a handful of plant species. Most of these studies have been undertaken in Arabidopsis, employing chromatin immunoprecipitation followed by hybridization to whole-genome tiling arrays (ChIP-chip) or sequencing (ChIP-seq) to identify transcription factor binding sites (Morohashi & Grotewold, 2009; Sun *et al.*, 2010; Ouyang *et al.*, 2011; Brandt *et al.*, 2012; Gregis *et al.*, 2013; Zhang *et al.*, 2013), DNaseI protection assays to identify protein-bound regions of the genome (Zhang *et al.*, 2012), microarrays and RNA-seq to quantify transcript levels (Giorgi *et al.*, 2013), and ribosome-bound transcript profiling to estimate levels of transcripts in active translation (Mustroph *et al.*, 2009). Hundreds or thousands of binding loci were identified for individual transcription factors assayed by ChIP-chip or ChIP-seq in Arabidopsis (Morohashi & Grotewold, 2009; Sun *et al.*, 2010; Ouyang *et al.*, 2011; Brandt *et al.*, 2012; Gregis *et al.*, 2013; Zhang *et al.*, 2013), indicating similar levels of complexity for transcriptional regulation in plants to those seen in animals. These tools are beginning to be extended to other plant species with fully sequenced genomes. For example, ChIP-seq of the maize (*Zea mays*) class I KNOX transcription factor KNOTTED1 (KN1) identified high-confidence KN1 binding sites associated with 4274 genes, with KN1 binding modestly correlated with transcript levels for those genes (Bolduc *et al.*, 2012).

One largely unaddressed question is how transcription factor binding loci have been gained and lost during plant evolution, including the fate of binding loci after genome duplication events commonly found in angiosperm lineages (Soltis *et al.*, 2009). For example, the *Populus* genome underwent a relatively recent whole-genome duplication event (salicoid duplication, estimated at 60–65 Myr ago), from which *c.* 8000 pairs of duplicated paralogs were retained (Tuskan *et al.*, 2006). Comparative analyses revealed that the salicoid paralogs were preferentially preserved in certain functional categories and have more conserved gene expression than expected by chance (Rodgers-Melnick *et al.*, 2012), but functional analysis of transcription factor binding in paralogous *Populus* genes is currently lacking. As data for transcription factor binding become available in additional species with different phylogenetic, developmental, and life history attributes, addressing basic questions regarding the evolution of transcriptional networks and associated biological traits will become tractable.

In this report, we describe experiments detailing attributes of ARK1 binding in the *Populus* genome, functional attributes of ARK1 binding, and the evolutionary history of ARK1 binding through comparison of binding both within and between species.

Materials and Methods

Plant cultivation and sample collection

Wild type and *ARK1* over-expression mutants (*35S:ARK1*) established in the hybrid aspen clone INRA 717-IB4 (*Populus tremula* × *Populus alba*) were grown in Magenta boxes (bioWORLD, Dublin, OH, USA) in controlled environmental chambers at 23°C with 18-h days as previously described (Groover *et al.*, 2006). Defoliated whole stems were collected for RNA-seq experiments of each genotype 6 wk after subculture. Samples for ChIP-seq were collected during active growth from mature *Populus trichocarpa* (Torr. & A. Gray) trees located in Westport, Oregon in late June 2011. Briefly, the bark was peeled from *c.* 1-m sections of stems, and vascular cambium and derivatives were collected by light scraping on both the xylem and phloem faces with double-edged razor blades and immediately processed. Tissues for ChIP-seq were fixed (0.4 M sucrose, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1% formaldehyde and 1 mM PMSF) under vacuum for 15 min. Glycine was added to a final concentration of 0.1 M for 5 min to terminate fixation. The tissues were then rinsed in distilled water before being frozen in the field in a dry-ice alcohol bath.

Immunology

Peptides that are both unique (i.e. not present in related KNOX transcription factors) and of predicted high antigenicity were identified in ARK1a (Potri.011G011100.1) and its paralog ARK1b (Potri.004G004700.1). Two peptides (ARK1_3738, EGNDRNASSEEELDV and ARK1_3940, DPQAEDQELK GQ) were selected as antigens for polyclonal antibody production (Supporting Information Fig. S1a), and were synthesized as Cys-conjugated peptides and used to immunize two rabbits per peptide by Pacific Immunology (Ramona, CA, USA). Antibodies were affinity-purified against the conjugated peptide, and evaluated for titer based on enzyme-linked immunosorbent assay (ELISA) of the peptide used for immunization.

For production of recombinant proteins, *ARK1a*, *ARK1b* and *ARK2* (Potri.002G113300.1) Protein Coding Sequences (CDS) were amplified using primers adding *Bam*HI and *Not*I sites (Supporting Information Table S1), and cloned into expression vector *pET23a*. Constructs were sequence confirmed and transformed into *Escherichia coli* BL21 cells (Invitrogen) for protein expression. Total protein was extracted with 1 × phosphate-buffered saline (PBS) buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4) and separated on 13% SDS PAGE. Proteins were transferred to nylon membranes, blocked with 5% nonfat milk in 1 × TBS buffer (50 mM Tris-Cl, pH 8.0, and 200 mM NaCl), and probed for 1 h with 0.4 μg ml^{-1}

anti-ARK1_3738 or with 0.4 μg ml^{-1} anti-ARK1_3940 in the blocking buffer (50 mM Tris-HCl, 200 mM NaCl, and 5% non-fat milk). Membranes were incubated with a peroxidase-conjugated anti-rabbit secondary antibody for 1 h, washed with 1 × TBST buffer (50 mM Tris-Cl, pH 8.0, 200 mM NaCl, and 0.05% Tween 20), developed with Pierce ECL Western Blotting Substrate (32209; Thermo Scientific, Waltham, MA, USA) and exposed to film.

ChIP-seq

Fixed *P. trichocarpa* vascular samples were ground to powder in liquid nitrogen. The nuclei were isolated with a CelLytic PN extraction kit (Sigma-Aldrich). Chromatin was fragmented to a size range of 200–500 bp by sonication (model VCXX130PB, 3-mm probe; Sonics & Materials Inc., Newtown, CT, USA) in lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 0.5% sodium deoxycholate, 0.3% SDS, proteinase inhibitor (Sigma; P9599-5ML) and 1 mM PMSF (Thermo Scientific; 36978)) on ice, with 40% amplitude and 10 s on/10 s off pulse. Overnight chromatin immunoprecipitations were performed using ChIP-IT® Express (Active Motif, Carlsbad, CA, USA; 53008) following the manufacturer's instructions. ChIP-seq libraries were prepared with an Illumina TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions, excluding gel purification and 15 PCR cycles for library amplification. Control 'input' libraries were prepared with *c*. 10 ng whole genomic DNA purified from sonicated chromatin. Libraries were sequenced using Solexa (Illumina) 50-bp single-end sequencing.

ChIP-seq data analysis

Raw reads were processed to trim adaptor sequences (using the scythe utility) and filter low-quality reads (using the sickle utility) (http://training.bioinformatics.ucdavis.edu/docs/2013/02/bootcamp/galaxy/qa-and-i.html). The processed reads were mapped to *P. trichocarpa* genome reference v3 (http://www.phytozome.net/poplar.php) using Bowtie2 (Langmead & Salzberg, 2012) with default parameters. Mapped reads were filtered to retain only those with a mapping quality score of 10 for peak calling (http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml). The ENCODE irreproducible discovery rate (IDR) pipeline (Li *et al.*, 2011) (https://sites.google.com/site/anshulkundaje/projects/idr) with peak caller macs2 (https://github.com/taoliu/MACS/) (Zhang *et al.*, 2008) was used for ChIP-seq data evaluation and peak calling, using input library as a control. Conservative thresholds were used to determine the final peaks: 0.02 threshold for self-consistency or comparison of true replicates and 0.0025 threshold for pooled-consistency analysis. ChIP-seq peaks were visualized using the Integrated Genome Browser (Thorvaldsdóttir *et al.*, 2013). Target genes and genome-wide binding patterns for ARK1 ChIP-seq peaks were calculated using the Bioconductor package ChIPpeakAnno (Zhu *et al.*, 2010). Each peak was assigned to the closest gene on either the positive or negative strand of a given chromosome, and these genes treated as putative ARK1 targets in follow-on analyses. The distance to the transcriptional start site (TSS) was calculated as the distance from the peak center to the nearest TSS on either the positive or negative strand. ARK1 binding in relation to target genes was classified into six possible categories: (1) upstream of TSS; (2) downstream of the 3′ end of the gene feature; (3) overlapStart, where the peak interval overlaps the TSS; (4) inside, where the peak interval is found completely inside the gene feature; (5) includeFeature, where the gene feature is found completely inside the peak interval; (6) overlapEnd, where the peak interval overlaps the 3′ end of the gene feature.

ARK1 salicoid paralog analyses and ARK1 comparison to KNOTTED1

Populus salicoid duplicated paralogs were identified from the *P. trichocarpa* V3 assembly using the algorithm previously described (Rodgers-Melnick *et al.*, 2012). The observed versus expected frequencies of ARK1 binding to salicoid paralogs were determined. Salicoid paralogs of ARK1 target genes were identified, and the percentage of those salicoid paralogs bound by ARK1 was determined. Random permutation tests were used to test the statistical significance of the percentage of ARK1 target genes with ARK1-bound salicoid paralogs. Syntenic mapping of the *Populus* v3 genome onto the maize v2 genome and extraction of putative orthologs were performed using the SynMap tool from CoGe (http://genomeevolution.org), with default options.

RNA-seq methods and data analysis

Three biological replicates were prepared for wild-type controls and *ARK1* over-expression mutants. Each biological replicate consisted of bulks of two defoliated stems. Total RNA was extracted with Trizol (Invitrogen), treated with DNase (Qiagen; 154903-4X2L) and then purified with the RNeasy Mini kit (Qiagen; 74104) following the manufacturer's protocol. mRNA sequencing libraries were prepared from total RNA using the Illumina TruSeq RNA Sample PrepKit and submitted for ultra-high-throughput Solexa (Illumina) 50-bp single-end sequencing.

Raw reads were processed using Scythe and Sickle with default settings. The processed reads were mapped to the *P. trichocarpa* v3 genome assembly using Tophat (Trapnell *et al.*, 2009) with default parameters. The raw mapped reads for each sample were counted using htseq-count (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html). Genes with zero read counts in all samples were removed from further analysis. The edgeR package was used for detection of differentially expressed genes (Robinson *et al.*, 2010), tmm was used for read count normalization (Robinson & Oshlack, 2010), and Fisher's exact test was used for evaluating statistical significance with a false discovery rate (FDR) of less than 0.05.

Yeast one-hybrid

To identify the upstream regulators of *popREVOLUTA* (*popREV*; Potri.009G014500), a yeast one-hybrid assay was performed according to the method of Deplancke *et al.* (2004), except that

vector 476 p5e-mcs (https://www.addgene.org/) was used instead of pDONR- P41r for initial cloning of the *popREV* promoter (*popREVp*). *popREVp* (1.5 kb upstream of the translational start site) was amplified from genomic DNA isolated from *P. trichocarpa*, using primers popREV-KpnI-5′ and popREV-SpeI-3′ (Table S1). The Deplancke *et al.* (2004) method allowed for sequential integration of the *popREVp*::*HIS3* and *popREVp*::*LacZ* reporter cassettes into the yeast genome, providing a more robust and natural chromatin setting for transcription factor binding and activation of reporter expression. Integration of the promoter::reporter constructs was verified by PCR, using primer popREV-KpnI-5′ with either HIS293RV or LacZ592RV. Transformed strains were selected by growth on minimal medium lacking histidine and uracil (the latter reflecting successful integration of the LacZ reporter), and positive colonies were then tested for auto-activation to identify the yeast bait strain exhibiting minimal auto-activation (background) to facilitate screening. The *popREVp* bait strain was screened against a pool of activation domain transcription factor (AD-TF) fusions representing 42 transcription factors up-regulated in differentiating xylem (Rodgers-Melnick *et al.*, 2012). Only those AD-TFs capable of activating both reporters were considered for further analysis. To identify the transcription factors that interacted with *popREVp*, colony PCR was performed with AD (Gal4AD) and terminator-specific (Gal4T) primers, followed by sequencing for those colonies yielding a single amplified PCR product. To retest the putative interactors, the *popREVp* bait strain was independently transformed with each relevant AD-TF vector and scored for reporter activation.

Gene ontology (GO) analysis

Over-representation of gene ontology (GO) terms was tested for gene sets derived from ARK1 target genes and differentially expressed genes using the BIOCONDUCTOR package GOSTATS (Falcon & Gentleman, 2007) with $P < 0.01$. GO annotation of poplar genes was obtained from the *P. trichocarpa* v3 genome annotation. Visualization of enriched GO categories was conducted using REVIGO (Supek *et al.*, 2011) with default settings.

Motif analysis

For *de novo* discovery of putative cis-motifs, sequences underlying ARK1 ChIP-seq peaks were extracted and analyzed using RSAT (http://rsat.ulb.ac.be/) using the peak-motifs tool (Thomas-Chollier *et al.*, 2012) with the cut peak sequence parameter set as 1000 bp on each side of peak center. Algorithms oligo-analysis, position-analysis, local-word-analysis, and dyad-analysis were chosen for motif prediction with oligomer lengths 6 nt, 7 nt, or 8 nt, and the top five best-scored motifs returned for each algorithm. The 1000-bp window for motif discovery encompassed > 95% of the peak intervals, and allowed evaluation of motif enrichment around the center of peak intervals. To determine whether a previously identified KNOX protein binding motif was enriched in ARK1 peak intervals, the previously reported 12-nt probability matrix of the maize class I KNOX KN1 binding motif (Bolduc *et al.*, 2012) was used in the Finding Individual Motif Occurrences (FIMO) tool from the MEME suite (http://meme.nhcr.net) (Grant *et al.*, 2011) to compare motif occurrence in ARK1 peak intervals versus the *Populus* genome. Motif occurrences were expressed in terms of the number per kilobase and considered significant if the *P*-value was below 1e-5.

Data archiving

All ARK1 ChIP-seq and RNA-seq sequences are archived in the NCBI Sequence Read Archive (SRA) under accession number SRP042635.

Results

Genome-wide identification of ARK1 binding loci

ChIP-seq was used to identify the genome-wide binding loci of ARBORKNOX1 (ARK1) (Groover *et al.*, 2006) in the vascular cambium and its recent derivatives of mature *P. trichocarpa* trees (see the 'Materials and Methods' section) following the ChIP-seq guidelines of the ENCODE consortia (Landt *et al.*, 2012). The *Populus* genome contains two ARK1 paralogs (ARK1a, Potri.011G011100; and ARK1b, Potri.004G004700). Two peptide-specific polyclonal antibodies were raised against endogenous ARK1 proteins (see the 'Materials and Methods' section) (Liu *et al.*, 2014). As shown in Fig. S1, antibody ARK1_3940 recognized both ARK1a and ARK1b proteins in denaturing western blots, while antibody ARK1_3738 recognized ARK1b but not ARK1a, presumably as a result of a single amino acid change from glutamic acid to aspartic acid in ARK1a versus ARK1b. Neither antibody recognizes the closely related *Populus* KNOX protein ARBORKNOX2 (Du *et al.*, 2009), showing high specificity of each antibody against ARK1 proteins.

Illumina libraries were sequenced for two independent ChIP-seq biological replicates for each antibody and for DNA 'input' libraries used as a control for ChIP-seq peak calling (see the 'Materials and Methods' section) (Liu *et al.*, 2014). Cross-correlation analysis (Landt *et al.*, 2012) with randomly subsampled sequencing reads from replicate ARK1_3738_r2 (Liu *et al.*, 2014) showed that at least 21 million mapped reads were required to reach the RSC (relative strand correlation) threshold suggested by ENCODE (Fig. S2a). Therefore, libraries were resequenced to obtain > 75 million sequencing reads for each replicate (Table S2). ChIP-seq replicates with antibody ARK1_3738 showed a higher RSC value than those with antibody ARK1_3940 in cross-correlation analysis (Fig. S2b).

Peak detection for the ChIP-seq replicates of two ARK1 antibodies showed different patterns of peak reproducibility when evaluated using IDR analysis (Supporting Information Notes S1) (Zhang *et al.*, 2008). Two replicates from ARK1_3738 ChIP-seq (ARK1_3738_r1 and ARK1_3738_r2) showed consistent detection of reproducible peaks (Fig. 1a, comparison 1 in the all peak plot) and low IDR scores until the number of peaks reached *c.* 15000 (Fig. 1a, comparison 1 in the IDR plot). Two replicates from ARK1_3940 (ARK1_3940_r1 and ARK1_3940_r2)

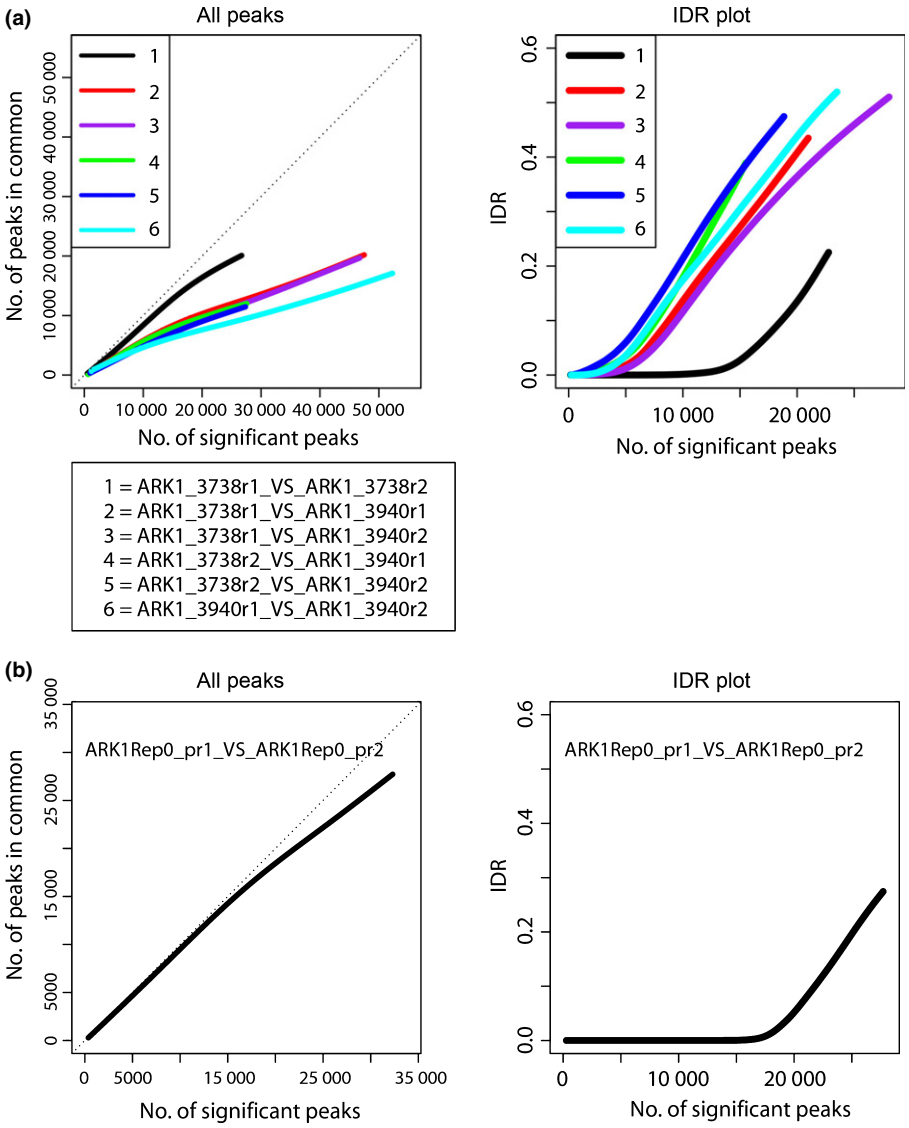

Fig. 1 Reproducibility and irreproducible discovery rate (IDR) analysis of *Populus* ARBORKNOX1 (ARK1) chromatin immunoprecipitation sequencing (ChIP-seq) from the vascular cambium and its recent derivatives. (a) Reproducibility plots of ARK1 ChIP-seq replicates. The inset box provides a key to ChIP-seq data sets by antibody and replicate name. The left plot compares the number of significant peaks commonly shared by each possible pairwise comparison of all replicates for increasing numbers of peaks included in the analysis (reproducibility profile). Theoretical perfect congruence between replicates is indicated by the dotted line. The right plot compares the IDR of increasing numbers of ChIP peaks for all possible pairwise comparisons of replicates. (b) Reproducibility plots of pooled ARK1 ChIP-seq replicates. ARK1Rep0_pr1 and ARK1Rep0_pr2 refer to pseudo-replicates of the pooled samples which were derived by randomly splitting the mapped reads into two samples. Left plot, the reproducibility profile between all the peaks identified in two replicates; right plot, the IDR at increasing numbers of peaks selected by IDR criterion.

showed less reproducible peaks and the IDR scores increased quickly after 5000 peaks when compared to each other (Fig. 1a, comparison 6 in all peak and IDR plots) and when compared to replicates from ARK1_3738 (Fig. 1a, comparisons 2–5 in all peak and IDR plots). The better performance of antibody ARK1_3738 presumably reflects differences in antibody recognition or performance in the ChIP protocol.

The final set of ChIP-seq peaks for follow-on analyses was calculated by pooling all four ARK1 replicates. Consistency analysis of the pooled data set showed detection of reproducible peaks with low IDR scores until *c.* 16 000 peaks (Fig. 1b, all peak and IDR plots). Based on a conservative threshold (see the 'Materials and Methods' section), 14 463 highly significant and reproducible ChIP-seq peaks were identified as putative ARK1 binding loci and used in the following studies (Table S3).

ARK1 binding loci are highly enriched in genic regions

ARK1 ChIP-seq peaks were highly concentrated around the TSS of genes (Fig. 2a). Using the distance of the peak center to the closest gene's TSS as a standard (see the 'Materials and Methods' section), 70% of the peaks localized within 0.5 kb of the TSS, and up to 95% localized within 3 kb of the TSS (Table 1). Up to 91% of the peaks overlapped with annotated gene features: 73% of peaks overlapped the closest gene's TSS, while only 8% resided upstream or downstream of the transcribed region of genes (Fig. 2b).

Each peak was assigned to the closest gene, which are referred to as ARK1 target genes hereafter. In total, 13 944 ARK1 target genes were identified, including 500 genes that had > 1 associated ChIP-seq peaks. There are no previously experimentally validated target genes for ARK1 in *Populus*, but KNOX protein binding to gibberellin-related Gibberellin 2-oxidase (*ga2ox*) has been experimentally validated in other species (Sakamoto *et al.*, 2001; Bolduc & Hake, 2009; Spinelli *et al.*, 2011; Bolduc *et al.*, 2012). As shown in Fig. 2(c), two *ga2ox* genes were identified as ARK1 target genes. Similar to KN1 (Bolduc & Hake, 2009; Bolduc *et al.*, 2012), the ARK1 binding loci are within the protein coding sequence of the *ga2ox* genes (Fig. 2c).

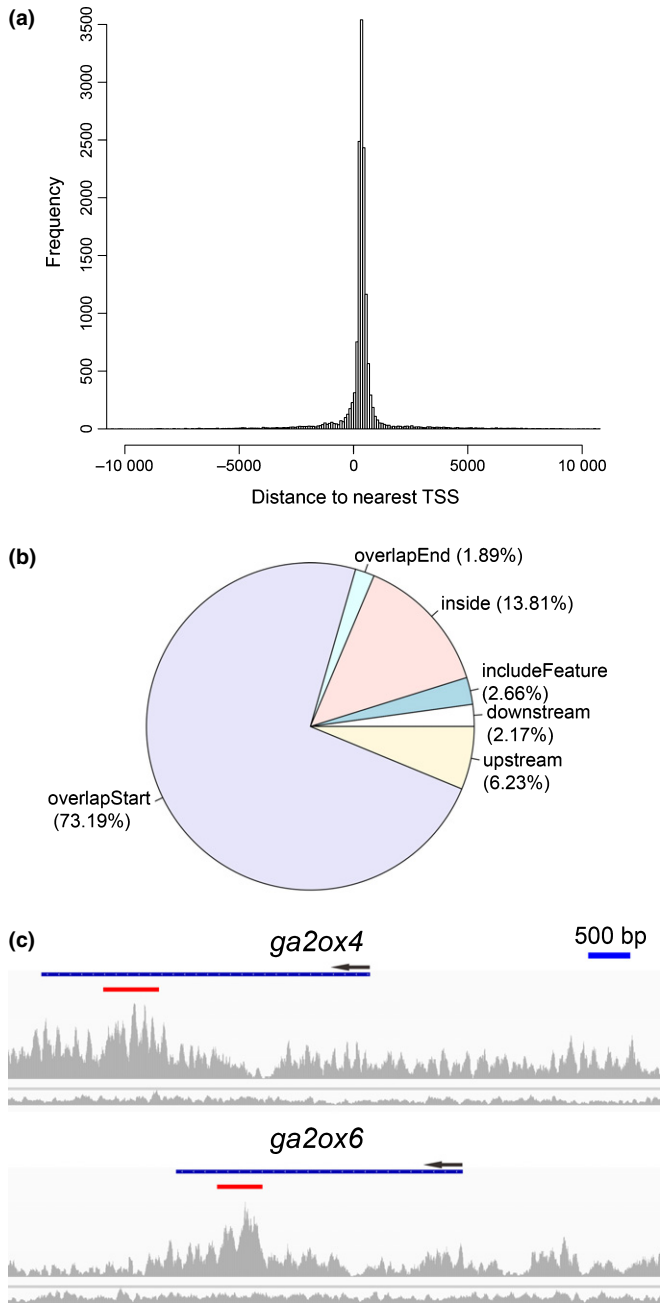

Fig. 2 Overview of ARBORKNOX1 (ARK1) binding loci in the *Populus* genome. (a) ARK1 binding loci are highly enriched in the promoter regions around the transcriptional start site (TSS) of genes. (b) ARK1 binding loci relative to gene features. (c) Significant ARK1 chromatin immunoprecipitation sequencing (ChIP-seq) peaks associated with two Gibberellin 2-oxidase (*ga2ox*) genes, homologs of which were characterized as direct targets of the class I KNOX protein KN1 in maize. For each gene, tracks from top to bottom represent the gene size and orientation, ARK1 ChIP-seq peak called by MACS2 (red bars), mapped reads of the pooled ARK1 ChIP-seq, and mapped reads of the input control used to subtract background differences in read mapping efficiencies from ChIP-peak calling, respectively. The black arrow indicates the gene TSS and orientation.

Putative ARK1 binding motifs

A total of 33 statistically overrepresented cis-element sequence motifs were identified in ARK1 ChIP-seq peaks (Notes S2) using the *de novo* regulatory sequence search tool peak-motifs (see the 'Materials and Methods' section) (van Helden, 2003; Thomas-Chollier *et al.*, 2008, 2011). The motif containing the core sequence 'AAACCCT/A' was most frequently identified. Only five out of the 33 significant motifs are found in the JASPAR core plants database and footprintDB-plants database (Mathelier *et al.*, 2014). Four high-confidence motifs were identified independently by different algorithms (Fig. 3a–d).

The previously characterized class I KNOX binding motif, which has a 'TGAC' core sequence (Sakamoto *et al.*, 2001; Bolduc & Hake, 2009; Spinelli *et al.*, 2011; Bolduc *et al.*, 2012), was not retrieved from the *de novo* search. To further determine whether this motif was present in ARK1 binding loci, we used FIMO (Grant *et al.*, 2011) to perform a directed search for the occurrence of this motif. As shown in Fig. 3(b), the frequency of this motif was approximately two times higher in ARK1 ChIP peaks compared with the whole genome, with a total of 560 occurrences in ARK1 binding loci.

Table 1 Distribution of ARBORKNOX1 (ARK1) chromatin immunoprecipitation sequencing (ChIP-seq) peaks relative to the closest transcriptional start site (TSS)

	< 100 bp[a]	0.1–0.5 kb[a]	0.5–1 kb[a]	1–2 kb[a]	2–3 kb[a]	> 3 kb[a]
Percentage of peaks[b]	3.7	66.8	17.9	4.6	2.4	4.5

[a]Distance from peak center to the TSS of the closest gene; [b]percentage of total peaks within each distance range.

ARK1 binding is enriched and retained in whole-genome duplication-derived paralogs

There were 7138 ARK1 target genes with at least one salicoid paralog in the genome (Rodgers-Melnick *et al.*, 2012), for which a total of 7141 corresponding salicoid paralogs were found, suggesting that ARK1 preferentially targets genes with salicoid paralogs in the genome ($P < 1e\text{-}5$). Whole-genome duplication-derived salicoid paralogs have more conserved gene expression than expected by chance (Rodgers-Melnick *et al.*, 2012), which indicates more conserved transcriptional regulation. Therefore, we hypothesized that if a gene was bound by ARK1, then its salicoid paralogs would have a higher chance to be bound by ARK1. Consistent with this hypothesis, 71% (5068) of the 7141 corresponding salicoid paralogs were bound by ARK1, a significant enrichment ($P < 1e\text{-}5$) as determined by a permutation test (Table 2). These results are consistent with overall conserved biological function and selection for retention of a significant fraction of ARK1 binding loci in the genome.

Comparison of ARK1 targets in *Populus* and KN1 targets in maize

Comparison of identified gene targets of ARK1 in *Populus* and the homologous targets of KNOTTED1 (KN1) in maize can offer an evolutionary perspective of class I KNOX binding across

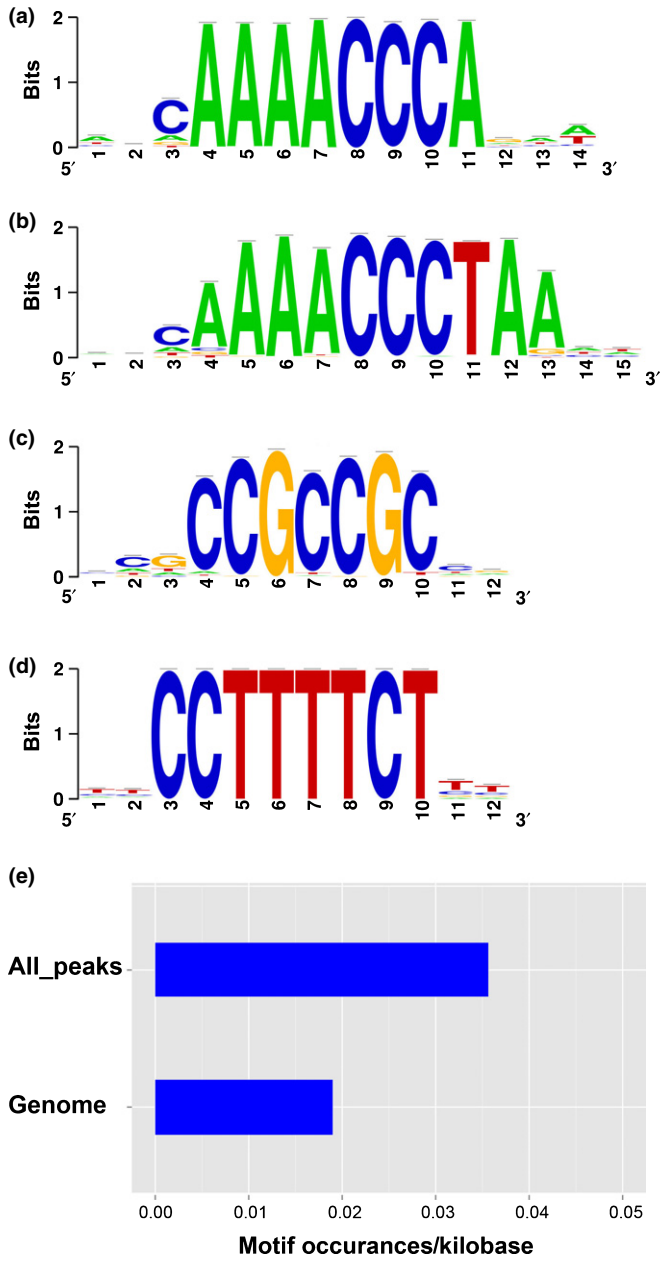

Fig. 3 Putative ARBORKNOX1 (ARK1) binding motifs. (a–d) Four high-confidence putative ARK1 binding motifs identified by different algorithms. (a) Positions_6nt_m1 (significance score = 300.00); (b) positions_6nt_m3 (significance score = 300.00); (c) positions_6nt_m5 (significance score = 300.00); (d) positions_6nt_m1 (significance score = 218.25). (e) Occurrence of a previously identified KNOX protein binding motif among ARK1 binding loci compared to the whole genome.

species. To that end, we performed syntenic mapping and retrieved the putative orthologous gene sets between *Populus* and maize (see the 'Materials and Methods' section).

At 52% or higher sequence similarity cutoff, 23 369 maize genes out of 39 657 (59%) mapped onto 23 676 *Populus* genes out of 41 335 (57%). Interestingly, the concentration of *Populus* genes with homologs in maize that were also ARK1 targets was much higher than expected. Specifically, out of the 13 944 ARK1 targets in *Populus*, 11 473 (82%) had maize homologs (i.e. were among those 23 676). Similarly, out of the 4274 KN1 targets in maize (Bolduc *et al.*, 2012), 3040 (71%) had homologs in *Populus*. A hypergeometric distribution test yielded a zero probability of obtaining these results by chance, showing that ARK1 and KN1 bind more selectively to genes that are evolutionarily conserved between the two species.

The gene homology relationships between *Populus* and maize are generally many-to-many, consistent with the known complex genome duplications and fractionations following their divergence. For example, at 52% sequence similarity cutoff, 3040 KN1-bound maize genes had 9769 homologs in *Populus*, of which 3688 were ARK1 targets; at 75% sequence similarity cutoff, 586 KN1-bound maize genes had 1056 *Populus* homologs, of which 475 were ARK1 targets; and at 80% sequence similarity cutoff, 107 KN1-bound maize genes had 183 *Populus* homologs, of which 83 were ARK1 targets (Table S4). As a consequence of the many-to-many relationships among orthologs, a more detailed statistical analysis of the enrichment of KN1 target homologs among ARK1 targets is not a well-defined task. However, ARK1 and KN1 have many common target genes which are homologous (Table S4) and include important regulators of plant development, such as transcription factor-encoding genes from the homeobox, NAC (NAM, ATAF1/2 and CUC2) domain, Myb (Myeloblastosis) domain, basic helix-loop-helix (bHLH), and MADS (MCM1, AG, DEF A and SRE)-box families; genes encoding ARGONAUTE proteins that are essential for small regulatory RNA processing; as well as genes encoding proteins involved in hormone perception and response (Table S4; see the next section).

ARK1 target genes participate in diverse biological functions

ARK1 target genes were significantly overrepresented within specific functional categories, as determined by GO analysis (see the 'Materials and Methods' section). In total, 605 GO enriched categories were identified, including 361 ontologies in

Table 2 Frequency of binding by ARBORKNOX1 (ARK1) to salicoid duplication-derived paralogs

	ARK1 target genes with paralog(s) [a]	Corresponding paralogs[b]	Corresponding paralogs bound by ARK1[c]	Corresponding paralogs bound by ARK1[d]	Expected by chance[e]	*P*-value[f]
Salicoid paralogs	7138	7141	5068	71%	33%	$P < 1e\text{-}5$

[a]Number of ARK1 target genes that have at least one salicoid paralog; [b]number of genes that are salicoid paralogs of ARK1 target genes from ([a]) in *Populus* genome; [c]number of genes from ([b]) that are also identified as ARK1 target genes; [d]the percentage of ([c]) in ([b]); [e]the percentage of all genes in the genome bound by ARK1; [f]*P*-value was derived from permutation test.

biological process (BP), 143 in molecular function (MF), and 100 in cellular component (CC) (Table S5). Overall, ARK1 was found to bind to genes participating in a wide variety of cellular activities and biological processes, such as the establishment of localization in cell, cellular macromolecular catabolism, carbohydrate derivative biosynthesis, and metabolism (Fig. S3). Notably, many of these enriched categories are related to DNA repair, RNA processing, and protein transport and modifications (Fig. 4).

Many genes encoding hormone receptors and their interacting proteins from different hormone pathways were found to be targeted by ARK1 (Table S6). For example, five cytokinin histidine kinase receptor genes in *Populus* were all identified as ARK1 targets; several F-box receptors of auxin (*transport inhibitor response 1* (*TIR1*), *TIR3*, *AUXIN SIGNALING F-BOX 2* (*AFB2*) and *AFB5*) (Greenham *et al.*, 2011; Calderon Villalobos *et al.*, 2012) and jasmonates (*coronatine-insensitive 1* (*COI1*)) (Ralhan *et al.*, 2012; Yan *et al.*, 2013), and an F-box component of gibberellin signaling (*Sleepy 1*) (Ariizumi *et al.*, 2011; Hauvermale *et al.*, 2014) are bound by ARK1. Multiple receptor genes of ethylene (*Ethylene receptor 1* (*ETR1*), *ETR2* and *ETHYLENE INSENSITIVE 4* (*EIN4*) (Hall *et al.*, 2012; Liu & Wen, 2012) and brassinosteroid (*brassinosteroid insensitive 1* (*BRI1*), *BRI1 Associated receptor Kinase 1* (*BAK1*) and *BR-signaling kinase 1* (*BSK1*)) (Friedrichsen *et al.*, 2000; Tang *et al.*, 2008; Wang *et al.*, 2012) signaling pathways were identified.

Class III *HD-ZIP* genes are targeted by ARK1

Class III *HD-ZIP* genes belong to a small plant-specific gene family, which contains five members in Arabidopsis (*REVOLUTA* (*REV*), *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), *CORONA/AtHB15* (*CNA*), and *Arabidopsis thaliana* HOMEOBOX GENE 8 (*AtHB8*)) and eight members in *Populus* (Ko *et al.*, 2006; Zhu *et al.*, 2013). Similar to ARK1, previous studies in *Populus* showed that class III *HD-ZIP* genes are involved in the initiation of vascular cambium, patterning of secondary vascular tissues, and differentiation of cambial daughter cells (Ko *et al.*, 2006; Du *et al.*, 2011; Robischon *et al.*, 2011; Zhu *et al.*, 2013). We thus asked whether there was any evidence of ARK1 regulation of these class III *HD-ZIP* genes affecting the cambium and secondary growth. As shown in Fig. 5(a), four (*PtrHB3*, *PtrHB1*, *PtrHB2/popREV*, and *PtrHB5/popPCN*) out of the eight *Populus* class III *HD-ZIP* genes were found to be bound by ARK1. ARK1 was also independently identified as binding to the promoter of *PtrHB2/popREV* in a yeast one-hybrid screen of upstream regulators of *PtrHB2/popREV* (Fig. 5b). Notably, beside ARK1, four other transcription factors were isolated as upstream regulators of *popREV* in the yeast one-hybrid screen (Fig. 5b), suggesting that ARK1 probably regulates transcription of *PtrHB2/popREV* through combinatorial binding in concert with other transcriptional regulators.

Effects of ARK1 binding on differentially expressed genes

To investigate the effects of ARK1 binding on transcript levels of target genes, genome-wide changes in transcript levels were compared using RNA sequencing (RNA-seq) for wild-type controls and an available *ARK1* over-expression mutant in hybrid aspen (Fig 6a) (see the 'Materials and Methods' section). Of the 3560 significantly differentially expressed genes (DEGs) identified, 2284 were up- and 1276 down-regulated in the *ARK1* over-expression mutant (Table S7). GO analysis found 121 categories enriched in these DEGs (Fig. S4; Table S8).

ARK1 binding was associated with expression modulation of only a subset of target genes in the genotypes and conditions tested. As shown in Fig. 6(b), there were 866 genes in common between DEGs and ARK1 target genes, representing 24% of the 3560 DEGs in *ARK1* over-expression mutants (Table S9), but only 6.2% of the 13 944 ARK1-bound genes ($P < 2.244e-35$), showing that ARK1 binding alone is not highly influential with target gene transcript levels. However, of the 866 genes that were both bound and differentially expressed, 683 were up-regulated while only 183 were down-regulated, suggesting that ARK1 is associated with transcription activation.

Gene ontology analysis of the DEGs bound by ARK1 indicated enrichment in categories of cellular protein modification processes, such as protein phosphorylation (76 genes), protein

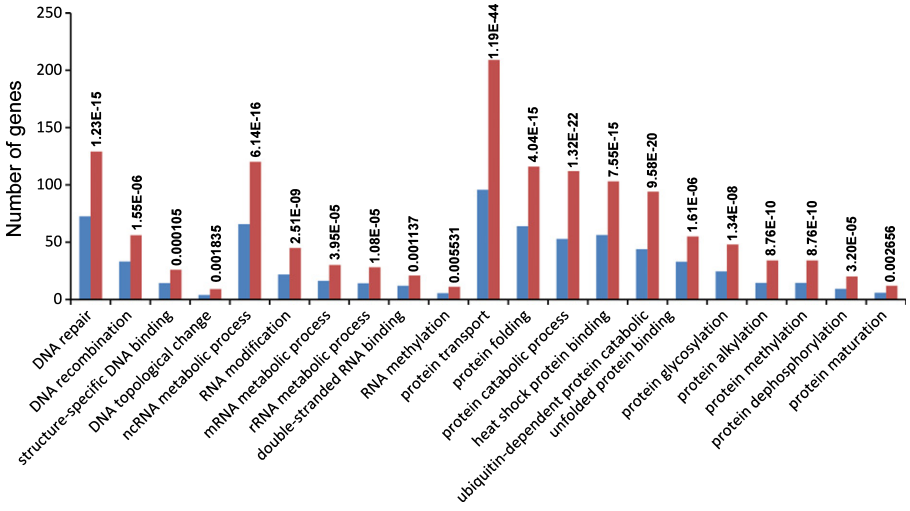

Fig. 4 Enrichment of ARBORKNOX1 (ARK1) target genes in gene ontology (GO) categories related to DNA, RNA, and protein catabolism. Numbers at the top of each histogram are *P*-values of the GO category enrichment. 'Expected' (blue bars) and 'Observed' (red bars) represent the numbers of expected and observed ARK1 chromatin immunoprecipitation sequencing (ChIP-seq) target genes in each GO category.

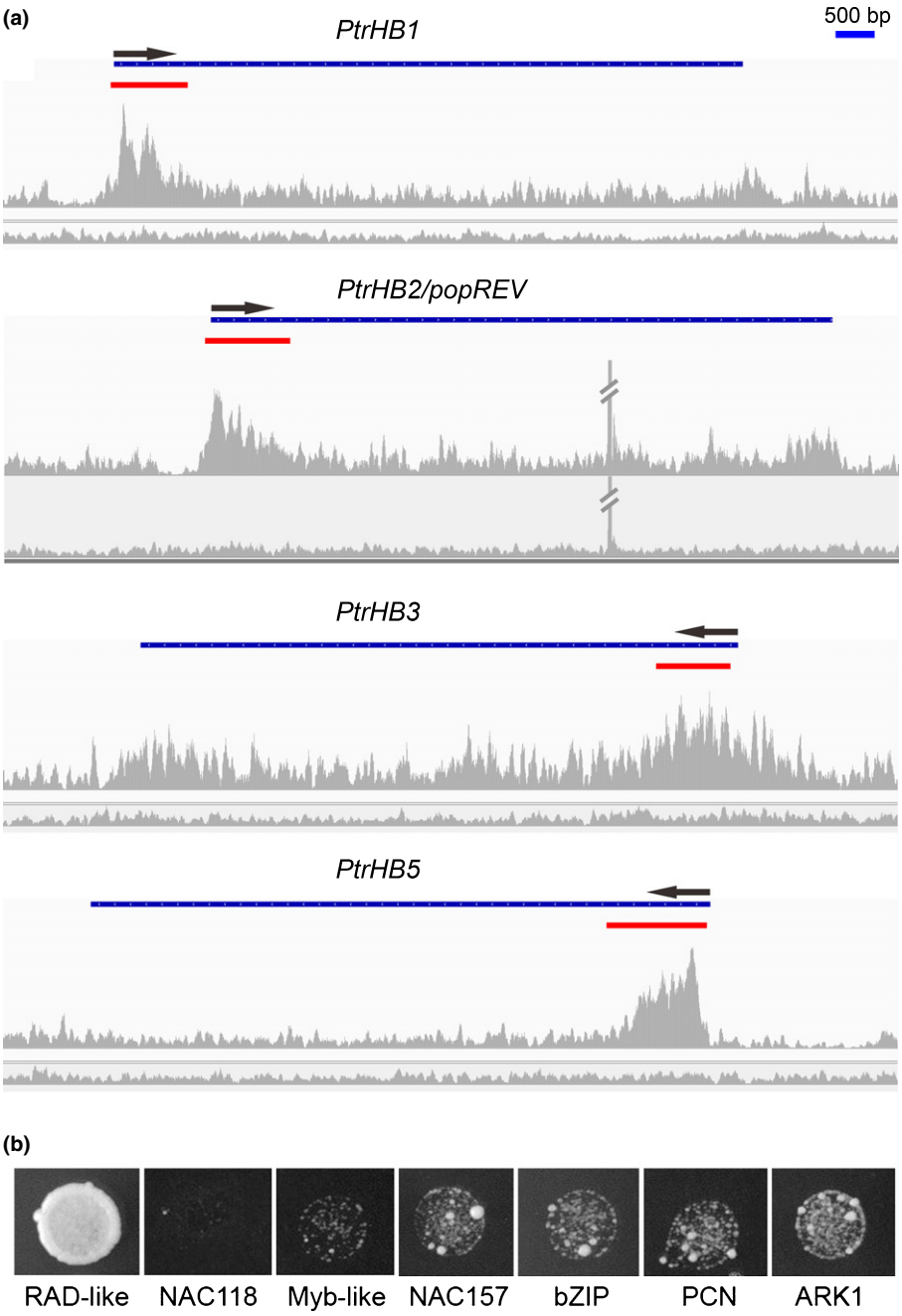

Fig. 5 ARBORKNOX1 (ARK1) binding to promoters of class III *HD ZIP* genes in *Populus*. (a) ARK1 chromatin immunoprecipitation sequencing (ChIP-seq) peaks associated with the promoters of class III *HD ZIP* genes. For each gene, tracks from top to bottom represent the gene size and orientation, ARK1 ChIP-seq peak called by MACS2 (red bars), mapped reads of the pooled ARK1 ChIP-seq, and mapped reads of the input control used to subtract background differences in read mapping efficiencies from ChIP-peak calling, respectively. The break lines in the mapped read tracks of *PtrHB2/popREV* (*Populus trichocarpa HOMEOBOX GENE 2/Populus REVOLUTA*) indicate that sequence read mapping in this region was off the scale, and is not enriched in ARK1 ChIP-seq reads compared with the input control. The black arrow indicates the gene transcriptional start site (TSS) and orientation. (b) Yeast one-hybrid assay identification of five *Populus* transcription factors, including ARK1, that putatively interact with the *popREV* promoter. RAD-like (*MYB protein RADIALIS-like*) (Potri.002G260000) and NAC118 (Potri.003G022800) are positive (strong interactor) and negative (noninteractor) controls, respectively. Four additional transcription factors were able to support weak to moderate growth: Myb-like, Potri.005G063200; NAC157, Potri.004G049300; bZIP (basic leucine zipper domain protein), Potri.008G018400; and POPCORONA (PCN), Potri.001G188800. Cells were diluted to OD A_{600} of 0.100 and grown for 3 d on minimal medium deficient in His, Ura, and Trp, and in the presence of 60 mM 3-Amino-1,2,4-triazole (3-AT).

ubiquitination (12 genes), and protein metabolic processes (Table S10). Other enriched categories included carbohydrate biosynthesis/metabolism and cell surface receptor signaling pathways (Fig. 6c). On a finer scale, some gene families showed significant overrepresentation among the 866 overlapping genes of the DEGs and ARK1 binding data sets. For example, 12 out of 17 up-regulated MAP kinase genes were bound by ARK1 ($P < 0.002$); all seven up-regulated cellulose synthase family genes were bound by ARK1 ($P < 4.630e\text{-}04$); as were two up-regulated poltergeist-like genes (*PLL1* and *PLL4*), which in Arabidopsis are dosage-sensitive regulators of meristem and organ development (Yu *et al.*, 2003; Song & Clark, 2005; Gagne *et al.*, 2008). These broad-acting and diverse genes that are bound and modulated by ARK1 are consistent with the strong and pleiotropic phenotype of ARK1 over-expression mutants.

Discussion

Transcriptional regulation is central to plant development, including the regulation of secondary growth and wood formation. To date, little is known about fundamental aspects of the interaction of transcription factors with their target genes in trees, including the identities of genome-wide binding targets, functional characteristics of binding, or how transcription factor binding correlates with transcript levels of target genes. Even less is known about the evolutionary gain and loss of transcription

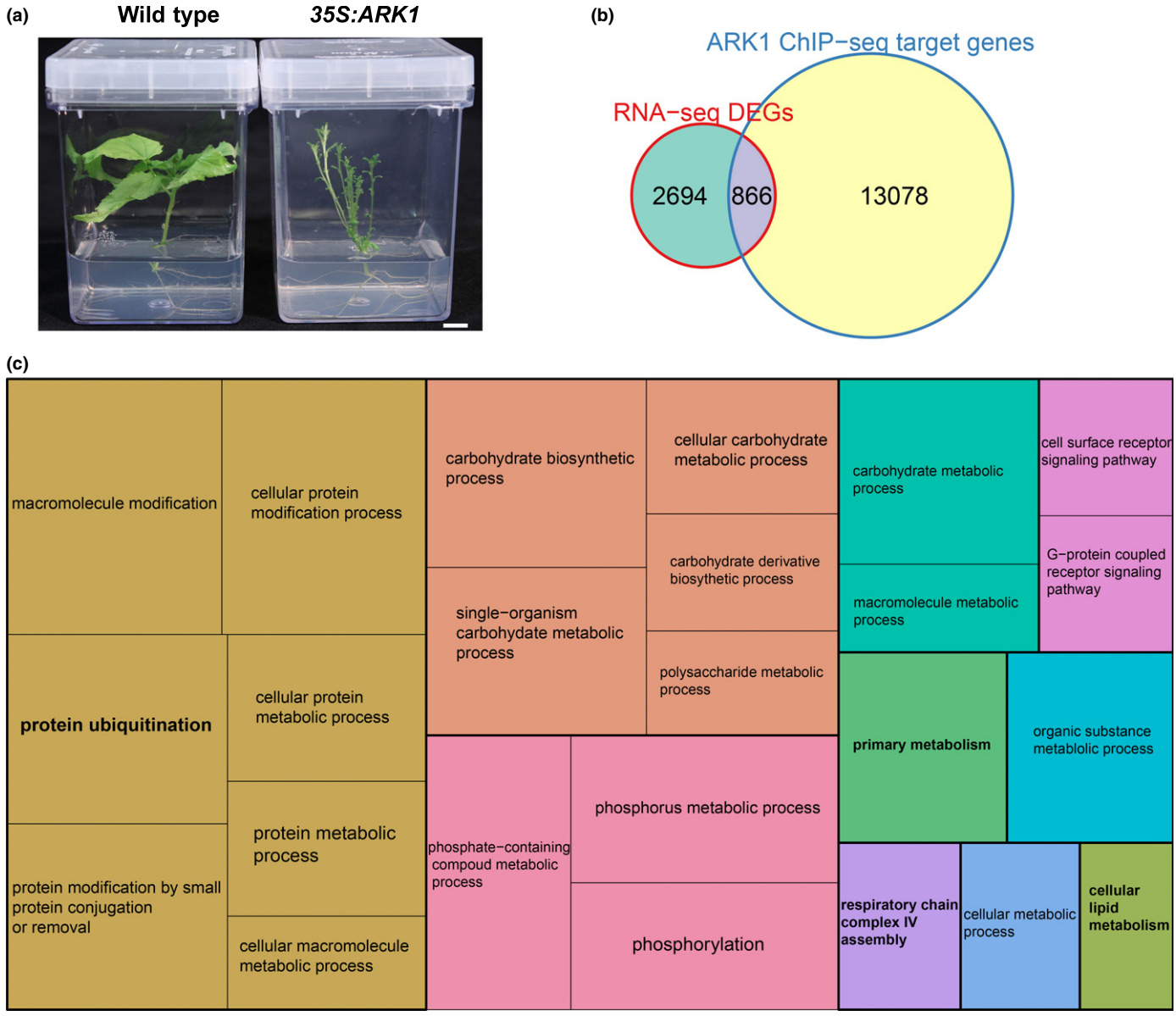

Fig. 6 Overlap between ARBORKNOX1 (ARK1) target genes and significantly differentially expressed genes in an *ARK1* over-expression mutant. (a) Whole-plant phenotype of wild-type control (left) and *ARK1* gain-of-function mutant (right) hybrid aspen (*Populus tremula* × *Populus alba*). Bar, 1.0 cm. (b) Venn diagram showing the overlap between ARK1 chromatin immunoprecipitation sequencing (ChIP-seq) target genes and RNA-seq significantly differentially expressed genes. (c) Visualization of enriched biological pathway gene ontology (GO) categories of the overlapping genes between ARK1 direct targets and significantly differentially expressed genes.

factor binding loci during plant evolution or after genome duplication events. In the study reported here, we explored these and other fundamental aspects of transcriptional regulation genome-wide for the class I KNOX transcription factor ARK1, which has been previously shown to regulate the vascular cambium and differentiation of cambium daughter cells (Groover *et al.*, 2006).

We performed ChIP-seq of cambium and its recent derivatives from mature, field-grown *P. trichocarpa* trees. This strategy involved raising polyclonal antibodies against peptides from ARK1 proteins, resulting in highly specific antibodies that can be used in any *Populus* genotype, including mature untransformed trees, thus bypassing the need for expression of epitope-tagged ARK1 for ChIP-seq. The success of this approach demonstrates that the ability to harvest large quantities of cambium and wood-forming cells in *Populus* can be exploited for mapping of transcription factor binding loci or chromatin marks using custom or commercially available antibodies. This is unique in meristem biology, given that harvesting of significant quantities of shoot or root meristems for ChIP or biochemical approaches is challenging.

Using rigorous standards established by the ENCODE project (Landt *et al.*, 2012), we identified 14 463 highly reproducible ARK1 binding sites in the *P. trichocarpa* genome. This large number of binding sites is roughly in keeping with what has been

reported for other plant as well as animal transcription factors (Morohashi & Grotewold, 2009; Sun *et al.*, 2010; Ouyang *et al.*, 2011; Brandt *et al.*, 2012; Gregis *et al.*, 2013; Zhang *et al.*, 2013). The challenge shared by this and other ChIP-seq studies is to determine the biological relevance of the large number of binding sites, which would seem excessive if each binding event has direct relevance to the regulation of gene expression. We used six criteria to evaluate the biological relevance of ChIP peaks, which collectively support the notion that the ARK1 binding revealed by ChIP-seq peaks is reflective of ARK1 function. First, ARK1 binding is not randomly distributed throughout the genome, but rather is highly concentrated proximal to the transcriptional start sites of genes. Secondly, salicoid duplication-derived paralogs have a higher frequency of ARK1 binding than unduplicated genes, and ARK1 binding to paralogous genes is highly correlated. Thirdly, ARK1 in *Populus* and the homologous KNOTTED in maize preferentially bind to evolutionarily conserved genes. Fourthly, ARK1 target genes are enriched in specific GO categories. Fifthly, enriched cis motifs are associated with ARK1 binding loci although, similar to findings for KN1 in maize (Bolduc *et al.*, 2012), there is no single consensus motif among all loci, indicating complexities for *in planta* binding specificity not revealed by previous *in vitro* binding studies for KNOX proteins. The sixth criterion evaluated ARK1 binding in relation to gene transcript levels, finding only a modest portion of ARK1 target genes differentially expressed in an *ARK1* over-expression mutant. This low overlap is not unlike findings for other transcription factors (Cheng *et al.*, 2012), but was probably further reduced by practical limitations of our experimental design. We used mature *P. trichocarpa* trees for ChIP-seq, to facilitate both harvest of sufficient cambium and recent derivatives for efficient ChIP of these specific tissues and efficient mapping of ChIP-seq reads to the *P. trichocarpa* reference genome. Unfortunately, *P. trichocarpa* performs poorly in transformation, and thus we used a readily transformed *P. tremula* × *alba* hybrid for transformation and RNAseq assay of transcript levels. These limitations will probably be addressed in the near future, as reference genomes become available for aspens and other *Populus* species. However, the modest correlation of ARK1 binding and transcript levels in the present study probably reflects other basic features of ARK1 function.

Our findings allow us to draw some important conclusions regarding the function of ARK1. Notably, binding by ARK1 alone is not sufficient to significantly modulate the expression of most target genes. While some ARK1 binding loci possibly represent 'scanning' of promoters (Biggin, 2011), it is likely that factors in addition to ARK1 are required to activate or repress transcription. Indeed, combinatorial regulation of gene expression through binding of multiple transcription factors to a promoter is common (Slattery *et al.*, 2011; Bochkis *et al.*, 2012; Zhang *et al.*, 2013), and facilitates integration of environmental and developmental cues into transcriptional outputs. The well-established heterodimerization of KNOX and BELL-like transcription factors (Bellaoui *et al.*, 2001; Muller *et al.*, 2001) demonstrates that ARK1 co-associates with other transcription factors on the promoters of target genes. Moreover, the binding of ARK1 to the promoter of the poplar *REVOLUTA* ortholog (*popREV*) was detected here by both ARK1 ChIP-seq and yeast one-hybrid assay, yet *popREV* is not significantly misexpressed in *ARK1* over-expression mutants. That yeast one-hybrid assays identified at least four other transcription factors that bind to the popREV promoter may be a reflection of the requirement for these or other transcription factors in combination with ARK1 as regulators of *popREV*.

The present study was successful in cataloging and describing genome-wide binding for an important transcription factor regulating plant meristems and growth, ARK1. However, the results also make clear that knowledge of binding of a single transcription factor is not sufficient to build robust, predictive models of transcriptional regulation of secondary growth. To meet this larger goal, it is likely that consideration of additional factors found by the ENCODE project will need to be addressed (Gerstein *et al.*, 2012). Critically, binding data for multiple transcription factors involved in secondary growth will be required to model combinatorial binding. Transcription factor binding data should be further augmented by mapping of active and repressive chromatin marks, as well as more direct measures of transcription than simple quantification of transcript levels. The relative cost and power of nucleic acid sequencing are making such comprehensive experiments increasingly tractable, and data thus obtained would be highly complementary to the growing genome and gene expression databases for forest trees.

Acknowledgements

We thank Courtney Castle and Gayle Dupper for plant transformation and propagation, and Victor Missirian and Trevor Ramsey for contribution of custom Python scripts and pipelines. We thank Kathy Haiby, Rich Shuren, and Brian Stanton of Greenwood Resources for facilitating harvesting of cambium and developing wood/bark samples. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. This work was supported by USDA NIFA Grant 2011-67013-30062 to A.G. and V.F., and by the Office of Science (BER), US Department of Energy, Grant no. DE-FG02-07ER64449 to E.P.B.

References

Ariizumi T, Lawrence PK, Steber CM. 2011. The role of two F-box proteins, SLEEPY1 and SNEEZY, in Arabidopsis gibberellin signaling. *Plant Physiology* **155**: 765–775.

Bellaoui M, Pidkowich MS, Samach A, Kushalappa K, Kohalmi SE, Modrusan Z, Crosby WL, Haughn GW. 2001. The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals. *Plant Cell* **13**: 2455–2470.

Biggin MD. 2011. Animal transcription networks as highly connected, quantitative continua. *Developmental Cell* **21**: 611–626.

Bochkis IM, Schug J, Ye DZ, Kurinna S, Stratton SA, Barton MC, Kaestner KH. 2012. Genome-wide location analysis reveals distinct transcriptional circuitry by paralogous regulators Foxa1 and Foxa2. *PLoS Genetics* **8**: e1002770.

Bolduc N, Hake S. 2009. The maize transcription factor KNOTTED1 directly regulates the gibberellin catabolism gene ga2ox1. *Plant Cell* **21**: 1647–1658.

Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'Connor D, Grotewold E, Hake S. 2012. Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes & Development* **26**: 1685–1690.

Brandt R, Salla-Martret M, Bou-Torrent J, Musielak T, Stahl M, Lanz C, Ott F, Schmid M, Greb T, Schwarz M *et al.* 2012. Genome-wide binding-site analysis of REVOLUTA reveals a link between leaf patterning and light-mediated growth responses. *Plant Journal* **72**: 31–42.

Calderon Villalobos LI, Lee S, De Oliveira C, Ivetac A, Brandt W, Armitage L, Sheard LB, Tan X, Parry G, Mao H *et al.* 2012. A combinatorial TIR1/AFB-Aux/IAA co-receptor system for differential sensing of auxin. *Nature Chemical Biology* **8**: 477–485.

Cha M, Zhou Q. 2014. Detecting clustering and ordering binding patterns among transcription factors via point process models. *Bioinformatics* **30**: 2263–2271.

Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, Yan KK, Dong X, Djebali S, Ruan Y *et al.* 2012. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Research* **22**: 1658–1667.

Consortium TEP. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Deplancke B, Dupuy D, Vidal M, Walhout AJ. 2004. A gateway-compatible yeast one-hybrid system. *Genome Research* **14**: 2093–2101.

Du J, Mansfield SD, Groover AT. 2009. The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *Plant Journal* **60**: 1000–1014.

Du J, Miura E, Robischon M, Martinez C, Groover A. 2011. The *Populus* Class III HD ZIP transcription factor *POPCORONA* affects cell differentiation during secondary growth of woody stems. *PLoS ONE* **6**: e17458.

Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**: 257–258.

Friedrichsen DM, Joazeiro CA, Li J, Hunter T, Chory J. 2000. Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiology* **123**: 1247–1256.

Gagne JM, Song SK, Clark SE. 2008. POLTERGEIST and PLL1 are required for stem cell function with potential roles in cell asymmetry and auxin signaling. *Communicative & Integrative Biology* **1**: 53–55.

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R *et al.* 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**: 91–100.

Giorgi FM, Del Fabbro C, Licausi F. 2013. Comparative study of RNA-seq- and microarray-derived coexpression networks in *Arabidopsis thaliana*. *Bioinformatics* **29**: 717–724.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Greenham K, Santner A, Castillejo C, Mooney S, Sairanen I, Ljung K, Estelle M. 2011. The AFB4 auxin receptor is a negative regulator of auxin signaling in seedlings. *Current Biology* **21**: 520–525.

Gregis V, Andres F, Sessa A, Guerra RF, Simonini S, Mateos JL, Torti S, Zambelli F, Prazzoli GM, Bjerkan KN *et al.* 2013. Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in *Arabidopsis*. *Genome Biology* **14**: R56.

Groover AT, Mansfield SD, DiFazio SP, Dupper G, Fontana JR, Millar R, Wang Y. 2006. The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Molecular Biology* **61**: 917–932.

Hall BP, Shakeel SN, Amir M, Ul Haq N, Qu X, Schaller GE. 2012. Histidine kinase activity of the ethylene receptor ETR1 facilitates the ethylene response in Arabidopsis. *Plant Physiology* **159**: 682–695.

Hauvermale AL, Ariizumi T, Steber CM. 2014. The roles of the GA receptors GID1a, GID1b, and GID1c in sly1-independent GA signaling. *Plant Signaling & Behavior* **9**: e28030.

van Helden J. 2003. Regulatory sequence analysis tools. *Nucleic Acids Research* **31**: 3593–3596.

Karczewski KJ, Snyder M, Altman RB, Tatonetti NP. 2014. Coherent functional modules improve transcription factor target identification, cooperativity prediction, and disease association. *PLoS Genetics* **10**: e1004122.

Ko JH, Prassinos C, Han KH. 2006. Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (*miR166*). *New Phytologist* **169**: 469–478.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P *et al.* 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* **22**: 1813–1831.

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

Larson PR. 1994. *The vascular cambium*. Berlin, Heidelberg, Germany: Springer-Verlag.

Li Q, Brown J, Huang H, Bickel P. 2011. Measuring eproducibility of highthroughput experiments. *Annals of Applied Statistics* **5**: 1752–1779.

Liu L, Missirian V, Zinkgraf M, Groover A, Filkov V. 2014. Evaluation of experimental design and computational parameter choices affecting analyses of ChIP-seq and RNA-seq data in undomesticated poplar trees. *BMC Genomics* **15**(Suppl 5): S3.

Liu Q, Wen CK. 2012. Arabidopsis ETR1 and ERS1 differentially repress the ethylene response in combination with other ethylene receptor genes. *Plant Physiology* **158**: 1193–1207.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H *et al.* 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research* **42**(Database issue): D142–D147.

Morohashi K, Grotewold E. 2009. A systems approach reveals regulatory circuitry for Arabidopsis trichome initiation by the GL3 and GL1 selectors. *PLoS Genetics* **5**: e1000396.

Muller J, Wang Y, Franzen R, Santi L, Salamini F, Rohde W. 2001. In vitro interactions between barley TALE homeodomain proteins suggest a role for protein-protein associations in the regulation of Knox gene function. *Plant Journal* **27**: 13–23.

Mustroph A, Zanetti ME, Jang CJH, Holtan HE, Repetti PP, Galbraith DW, Girke T, Bailey-Serres J. 2009. Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* **106**: 18843–18848.

Ouyang X, Li J, Li G, Li B, Chen B, Shen H, Huang X, Mo X, Wan X, Lin R *et al.* 2011. Genome-wide binding site analysis of FAR-RED ELONGATED HYPOCOTYL3 reveals its novel function in *Arabidopsis* development. *Plant Cell* **23**: 2514–2535.

Ralhan A, Schottle S, Thurow C, Iven T, Feussner I, Polle A, Gatz C. 2012. The vascular pathogen *Verticillium longisporum* requires a jasmonic acid-independent COI1 function in roots to elicit disease symptoms in Arabidopsis shoots. *Plant Physiology* **159**: 1192–1203.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Robischon M, Du J, Miura E, Groover A. 2011. The *Populus* class III HD ZIP, *popREVOLUTA*, influences cambium initiation and patterning of woody stems. *Plant Physiology* **155**: 1214–1225.

Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, Difazio SP. 2012. Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Research* **22**: 95–105.

Sakamoto T, Kamiya N, Ueguchi-Tanaka M, Iwahori S, Matsuoka M. 2001. KNOX homeodomain protein directly suppresses the expression of a gibberellin biosynthetic gene in the tobacco shoot apical meristem. *Genes & Development* **15**: 581–590.

Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**: 240–248.

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, Sandberg G. 2004. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* **16**: 2278–2292.

Slattery M, Ma L, Negre N, White KP, Mann RS. 2011. Genome-wide tissue-specific occupancy of the Hox protein Ultrabithorax and Hox cofactor Homothorax in *Drosophila*. *PLoS ONE* **6**: e14686.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.

Song SK, Clark SE. 2005. POL and related phosphatases are dosage-sensitive regulators of meristem and organ development in Arabidopsis. *Developmental Biology* **285**: 272–284.

Spinelli SV, Martin AP, Viola IL, Gonzalez DH, Palatnik JF. 2011. A mechanistic link between STM and CUC1 during Arabidopsis development. *Plant Physiology* **156**: 1894–1904.

Sun Y, Fan XY, Cao DM, Tang W, He K, Zhu JY, He JX, Bai MY, Zhu S, Oh E *et al.* 2010. Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis. *Developmental Cell* **19**: 765–777.

Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**: e21800.

Tang W, Kim TW, Oses-Prieto JA, Sun Y, Deng Z, Zhu S, Wang R, Burlingame AL, Wang ZY. 2008. BSKs mediate signal transduction from the receptor kinase BRI1 in Arabidopsis. *Science* **321**: 557–560.

Teng L, He B, Gao P, Gao L, Tan K. 2014. Discover context-specific combinatorial transcription factor interactions by integrating diverse ChIP-Seq data sets. *Nucleic Acids Research* **42**: e24.

Thomas-Chollier M, Darbo E, Herrmann C, Defrance M, Thieffry D, van Helden J. 2012. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nature Protocols* **7**: 1551–1568.

Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J. 2011. RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Research* **39**(Web Server issue): W86–W91.

Thomas-Chollier M, Sand O, Turatsinze JV, Janky R, Defrance M, Vervisch E, Brohee S, van Helden J. 2008. RSAT: regulatory sequence analysis tools. *Nucleic Acids Research* **36**(Web Server issue): W119–W127.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**: 178–192.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.

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

Wang ZY, Bai MY, Oh E, Zhu JY. 2012. Brassinosteroid signaling network and regulation of photomorphogenesis. *Annual Review of Genetics* **46**: 701–724.

Yan J, Li H, Li S, Yao R, Deng H, Xie Q, Xie D. 2013. The Arabidopsis F-box protein CORONATINE INSENSITIVE1 is stabilized by SCFCOI1 and degraded via the 26S proteasome pathway. *Plant Cell* **25**: 486–498.

Yu LP, Miller AK, Clark SE. 2003. *POLTERGEIST* encodes a protein phosphatase 2C that regulates CLAVATA pathways controlling stem cell identity at *Arabidopsis* shoot and flower meristems. *Current Biology* **13**: 179–188.

Zhang W, Zhang T, Wu Y, Jiang J. 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in Arabidopsis. *Plant Cell* **24**: 2719–2731.

Zhang Y, Liu T, Meyer C, Eeckhoute J, Johnson D, Bernstein B, Nusbaum C, Myers R, Brown M, Li W *et al.* 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**: R137.

Zhang Y, Mayba O, Pfeiffer A, Shi H, Tepperman JM, Speed TP, Quail PH. 2013. A quartet of PIF bHLH factors provides a transcriptionally centered signaling hub that regulates seedling morphogenesis through differential expression-patterning of shared target genes in Arabidopsis. *PLoS Genetics* **9**: e1003244.

Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, Green MR. 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**: 237.

Zhu Y, Song D, Sun J, Wang X, Li L. 2013. *PtrHB7*, a class III HD-Zip gene, plays a critical role in regulation of vascular cambium differentiation in *Populus*. *Molecular Plant* **6**: 1331–1343.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 ARK1-specific antibodies for ChIP-seq.

Fig. S2 Cross-correlation analysis of ARK1 ChIP-seq replicates.

Fig. S3 Visualization of enriched biological pathway (BP) GO categories of ARK1 target genes.

Fig. S4 Visualization of enriched biological pathway (BP) GO categories of significantly differentially expressed genes in *ARK1* over-expression mutants.

Table S1 Primers used for cloning and PCR verification

Table S2 Summary of ARK1 ChIP-seq libraries

Table S3 ARK1 ChIP-seq binding loci and target genes

Table S4 Comparison of ARK1 and KN1 target genes

Table S5 GO enrichment of ARK1 target genes

Table S6 Hormone perception-related genes in ARK1 targets

Table S7 Significantly differentially expressed genes in *ARK1* over-expression mutant

Table S8 GO enrichment of significantly differentially expressed genes in *ARK1* over-expression mutant

Table S9 Overlap of ARK1 target genes and significantly differentially expressed genes in *ARK1* over-expression mutant

Table S10 GO enrichment of overlapping genes between ARK1 targets and significantly differentially expressed genes in *ARK1* over-expression mutant

Notes S1 Commands used for ARK1 ChIP-seq peak calling and IDR analysis.

Notes S2 Putative ARK1 binding motifs identified with peak-motifs.

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