PIF4 enhances DNA binding of CDF2 to co-regulate target gene expression and promote Arabidopsis hypocotyl cell elongation

He Gao1, Wen Song2,3, Edouard Severing1,3, Alice Vayssières1, Bruno Huettel1, Rainer Franzen1, René Richter1, Jijie Chai1,2 and George Coupland1,2,✉

How specificity is conferred within gene regulatory networks is an important problem in biology. The basic helix-loop-helix PHYTOCHROME-INTERACTING FACTORS (PIFs) and single zinc-finger CYCLING DOF FACTORS (CDFs) mediate growth responses of Arabidopsis to light and temperature. We show that these two classes of transcription factor (TF) act cooperatively. CDF2 and PIF4 are temporally and spatially co-expressed, they interact to form a protein complex and act in the same genetic pathway to promote hypocotyl cell elongation. Furthermore, PIF4 substantially strengthens genome-wide occupancy of CDF2 at a subset of its target genes. One of these, YUCCA8, encodes an auxin biosynthesis enzyme whose transcription is increased by PIF4 and CDF2 to contribute to hypocotyl elongation. The binding sites of PIF4 and CDF2 in YUCCA8 are closely spaced, and in vitro PIF4 enhances binding of CDF2. We propose that this occurs by direct protein interaction and because PIF4 binding alters DNA conformation. Thus, we define mechanisms by which PIF and CDF TFs cooperate to achieve regulatory specificity and promote cell elongation in response to light.

Cellular responses to environmental and developmental signals require activation of gene regulatory networks by recruitment of transcription factors (TFs) to specific genes. Plant genomes encode relatively large numbers of TFs, emphasizing the importance of transcriptional regulation, and recruitment of combinations of TFs to the same gene can integrate different signals and enhance specificity4. Nevertheless, TFs usually recognize simple DNA sequences in vitro, and it remains unclear how they are recruited to specific genes and implement unique functions in vivo. In plants, DOF (DNA-binding with one finger) TFs, which contain a conserved CXCX2CX2C motif, regulate a wide range of developmental and environmental responses by binding to specific target genes in vivo, but their DNA-binding site has only been described by a simple consensus motif, AAAG or [T/A]AAAG motif, which occurs very widely in plant genomes. Within this family, CYCLING DOF FACTORS (CDF) are temporally regulated by the circadian clock to repress photoperiodic flowering and tuberisation9–11, and they also promote environmental responses by binding to specific target genes in vivo6, but it remains unclear how these interactions influence binding-site selection. Here we use a combination of in vivo and in vitro approaches to show that in Arabidopsis, CDF2 physically interacts with PHYTOCHROME-INTERACTING 4 (PIF4), an intensively studied basic helix-loop-helix (bHLH) TF with well-established functions in promoting growth in response to light and temperature16–20. PIF4 interacts directly with phytochrome and cryptochrome photoreceptors18,19,21,22, which regulate its activity in response to red/far-red and blue light, respectively. We find that PIF4 and CDF2 promote hypocotyl cell elongation, that the proteins directly interact and that PIF4 binding increases the strength and alters the specificity of CDF2 binding to a subset of target genes in vivo and in vitro. Therefore, combinatorial functions of PIF4 and CDF2 increase transcription of their mutual target genes, and provide a mechanism by which PIF4 enables CDF2 to activate specific target genes to promote hypocotyl cell growth.

Results
PIF4 and CDFs promote elongation of hypocotyl cells. Under short-day (SD) photoperiods, the hypocotyl of Arabidopsis seedlings grows rhythmically with a peak in growth rate at dawn23,24. CDF and PIF TFs promote hypocotyl elongation under SDs23,25,26. Inactivation of the partially redundant CDF1, CDF2, CDF3 and CDF5 genes in the cdf1235 quadruple (cdfq) mutant or of the PIF4 and PIF5 genes in the pif4 pif5 double mutant reduces hypocotyl growth under SDs4,14,23,25,26. To test whether CDFs and PIF4 promote growth in the same genetic pathway, the cdfq pif4 quintuple mutant was generated. In SDs, no differences were observed in hypocotyl length among the pif4, cdfq and cdfq pif4 genotypes, although they all produced shorter hypocotyls than those of wild-type (Col-0) plants (Extended Data Fig. 1a). After germination in the dark, the length of the hypocotyl of pif4, cdfq and cdfq pif4 mutants was indistinguishable from that of Col-0 (Extended Data Fig. 1b–d). Therefore, CDFs and PIF4 promote hypocotyl elongation under SDs in a non-additive, light-dependent manner, suggesting that they promote growth in the same genetic pathway.

To understand the histological differences underlying variation in hypocotyl length among these genotypes, the size and number of cells in the epidermis were measured. Non-dividing cell files were examined to assess the effect of the mutations on cell growth27 (Fig. 1a). Analysis of confocal microscopy images showed that the

1Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany. 2Institute of Biochemistry, University of Cologne, Cologne, Germany. 3These authors contributed equally: Wen Song, Edouard Severing. ✉E-mail: coupland@mpipz.mpg.de
numbers of cells in the non-dividing files were highly similar among all genotypes (Fig. 1b), although the hypocotyls of pif4, cdfq and pif4 cdfq mutants were shorter than those of Col-0 (Extended Data Fig. 1a). In each genotype, the length of cells in the non-dividing files increased basipetally from the shoot apical meristem to the cotlet, particularly between cells 8 and 14 (Fig. 1a and Extended Data Fig. 1e), as described for dark-grown Col-0 seedlings27. In pif4, cdfq and pif4 cdfq mutants, the mean cell length in these files was shorter than in Col-0, particularly between cells 5 and 10 (Extended Data Fig. 1e), and there was no significant difference in cell length among the mutants (Fig. 1c). The cell width of the non-dividing files decreased basipetally (Extended Data Fig. 1f). The mean cell width in the non-dividing files in pif4 cdfq was slightly narrower than that of Col-0, cdfq and pif4 plants (Fig. 1d). These histological analyses suggest that PIF4 and the CDFs act in the same genetic pathway to promote elongation of hypocotyl cells in non-dividing files.

CDF2 and PIF4 are co-expressed and physically interact. The temporal and spatial expression patterns of PIF4 and CDF2 were then compared. In a functional transgenic CDF2::HA-CDF2 cdf2-1 line (Extended Data Fig. 2a–c) grown under SDs, HA-CDF2 messenger RNA (mRNA) and protein exhibited similar diurnal cycles to those described for endogenous CDF2 (ref. 11), reaching maximum levels early in the light period (Fig. 2a,c). Under the same conditions, PIF4-HA mRNA and protein in transgenic PIF4::PIF4-HA pif4-101 plants28 also showed diurnal rhythms (Fig. 2b,d). Notably, the diurnal patterns of PIF4-HA and HA-CDF2 overlapped early in the light period, and were subsequently co-expressed for several hours (Fig. 2c,d).

To visualize the spatial accumulation of CDF2 and PIF4, transgenic lines expressing CDF2-mVenus and mScarlet-I-PIF4 fluorescent protein fusions from their native gene promoters were generated in cdf2-1 and pif4-2 mutants, respectively. Signals of both CDF2-mVenus and mScarlet-I-PIF4 were detected in the nuclei of epidermal cells of cotyledons and hypocotyls (Fig. 2e,f), consistent with the induction of hypocotyl growth by the epidermal-specific expression of PIF4 (ref. 17).

Whether PIF4 could physically interact with CDF2 in vivo was then tested. Plants that co-expressed 35S::PIF4-TAP (9Myc-6His-3Flag)22 and CDF2::HA-CDF2 were generated, and HA-CDF2 was
PIF4 and CDF2 interact in vitro and in vivo.

**Fig. 2 | Spatio-temporal expression patterns of CDF2 and PIF4, and physical interaction between the two proteins.** a, b, RT-qPCR analysis of levels of HA-CDF2 (a) and PIF4-HA (b) mRNA. Data are represented as means ± SEM of three independent amplifications. All values are normalized to APA1 levels. c, d, Western blots comparing the accumulation of HA-CDF2 (c) and PIF4-HA (d) proteins. Time (h) is expressed as hours from dawn (ZT, zeitgeber). Actin served as the loading control. RNA and protein were extracted throughout a SD in 6-day-old seedlings. Western blots in b and c were performed twice with similar results. e, f, Confocal microscopy analysis of epidermal cells of 6-day-old CDF2::CDF2-mVenus (e) and PIF4::mScarlet-1-PIF4 (f) transgenic plants grown under SD. Scale bars, 30 μm. g, HA-CDF2 protein co-immunoprecipitates with PIF4-TAP (9xMyc-6xHis-3xFlag) from 6-day-old SD-grown seedlings. Co-IP experiments in b and c were performed twice with similar results. h, Box diagram of various fragments of CDF2 and PIF4 used in Fig. 2. i, PIF4-Myc (C-terminal) interacts with HA-CDF2 (N-terminal) in vitro in the light. PIF4-Myc, HA-CDF2 and their truncation proteins were synthesized in a system attached to epitope tags (Fig. 2h). In vitro, PIF4-Myc and HA-CDF2 interact in vitro in the light. PIF4-Myc, HA-CDF2 and their truncation proteins were synthesized in a system attached to epitope tags (Fig. 2h).

co-immunoprecipitated with PIF4-TAP at ZT-0.5 in nuclear extracts from SD-grown seedlings (Fig. 2g). To understand the interaction domains between the two TFs, full-length CDF2 and PIF4 proteins as well as truncated versions were synthesized in a cell-free system attached to epitope tags (Fig. 2h). In vitro, PIF4-Myc and PIF4-C-Myc were co-immunoprecipitated with HA-CDF2-FL (full length) and HA-CDF2-N using an anti-HA antibody (αHA-IP). However, no immunoprecipitation was detected using HA-CDF2-C or PIF4-N-Myc. These results indicate that direct physical interaction occurred through the PIF4-C and CDF2-N-terminal regions, which contained the PIF4BHHLH and CDF2DOF DNA binding domains, respectively (Fig. 2i). Collectively, these experiments demonstrate that CDF2 and PIF4 are spatially and temporally co-expressed, and that they interact in vivo and in vitro.

PIF4 and CDF2 bind to and co-regulate common target genes.

PIF4 directly interacts with other TFs through their DNA-binding domains to recognize promoters of common target genes. We performed chromatin immunoprecipitation sequencing (ChIP-seq) to identify the in vivo binding sites of HA-CDF2 and to compare these with previously identified PIF4 binding sites. A total of 9,027 CDF2 binding peaks were identified and associated with 12,308 neighbouring genes (Supplementary information, Extended Data Fig. 3a and Supplementary Table 1). The majority (81.6%) of the peaks were within 3 kb of sequence 5′ to the transcription start site of a gene (Fig. 3a), consistent with the action of CDF2 as a transcriptional regulator. The canonical DOF-binding motif AAAAG was overrepresented (E value 1.7×10−14) in the centre of the ChIP-seq peaks (Fig. 3b,c), but the G-box (CAGCTG), which is recognized by PIF419,21, and closely related sequences were identified as the most enriched motifs (E value 8.5×10−29). About 20% of CDF2-binding peaks contained one G-box and approximately 9% contained more than one (Fig. 3d), with a peak in spacing distance of 25 bp (Fig. 3e). Similarly, 87% of CDF2-binding peaks contained two or more DOF-binding motifs, with a maximum of three per peak (Fig. 3f), and a most frequent spacing distance of 15 bp (Fig. 3g). To test whether PIF4 recognizes the G-boxes at CDF2 targets, we reanalyzed ChIP–seq data of PIF421 (Supplementary Table 2). Similar to the findings of previous studies21, the highest frequency (88.19%) of PIF4 occupancy was located within 3 kb of sequence 5′ to the transcription start sites of genes (Extended Data Fig. 4a), and G-boxes were remarkably enriched (E value 3.5×10−39) in the centre of PIF4-occupancy regions (Extended Data Fig. 4b). More than 480 (19%) PIF4 peaks contained at least two closely spaced G-boxes (Extended Data Fig. 4c), and these motifs showed a most frequent spacing distance of 30 bp, similar to the arrangement of G-boxes found in CDF2 targets (Extended Data Fig. 4d and Fig. 3e).

The target genes and occupancy regions of these two TFs were then compared. The overlap among CDF2 and PIF4 target genes was highly significant (Fig. 3h and Supplementary Table 3) (P value < 2.2×10−16). In total, 1,744 common peaks of CDF2 and PIF4 were identified (Extended Data Fig. 4e), and were closely spaced...
Fig. 3 | PIF4 and CDF2 bind to overlapping genomic targets and cooperatively regulate their expression. a, Positional distribution of CDF2 ChIP-seq peaks relative to the gene body. The observed distribution (black), the 95% confidence interval (shaded blue) and mean (dashed line) of 1,000 random peak sets are shown. b, Position distribution of G-box motifs relative to the CDF2 ChIP-seq peak centres. c, Position distribution of DOF-binding motifs relative to the centre of CDF2 ChIP-seq peaks. d, Frequency distribution of the number of G-box motifs observed in CDF2 ChIP-seq peaks. e, Density plots of the distance between consecutive G-box motifs in the observed CDF2 ChIP-seq peaks (blue) and the promoters of all non-CDF2 targets (red). f, Frequency distribution of the number of DOF motifs observed in CDF2 ChIP-seq peaks. g, Density of the distances between consecutive DOF motifs in observed CDF2 ChIP-seq peaks (blue) and the promoters of non-CDF2 targets (red). h, Overlap between target genes of CDF2 and PIF4. i, Density plot showing the distribution of distances between PIF4 and CDF2 ChIP-seq peaks. j, Positional distribution of G-box motifs relative to centres of PIF4/ CDF2 common ChIP-seq peaks. k, Frequency of G-box motifs observed in ChIP-seq peaks shared between PIF4 and CDF2. l, Density distribution of the distances between consecutive G-box motifs in ChIP-seq peaks common to the PIF4 and CDF2 peak sets (blue). The distribution is compared with that obtained by examining the promoters of all non-targets (red). m, Volcano plot describing the differential binding analysis of DiffBind package, by plotting the log2 fold change in binding strength against the $-\log_{10}P$ value of the differential binding test. The confidence threshold: false discovery rate ≤0.05. n, Venn diagram consistency in the direction of gene expression change in cdqF and pif4 mutants relative to that in Col-0. o, p, Heatmap showing the z-score normalized expression values of selected genes in the cotyledon (o) and hypocotyl (p) of Col-0, cdqF and pif4-2 mutants.
G-boxes were present in the coding region and other DOF-binding sites (AAAAG) located throughout the whole gene body were within closed chromatin regions and were not detected in the ChIP−seq of PIF4 and CDF2, respectively (Fig. 4a). The binding affinity of CDF2 in the YUC8 promoter region was much reduced when PIF4 was absent (Fig. 4a and Supplementary Table 4), supporting the notion that PIF4 recruits CDF2 to their common targets. To initiate gene transcription, RNA polymerase II (Pol II) assembles with general initiation factors at the promoter regions of genes to form the pre-initiation complex. During pre-initiation complex assembly, the Mediator coactivator complex bridges upstream TFs and RNA Pol II14, and in tomato, PIF4 induces transcription via interaction with the Mediator subunit 25 (MED25)15. Similarly, the enrichment of RNA Pol II along the transcribed region of YUC8 showed a significant decrease when PIF4 was absent (Fig. 4b), consistent with the lower level of YUC8 mRNA in the CDF2::HA-CDF2 pif4-2 cdf2-1 mutant (Figs. 3o and 4c).

Molecular basis for CDF2DOF binding to the YUCCA8 promoter. CDF2 is predicted to be highly disordered and apart from the DOF DNA-binding domain, which is highly conserved in all members of the family (Fig. 4e), no other structured domain(s) were predicted by AlphaFold16 and I-TASSER Suite17 (Extended Data Fig. 6a–c). Therefore, to understand in more detail how it binds to DNA, a structural model of the CDF2DOF domain was made based on the crystal structure of a zinc-finger (Zif268)−DNA complex18 (Methods and Fig. 4d). Similar to what is generally found in classical zinc-finger (ZF) proteins, one α-helix, which is inferred to contribute to DNA binding, and two β-sheets were also predicted in the CDF2DOF domain (Fig. 4d,e). Similar to the results of previous studies19, our modelling of the CDF2DOF domain showed that four cysteine residues (C140, C143, C165 and C168) in the conserved CXCXXCXXC motif are likely to bind a metal ion (probably Zn) (Fig. 4d,e and Extended Data Fig. 7a). To stabilize the CDF2 DOF DNA-binding domain (CDF2DOF), an N-terminal malone binding protein (MBP) was fused with it and the MBP-CDF2DOF protein was purified from Escherichia coli (Methods). Gel-filtration results showed that MBP-CDF2DOF protein was purified as a monomer (Extended Data Fig. 7b). To address the specific binding of CDF2DOF to DNA, electrophoretic mobility shift assays (EMSA) were performed using DNA probes from the YUC8 promoter (Fig. 4a,f). Fragment ‘a’, which contains five AAAAG motifs, was not bound by CDF2 in vivo (Fig. 4a), but was strongly bound by CDF2DOF in vitro (Fig. 4f,g), supporting the notion that an in vivo open chromatin status is critical for accessibility of CDF2 to DNA. Fragment ‘b’, which is within the CDF2- and PIF4-binding peak regions on YUC8 (Fig. 4a) and contains two G-boxes in addition to one AAAAG and one AAAG motif, was also bound by CDF2DOF (Fig. 4f,g). Mutation of the G-boxes did not affect binding of CDF2DOF (Fig. 4f,g), indicating that CDF2DOF does not interact with the G-box directly in vitro. By contrast, mutation of both the AAAAG and AAAG motifs abolished CDF2DOF binding, whereas mutation of the single motifs demonstrated that AAAAG was bound by CDF2DOF much more strongly than AAAG (Fig. 4f,g). The EMSA assay was then used to further identify the base pairs that are bound by CDF2DOF. The results indicated that the 5′-bp core of the DOF-binding motif T(A/A)AAAG led to the maximum binding affinity, explaining why the AAAAG motif in YUC8 is recognized more efficiently than the AAAG motif, and that the position of the 3′ G nucleotide is critical for the strength of binding of CDF2DOF to DNA in vitro (Fig. 4h,i).

We then tested the structural model by mutating the CDF2DOF protein sequence. Binding of the metal ion to the four cysteines was predicted to strongly stabilize the CDF2DOF domain in an appropriate conformation for DNA interaction. Consistent with this, adding divalent metal chelator ethylenediaminetetraacetic acid (EDTA) or mutating the four cysteines (C140, C143, C165 and

Open chromatin at common targets of PIF4 and CDF2. To understand in more detail how PIF4 and CDF2 coordinate transcriptional regulation, we focused on YUC8. Assay for Transposase-Accessible Chromatin using sequencing analysis showed that the binding peaks of PIF4 and CDF2 on YUC8 (ref. 20) were located in an open chromatin region (Fig. 4a). By contrast, the G-box (CAGCTG) in the coding region and other DOF-binding sites (AAAAG) located throughout the whole gene body were within closed chromatin regions and were not detected in the ChIP−seq of PIF4 and CDF2, respectively (Fig. 4a). The binding affinity of CDF2 in the YUC8 promoter region was much reduced when PIF4 was absent (Fig. 4a and Supplementary Table 4), supporting the notion that PIF4 recruits CDF2 to their common targets. To initiate gene transcription, RNA polymerase II (Pol II) assembles with general initiation factors at the promoter regions of genes to form the pre-initiation complex. During pre-initiation complex assembly, the Mediator coactivator complex bridges upstream TFs and RNA Pol II14, and in tomato, PIF4 induces transcription via interaction with the Mediator subunit 25 (MED25)15. Similarly, the enrichment of RNA Pol II along the transcribed region of YUC8 showed a significant decrease when PIF4 was absent (Fig. 4b), consistent with the lower level of YUC8 mRNA in the CDF2::HA-CDF2 pif4-2 cdf2-1 mutant (Figs. 3o and 4c).

Molecular basis for CDF2DOF binding to the YUCCA8 promoter. CDF2 is predicted to be highly disordered and apart from the DOF DNA-binding domain, which is highly conserved in all members of the family (Fig. 4e), no other structured domain(s) were predicted by AlphaFold16 and I-TASSER Suite17 (Extended Data Fig. 6a–c). Therefore, to understand in more detail how it binds to DNA, a structural model of the CDF2DOF domain was made based on the crystal structure of a zinc-finger (Zif268)−DNA complex18 (Methods and Fig. 4d). Similar to what is generally found in classical zinc-finger (ZF) proteins, one α-helix, which is inferred to contribute to DNA binding, and two β-sheets were also predicted in the CDF2DOF domain (Fig. 4d,e). Similar to the results of previous studies19, our modelling of the CDF2DOF domain showed that four cysteine residues (C140, C143, C165 and C168) in the conserved CXCXXCXXC motif are likely to bind a metal ion (probably Zn) (Fig. 4d,e and Extended Data Fig. 7a). To stabilize the CDF2 DOF DNA-binding domain (CDF2DOF), an N-terminal malone binding protein (MBP) was fused with it and the MBP-CDF2DOF protein was purified from Escherichia coli (Methods). Gel-filtration results showed that MBP-CDF2DOF protein was purified as a monomer (Extended Data Fig. 7b). To address the specific binding of CDF2DOF to DNA, electrophoretic mobility shift assays (EMSA) were performed using DNA probes from the YUC8 promoter (Fig. 4a,f). Fragment ‘a’, which contains five AAAAG motifs, was not bound by CDF2 in vivo (Fig. 4a), but was strongly bound by CDF2DOF in vitro (Fig. 4f,g), supporting the notion that an in vivo open chromatin status is critical for accessibility of CDF2 to DNA. Fragment ‘b’, which is within the CDF2- and PIF4-binding peak regions on YUC8 (Fig. 4a) and contains two G-boxes in addition to one AAAAG and one AAAG motif, was also bound by CDF2DOF (Fig. 4f,g). Mutation of the G-boxes did not affect binding of CDF2DOF (Fig. 4f,g), indicating that CDF2DOF does not interact with the G-box directly in vitro. By contrast, mutation of both the AAAAG and AAAG motifs abolished CDF2DOF binding, whereas mutation of the single motifs demonstrated that AAAAG was bound by CDF2DOF much more strongly than AAAG (Fig. 4f,g). The EMSA assay was then used to further identify the base pairs that are bound by CDF2DOF. The results indicated that the 5′-bp core of the DOF-binding motif T(A/A)AAAG led to the maximum binding affinity, explaining why the AAAAG motif in YUC8 is recognized more efficiently than the AAAG motif, and that the position of the 3′ G nucleotide is critical for the strength of binding of CDF2DOF to DNA in vitro (Fig. 4h,i).

We then tested the structural model by mutating the CDF2DOF protein sequence. Binding of the metal ion to the four cysteines was predicted to strongly stabilize the CDF2DOF domain in an appropriate conformation for DNA interaction. Consistent with this, adding divalent metal chelator ethylenediaminetetraacetic acid (EDTA) or mutating the four cysteines (C140, C143, C165 and
C168) to alanine (CDF2	extsuperscript{DOF} (Mut1)) diminished or abolished interaction between CDF2	extsuperscript{DOF} and DNA (Fig. 4e and Extended Data Fig. 7c). The structural model showed that the α-helix of CDF2	extsuperscript{DOF} could fit into the DNA major groove (22 Å), and contribute to DNA binding. Mutations in Y171 (Mut6: Y171A) and W172 (Mut7: W172A) within the α-helix abolished DNA binding (Fig. 4d,e,j), as observed previously for conserved residues in DOF AOBP (ascorbate oxidase promoter-binding protein)	extsuperscript{34}, and consistent with these residues contributing to DNA recognition. However, our modelled CDF2	extsuperscript{DOF}–DNA complex suggested that additional residue(s) in the
putative α-helix might interact with DNA (Fig. 4d). Mutation of K167 (Mu3: K167A) and Q169 (Mu4: Q169A) had no detectable effect on binding, but mutation of K166 (K166A: Mu2) and R170 (R177A: Mu5) strongly reduced DNA binding (Fig. 4e,j). These two conserved residues, which were recently identified to be important for DNA binding, are very close to C165 and C168 and conformational changes induced by metal binding might influence their accessibility to DNA. Similar to Mu7, mutation of the three residues of CDF2DOF interact with the five-nucleotide AAAAG motif within the major groove.

Molecular basis for PIF4bHLH binding to the YUCCA8 promoter.

Similar to CDF2, structural modelling predicted that PIF4 protein is not well ordered (Extended Data Fig. 6d–f), except for the bHLH–DNA-binding domain, which showed a strong preference for binding the G-box (5'-CACGTG-3') (Fig. 5a,b). To further understand the molecular basis of the interaction between the DNA-binding domain of PIF4 (PIF4bHLH) and DNA, we performed protein structure modelling based on the crystal structure of the MYC2 bHLH–DNA complex with G-box (Methods). This approach predicted that PIF4bHLH binds DNA as a homodimer and two interfaces in the basic region of PIF4bHLH (Extended Data Fig. 5a). EMSA experiments were performed with fragment 'b' that was used previously for the CDF2 bHLH (Extended Data Fig. 8) or by mutation of single G-boxes (Fig. 5f). Mutation of both G-boxes in the fragment because it could be reduced by using shorter DNA fragments containing only one G-box (Extended Data Fig. 5b). To verify our structural modelling, PIF4bHLH was purified via an N-terminal MBP fusion that conferred greater solubility on the protein. EMSA experiments were performed with fragment 'b' that was used previously for the CDF2 bHLH (Extended Data Fig. 8). To verify our structural modelling, PIF4bHLH was purified via an N-terminal MBP fusion that conferred greater solubility on the protein. EMSA experiments were performed with fragment 'b' that was used previously for the CDF2 bHLH (Extended Data Fig. 8). To verify our structural modelling, PIF4bHLH was purified via an N-terminal MBP fusion that conferred greater solubility on the protein. EMSA experiments were performed with fragment 'b' that was used previously for the CDF2 bHLH (Extended Data Fig. 8). To verify our structural modelling, PIF4bHLH was purified via an N-terminal MBP fusion that conferred greater solubility on the protein. EMSA experiments were performed with fragment 'b' that was used previously for the CDF2 bHLH (Extended Data Fig. 8). To verify our structural modelling, PIF4bHLH was purified via an N-terminal MBP fusion that conferred greater solubility on the protein. EMSA experiments were performed with fragment 'b' that was used previously for the CDF2 bHLH (Extended Data Fig. 8). To verify our structural modelling, PIF4bHLH was purified via an N-terminal MBP fusion that conferred greater solubility on the protein.
or Interface 2 (N263A, S265A, S266A, R269A and R270AA; PIF4bHLH(Mu1)) prevented binding to DNA (Fig. 5b.e,f), consistent with the bHLH homotetramer based on MYC2–DNA complex structure (Fig. 5g), and a mutant protein (PIF4bHLH(Mu3)) was designed to form tetramers in solution (Fig. 5c,d). Previously, MYC2, a bHLH TF that functions in jasmonate signalling, was shown to independently of strong interaction between them, EMSA was performed three times with similar results. Proposed model for the role of the PIF4-CDF2 module in regulating gene transcription in the light. Upon chromatin opening, the E-box and DOF-binding motifs are accessible. Binding of the G-boxes by PIF4, and the DOF-binding motifs near those G-boxes by CDF2 occurs. Interaction between PIF4 and CDF2 can occur when they are bound to DNA. Binding of PIF4 to the G-boxes strengthens CDF2 binding and allows it to bind to DNA independently of the DOF-binding sites. Therefore, PIF4 recruits CDF2 to the YUCCA8 promoter. After binding of the PIF4–CDF2 module to chromatin, RNA polymerase II is recruited by PIF4 via the Mediator complex to induce gene transcription.

**Fig. 6 | Binding of PIF4 on the G-boxes enhances CDF2-binding strength on DNA.** a. Binding strengths for CDF2 peaks varied significantly with the numbers of G-box motifs (CACGTG, Kruskal–Wallis test, $P = 4.588 \times 10^{-4}$). b. Binding strengths for peaks with different numbers of DOF-binding motifs (AAAGAG, Kruskal–Wallis test, $P = 1.964 \times 10^{-4}$). Box plots in panel a and b show the minimum, 25th percentile, median, 75th percentile and maximum of data points. Letters in panels a and b show significant differences among groups (adjusted $P < 0.05$) using pairwise Wilcoxon tests. Groups that share at least one identical letter are not significantly different. c. Gel-shift analysis of the interactions between single PIF4bHLH or CDF2DOF WT proteins with WT or mutant DNA probes. d. Gel-shift analysis of the interactions between single PIF4bHLH or CDF2DOF or combination of PIF4bHLH(Mu2) WT or PIF4bHLH(Mu3) mutant proteins and CDF2DOF WT protein with WT or mutant DNA probes. e. Interactions between single PIF4bHLH or CDF2DOF or combinations of PIF4bHLH WT and CDF2DOF WT or CDF2DOF(Mu2 and Mu3) mutant proteins with WT or mutant DNA probes analyzed by EMSA. EMSA assays in c, d and e were performed three times with similar results. f. Proposed model for the role of the PIF4–CDF2 module in regulating gene transcription in the light. Upon chromatin opening, the E-box and DOF-binding motifs are accessible. Binding of the G-boxes by PIF4, and the DOF-binding motifs near those G-boxes by CDF2 occurs. Interaction between PIF4 and CDF2 can occur when they are bound to DNA. Binding of PIF4 to the G-boxes strengthens CDF2 binding and allows it to bind to DNA independently of the DOF-binding sites. Therefore, PIF4 recruits CDF2 to the YUCCA8 promoter. After binding of the PIF4–CDF2 module to chromatin, RNA polymerase II is recruited by PIF4 via the Mediator complex to induce gene transcription.

**Binding of PIF4bHLH to YUCCA8 enhances CDF2DOF binding.** ZF proteins usually contain several tandemly arranged ZF motifs that strengthen interaction of the protein with DNA. CDF2 and other DOF proteins contain only a single ZF motif. However, the ChIP-seq analysis showed that CDF2 binding was highly correlated with the presence of G-boxes (Fig. 6a) but less so with AAAAG motifs (Fig. 6b), suggesting that PIF4 may enhance binding of CDF2 to adjacent AAAAG motifs, and thereby strengthen its specificity for particular genomic regions. On the other hand, PIF4 might also alter CDF2 specificity, because not all CDF2-binding regions that contained G-boxes also contained an adjacent AAAAG motif. Although the precise mechanism remains unknown, interaction between the two TFs might have a role in both scenarios. The involvement of protein interaction is consistent with the CDF2–PIF4 interaction observed in vivo and in vitro, which could occur through CDF2-N-terminal (1–201 amino acids (aa)) and PIF4–C-terminal (248–431 aa) regions (Fig. 2i). However, interaction between CDF2DOF (134–201 aa) and PIF4bHLH(Mu2) (248–317 aa) domains used in the EMSA experiments was not detected by gel-filtration in vitro (Extended Data Fig. 10) indicating that the DNA-binding domains alone do not interact or do so very weakly. To test whether PIF4bHLH enhanced binding of CDF2DOF in vitro independently of strong interaction between them, EMSA was performed with the PIF4bHLH and CDF2DOF domains in combination. A supershift was detected with both proteins (Fig. 6c.1), indicating that they can bind to the same YUCCA8 fragment. Unexpectedly, a supershift was still detected when the AAAAG and AAAG motifs were mutated, although to a lesser extent (Fig. 6c.2). By contrast, the supershift was strongly reduced when both CAGCTG G-boxes were
mutated (Fig. 6c,3). No shift or supershift was detected when both G-boxes and DOF-binding motifs were mutated (Fig. 6c,4). These results indicate that PIF4<sup>BHLH</sup> binding to G-boxes is a determinant for the supershift, but that the DOF-binding motifs are not required. Consistently, the supershift was significantly reduced when mutant PIF4<sup>BHLH</sup> (Mut1) protein that cannot bind DNA was combined with wild-type DNA probe and CDF2<sup>DOF</sup> (Figs. 5 and 6,d, left). Also, use of PIF4<sup>BHLH</sup> (Mut2) abolished the supershift observed with PIF4<sup>BHLH</sup> and CDF2<sup>DOF</sup> on the mutated DOF-binding motif DNA (Fig. 6,d, panel right). Furthermore, no supershift was detected when PIF4<sup>BHLH</sup> was combined with CDF2<sup>DOF</sup> (Mut1) or CDF2<sup>DOF</sup> (Mut2), regardless of the presence of the DOF-binding motif (Fig. 4 and Extended Data Fig. 6). Therefore, the α-helix that is required for DNA binding by CDF2<sup>DOF</sup> (Fig. 4) is required for the supershift with PIF4<sup>BHLH</sup>, even for DNA fragments that do not contain the AAAAG and AAAG motifs, and although these two truncated proteins do not detectably interact in vitro. These results suggest that binding of PIF4<sup>BHLH</sup> to DNA induces other potential DNA interaction interface(s) of CDF2<sup>DOF</sup> to access DNA, a process related to DNA allostery that was previously described<sup>14,25</sup>.

**Discussion**

Hypocotyl cell elongation in the light requires cooperation between CDF2 and PIF4 to strongly activate transcription of YUC8 in cotyledons, consistent with the previous observation that increased auxin biosynthesis in cotyledons through the action of YUCCA enzymes promotes hypocotyl growth<sup>9,14</sup>. We show that in vivo, CDF2 binding strength and specificity are increased in the presence of PIF4 at a subset of common target genes, including YUC8, and this probably involves direct interaction between the proteins. Furthermore, in vitro, in the presence of PIF4<sup>BHLH</sup>, CDF2<sup>DOF</sup> binds to a DNA fragment in which both DOF-binding sites are mutated, and as interaction of these two truncated proteins was undetectable by gel-filtration, this suggests that the binding of PIF4<sup>BHLH</sup> may enhance the general affinity of CDF2<sup>DOF</sup> for PIF4<sup>BHLH</sup>–DNA complex by DNA allostery<sup>14</sup>. TFs recognize their binding sites by directly interacting with specific bases, and by recognizing features of local DNA shape, such as DNA bending or unwinding<sup>14</sup>. Thus, we propose that PIF4 increases the strength and specificity of CDF2 DNA binding through protein–protein interactions that enhance sequence-specific DNA binding, and by altering local DNA shape. Whether CDFs influence PIF4 binding in vivo remains to be tested. We demonstrated that PIF4 forms tetramers and these may facilitate DNA looping, as demonstrated for MYC2 tetramers<sup>19</sup>, and thereby strengthen CDF2 binding at adjacent sites. The combinatorial interactions between PIF4 and CDF2 that we elucidated contribute to auxin biosynthesis and hypocotyl elongation in the light. Our results also enhance understanding of the transcriptional code that regulates plant gene expression in response to light and how this contributes to cell elongation.

**Methods**

**Plant material and growth conditions.** The Arabidopsis thaliana (Col-0) Columbia ecotype was used as the main experimental organism. Seeds of Col-0, pif4-2 (SAIL_1288_E07), cdf2<sub>1</sub>,<sub>2</sub>,<sub>3</sub>,<sub>5</sub> (cdf5) and pif4-2 cdf5 were surface-sterilized with 70% ethanol for 10 min, rinsed with 99% ethanol for 5 min, air-dried and stratified at 4 °C for 3 days. Plants were grown on soil under SD conditions (8 h light/16 h dark) or were grown vertically on plates containing 1% agar (ethylenedinitrilotetraacetic acid, Sigma-Aldrich) pH 8.0, 4% SDS (sodium dodecyl sulfate, Sigma-Aldrich), 20% glycerol, 20 mM β-mercaptoethanol (Sigma-Aldrich), 20 mM DTT (DL-dithiothreitol, Sigma-Aldrich), 2 mM PMSF (phenylmethylsulfonyl fluoride, Sigma-Aldrich), 1x Protease Inhibitor Cocktail (Sigma-Aldrich), 80 µM MG132 (Sigma-Aldrich), 1% Halt Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) and 0.01% bromophenol blue in a 1:5 tissue-buffer (w/v) ratio by boiling for 10 min at 90 °C. Samples were centrifuged at 16,000 g for 5 min at room temperature and the supernatants were electrophoresed on a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) mini-gel to separate the proteins.

**Immunoblot assay.** For SD time-course western blots, about 20 mg of tissue from 6-day-old seedlings was ground into fine powder in liquid nitrogen with a TissueLyser system (QIAGEN). Total proteins were extracted using denaturing buffer (100 mM Tris-HCl pH 7.5, 100 mM NaCl, 30 mM EDTA (ethylenedinitrilotetraacetic acid, Sigma-Aldrich) pH 8.0, 20% glycerol, 20 mM β-mercaptoethanol (Sigma-Aldrich), 20 mM DTT (DL-dithiothreitol, Sigma-Aldrich), 2 mM PMSF (phenylmethylsulfonyl fluoride, Sigma-Aldrich), 1x Protease Inhibitor Cocktail (Sigma-Aldrich), 80 µM MG132 (Sigma-Aldrich), 1% Halt Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) and 0.01% bromophenol blue) in a 1:5 tissue-buffer (w/v) ratio by boiling for 10 min at 90 °C. Samples were centrifuged at 16,000 g for 5 min at room temperature and the supernatants were electrophoresed on a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) mini-gel to separate the proteins. The immunoblotting separated proteins were transferred onto a polyvinylidene fluoride membrane by the Trans-Blot Turbo Transfer System (Bio-rad). Blots were probed with anti-HA (21D3, Roche) or anti-Actin (sc-47778, Santa Cruz) antibodies conjugated to horseradish peroxidase (HRP). The blots were developed with a 1:1 mix of SuperSignal West Femto Maximum Sensitivity and SuperSignal West Dura Extended Duration Substrates and signals were detected on a ChemiDoc MP Imaging System (Bio-rad). Anti-HA (HRP) and anti-Actin (HRP) antibodies were used at 1:2,000 and 1:4,000-fold dilutions, respectively.

**In vitro pull-down assay.** To express proteins in a cell-free system, 2H-ACDF2, PIF4–6My and the truncated DNA fragments were amplified by overlap PCR, then cloned into EcoRI-digested pTrX1 vector with the In-Fusion HD Cloning Kit (Stratagene). For each construct 2H-ACDF2, PIF4–6My and truncated proteins (CDF2–N, 1–201 aa; CDF2–C, 202–457 aa; PIF4–N, 1–248 aa; PIF4–C, 249–431 aa)), 1 µg plasmid was expressed via the SP6 promoter in a cell-free system at 30 °C for 2 h in a thermocycler using the TriT Coupled Wheat Germ Extract System (Promega), according to the manufacturer’s instructions. A small amount of the reaction (2 µl) was used to verify expression of the target proteins by western blotting and the remaining extract (48 µl) was ‘snap-frozen’ in liquid nitrogen and stored at ~80 °C.

For protein pull-downs, 60 µl extract (30 µl of 2HA–CDF2 or 30 µl of 2HA–CDF2 and 30 µl of PIF4–6My) was mixed with 540 µl IP buffer (22 mM Tris-HCl, pH 7.5; 84 mM NaCl, 1 mM EDTA; 0.11% Triton X-100 and 0.1% Plant Protease Inhibitor Cocktail (Sigma)) and rotated in the dark at 4 °C for 30 min. To pull down proteins, 30 µl of Dynabeads Protein G beads (Thermo Fisher Scientific) coated with 10 µl anti-Myc antibody (Cell Signaling Technology) was added to the diluted protein solution and was rotated for 30 min at 4 °C. The beads were washed five times with wash buffer (20 mM Tris-HCl, pH 7.5; 150 mM NaCl, 1 mM EDTA; 0.5% Triton X-100 and 0.1% Plant Protease Inhibitor Cocktail). Proteins were eluted from the beads with 2x SDS–PAGE sample buffer and then subjected to immunoblotting analysis. Anti-HA (HRP) and anti-Myc (HRP-conjugated, 2040S, CST) antibodies were used at 1:5,000–fold dilution.

**In vivo co-immunoprecipitation assays.** The in vivo co-immunoprecipitation (Co-IP) assays were performed as previously described, with minor modifications<sup>17</sup>. In brief, 1 g of 6-day-old SD-grown F<sub>1</sub> seedlings (35S::PIF4–TAP x CDF2::3HA–CDF2) was harvested at ZT-1. The seedlings were ground to fine powder in liquid nitrogen, semi-pure nuclei extractions were performed and nuclear proteins were released by a short sonication.
For co-immunoprecipitation, 30 μl of Dynabeads Protein G beads coated with 10 μl anti-Myc antibody was added to the diluted nuclear protein solution (0.5% Triton X-100, 1 mM EDTA, 20 mM Tris-HCl, pH 7.5, and 100 mM NaCl and 1× Protease Inhibitor Cocktail (Sigma–Aldrich)) and rotated for 45 min at 4°C. The beads were washed five times with IP buffer. Proteins were eluted from the beads with 2× SDS–PAGE sample buffer and then subjected to immunoblotting analysis. For immunoblotting, anti-HA (HRP) and anti-Myc (HRP) antibodies were used at 1:2,500 fold dilution.

Chromatin immunoprecipitation. ChIP methods were described previously with minor modifications38. For ChIP–seq of CDF2, 9g above-ground tissue of 6-day-old SD-grown seedlings was harvested at ZT-1 and cross-linked for 10 min by vacuum filtration in phosphate-buffered saline solution containing 1% formaldehyde. For chromatin immunoprecipitation, 50 μl of Dynabeads Protein G beads (coated with 5 μl of a 1× antibody coated with 5 μl of antibody (Abcam)) was incubated for 4h with 5 ml of the diluted chromatin solution (1% Triton X-100, 1 mM EDTA, 0.08% SDS, 15 mM Tris-HCl, pH 8.0, and 150 mM NaCl). After washing three times with wash buffer (1% NP-40, 1 mM EDTA, 0.1% SDS, 0.1% DOC (sodium deoxycholate, Sigma–Aldrich), 20 mM Tris-HCl, pH 8.0, and 150 mM NaCl), the immune complex was eluted from the beads in 400 μl elution buffer (1% SDS and 0.1 M NaHCO3). Next, samples were reverse cross-linked with 5 μl Proteinase K and 20 μl 5 M NaCl at 65°C overnight and DNA was purified by a MinElute PCR Purification Kit (Qiagen). Amounts of input DNA were quantified by fluorometry (Quantus, Promega) and the size of the fragments was analyzed by ultra-sensitive capillary electrophoresis (Agilent FEMTOpulse). ChIP–seq libraries were generated according to Ovation Ultralow System (Bio-Rad), the primers used for DNA probes are listed in Supplementary Table 1. Further information on research design is available in the paper (RT–qPCR) are listed in Supplementary Table 1.

CentriMo was used to determine the enrichment of motifs in the centre of peaks. Distances between consecutive motifs and the number of motifs per peak were obtained using custom python scripts. Differential binding assays between CDF2 peak locations between CDF2:HA-CDF2; cdf2-1 and CDF2: HA-CDF2; cdf2-1 pif4-2 were performed with DiffBind (Biocomputer, https://biocomputer.org/package/release/biochip/DiffBind.html).

ChIP–qPCR of RNA polymerase II, 9g of above-ground tissue from 6-day-old SD-grown seedlings was harvested at ZT-1, cross-linked with 1 mM formaldehyde. The immune complex with G-box (PDB ID 5CGN)39. The structure data were processed using the program Coot and PyMOL softwares.

Gel-shift assay (EMSA). The long double-stranded DNA probe (95 bp) covering the two G-boxes and two DOF-binding sites was synthesized by PCR using 5′-Cy5-labelled oligo primers. The short double-stranded DNA probe (38 bp) covering one G-box and one DOF-binding site was synthesized by annealing single-stranded 5′-Cy5-labelled oligo in annealing buffer (10 mM Tris (pH 8.0), 50 mM NaCl, and 1 mM EDTA (pH 8.0)). Binding reactions were carried out in buffer containing 10 mM Tris, 50 μg/ml of Poly (dI-dC), 50 mM KCl, 10 mM MgCl2, 1 mM DTT, 5% glycerol and 0.1% NP-40. Samples were kept in the light on ice for 30 min and were then loaded onto 6% DNA Retardation Gels (Thermo Fisher Scientific) and run in 0.5X Tris/Borate/EDTA buffer at room temperature at 90 min at 70 V. Binding signals were visualized using a ChemiDoc MP Imaging System (Bio-Rad). The primers used for DNA probes are listed in Supplementary Table 1.

Protein expression and purification. Codons of the coding sequences of CDF2 and PIF4 domains from Arabidopsis thaliana were optimized to E. coli and cloned into pMAL-c5X-His Vector (NEB). CDF2 domain was between 133 and 201 aa, whereas the PIF4 domain was between 248 and 317 aa. The wild-type CDF2, PIF4 and mutant proteins were induced by 0.7 m IFPTG (Sigma) and expressed in Arcithalian Express cells (Agilent Technologies) at 12°C, overnight. The E. coli cells were collected by centrifugation, resuspended in wash buffer (25 mM Bis-Tris pH 8.0, 150 mM NaCl and 15 mM imidazole) and sonicated to prepare cell lysates. The proteins were purified using Ni-NTA beads (GE), the bound proteins were washed five times with wash buffer and eluted using elution buffer (25 mM Bis-Tris pH 8.0, 150 mM NaCl and 250 mM imidazole). The eluted proteins were further purified by size-exclusion chromatography (HLoad 200, GE Healthcare) in buffer containing 25 mM Bis-Tris pH 8.0, and 150 mM NaCl.

Structural modelling. The structures of CDF2 and PIF4 full length, and CDF2 DOF and PIF4 bHLH domain were predicted using AlphaFold40 and I-TASSER Suite41, respectively. The modelled structure of CDF2–DOF–DNA complex was based on the zif288–DNA complex (PDB ID 1ZAA)42. The dimer and tetramer modelled structures of PIF4–bHLH were based on MYC2–bHLH–DNA complex with G-box (PDB ID 5CNY)43. The structure data were processed using the program Coot and PyMOL softwares.

Received: 13 October 2021; Accepted: 4 July 2022; Published online: 15 August 2022

References
1. Stathopoulos, A. & Levine, M. Genomic regulatory networks and animal development. Dev. Cell 9, 449–462 (2005).
2. Riehmann, J. L. et al. Arabidopsis transcription factors: genome-wide integrates brassinosteroid and environmental responses. Nat. Cell Biol. 802–809 (2012).
3. Riechmann, J. L. et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290, 2105–2110 (2000).
4. Zinnen, R., Pirrung, C., Gagneur, J., Braun, M. & Furlong, E. E. Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature 462, 65–70 (2009).
5. Oh, E., Zhu, J. Y. & Wang, Z. Y. Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses. Nat. Cell Biol. 14, 802–809 (2012).
6. Slattery, M. et al. Absence of a simple code: how transcription factors read sequences of Dof transcription factors. Plant Syst. 17, 209–214 (1999).
8. Yanagisawa, S. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol. 45, 386–391 (2004).
9. Kloosterman, J. et al. Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495, 246–250 (2013).
10. Imazumi, T., Schultz, T. F., Harmon, F. G., Ho, L. A. & Kay, S. A. FK506 F-box protein mediates cytidic degradation of a repressor of CONSTANS in Arabidopsis. Science 309, 293–297 (2005).
11. Fornara, F. et al. Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. Dev. Cell 17, 75–86 (2009).
12. Martin, G. et al. Circadian waves of transcriptional repression shape PIF-regulated photoperiod-responsive growth in Arabidopsis. Curr. Biol. 28, 311–318 (2018).
13. Correllas, A. R. et al. Characterization of tomato Cycling Dof Factors reveals conserved and new functions in the control of flowering time and abiotic stress responses. J. Exp. Bot. 65, 995–1012 (2014).
14. Fornara, F. et al. The GI-CDF module of Arabidopsis affects freezing tolerance and growth as well as flowering. Plant J. 81, 695–706 (2015).
15. Slattery, M. et al. Cofactor binding evokes latent differences in RNA binding specificity between Hox proteins. Cell 147, 1270–1282 (2011).
16. Lee, S., Wang, W. L. & Huq, E. Spatial regulation of thermomorphogenesis by HY5 and PIF4 in Arabidopsis. Nat. Commun. 12, 3656 (2021).
17. Kim, S. et al. The epidermis coordinates thermoresponse growth through the phyB-PIF4auxin pathway. Nat. Commun. 11, 1053 (2020).
18. Huq, E. & Quail, P. H. PIF4, a phytochrome-interacting, bHLH factor, functions as a negative regulator of phytochrome B signaling in Arabidopsis. EMBO J. 21, 2441–2450 (2002).
19. Hornitschek, P. et al. Phytochrome interacting factors 4 and 5 control seedling growth in changing light conditions by directly controlling auxin signaling. Plant J. 71, 699–711 (2012).
20. Komi, M. A. et al. High temperature-mediated adaptations in plant architecture require the bHLH transcription factor PIF4. Curr. Biol. 19, 408–413 (2009).
21. Pedmale, U. V. et al. Cryptochrome interact directly with PIFs to control plant growth in limiting blue light. Cell 164, 233–245 (2016).
22. Ma, D. B. et al. Cryptochrome 1 interacts with PIF4 to regulate high temperature-mediated hypocotyl elongation in response to blue light. Proc. Natl Acad. Sci. USA 113, 224–229 (2016).
23. Nozue, K. et al. Rhythmic growth explained by coincidence between internal clock and environmental cues. Nature 448, 358–361 (2007).
24. Greenham, K. & McCllng, C. R. Integrating circadian dynamics with physiological processes in plants. Nat. Rev. Genet. 16, 596–610 (2015).
25. Martin, G. et al. The photoperiodic response of hypocotyl elongation involves regulation of CDF1 and CDF5 activity. Physiol. Plant. 169, 480–490 (2020).
26. Ma, D. B. et al. Cryptochrome 1 interacts with PIF4 to regulate photoperiodic response mediated by CDF1 and CDF5. Cell 164, 233–245 (2016).
27. Nomoto, Y., Kubozono, S., Yamashino, T., Nakamichi, N. & Mizuno, T. Circadian clock- and PIF4-controlled plant growth: a coincidence model that integrates circadian gating of thermoresponsive growth in Arabidopsis. Plant Signal Behav. 8, e12309 (2013).
28. Zhu, J. Y., Oh, E., Wang, T. & Wang, Z. Y. TOC1-PIF4 interaction mediates the circadian gating of thermomorphogenesis in Arabidopsis. Nat. Commun. 7, 13692 (2016).
29. Lu, Z. et al. Linking photoreceptor excitation to changes in plant architecture. Gene Dev. 26, 785–790 (2012).
30. Wang, Z. Y. & Tobin, E. M. Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. Cell 93, 1207–1217 (1998).
31. Lu, Z. et al. The prevalence, evolution and chromatin signatures of plant regulatory elements. Nat. Plants 5, 1250–1259 (2019).
32. Osman, S. & Garner, P. Structural biology of RNA polymerase II transcription: Cells Mol. Biol. 56, 1–17 (2020).
33. Sun, W. et al. Mediator subunit MED25 physically interacts with PHOTOTRACE INTERACTING FACTOR4 to regulate shade-induced hypocotyl elongation in tomato. Plant Physiol. 184, 1549–1562 (2020).
34. Jumpers, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
35. Yang, J. et al. The 1-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8 (2015).
36. Pavletich, N. P. & Babo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252, 809–817 (1991).
Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41477-022-01213-y.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-022-01213-y.
Correspondence and requests for materials should be addressed to George Coupland.
Peer review information Nature Plants thanks Enamul Huq and Xiao-Dong Su for their contribution to the peer review of this work.
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.
© The Author(s) 2022
Extended Data Fig. 1 | CDF1, 2, 3 and 5 function in the same genetic pathway with PIF4 to induce hypocotyl elongation. CDF1, 2, 3 and 5 function in the same genetic pathway with PIF4 to induce hypocotyl elongation. **a**, Hypocotyl length of 9-day-old seedlings of wild type, pif4, cdfq and pif4 cdfq in short-day conditions. **b**, Photographs of 4-day-old dark-grown seedlings. Scale bar = 1 mm. **c**, Hypocotyl length and **d**, hook curvature of 4-day-old dark-grown seedlings of WT, pif4, cdfq and pif4 cdfq. **e**, Length and **f**, width of cells in the non-dividing cell files of hypocotyls. Box plots in panel **a**, **b**, **d**, **e** and **f** show the minimum, 25th percentile, median, 75th percentile and maximum of data points. In **a** and **c**, n = 29 and 28 hypocotyls. In **d** and **e**, n = 8 cell files examined over four hypocotyls (cell numbers are presented in Fig. 1b) in **e** and **f**. Letters a–c in panels **a**, **b** and **d** show significant differences between genotypes (P < 0.05, using one-way ANOVA followed by Tukey test, two sided). P = 1.96 *10⁻²⁸ in **a**, P = 0.16 in **b** and P = 0.067 in **d**.
Extended Data Fig. 2 | Flowering time and hypocotyl length of transgenic plants carrying CDF2::HA-CDF2. Flowering time and hypocotyl length of transgenic plants carrying CDF2::HA-CDF2. a, Flowering time of WT, cdf2-1 and CDF2::HA-CDF2 transgenic lines in long-day conditions. The total leaf number was counted for plants of WT, cdf2-1, and T1 homozygous populations of three independent transgenic lines. Data are presented as means ± SEM; n = 14 for WT, cdf2-1, #7 and #3 and n = 13 for #10 in a. Box plots in panel a show the minimum, 25th percentile, median, 75th percentile and maximum of data points. n = (14, 14, 14, 14). Letters in panels a show significant differences between genotypes (P < 0.05, using one-way ANOVA followed by Tukey test, two sided), P = 7.254×10⁻⁴⁵. b, Hypocotyl length of 9-day-old seedlings of WT, cdf2-1, CDF2::HA-CDF2 (#3), cdf2-1 and CDF2::HA-CDF2 (#3) cdf2-1 pif4-2 in short-day conditions. Box plots in panel b show the minimum, 25th percentile, median, 75th percentile and maximum of data points. n = (30, 29, 29, 30). Letters in panels b show significant differences between genotypes (P < 0.05, using one-way ANOVA followed by Tukey test, two sided), P = 1.298×10⁻⁴². c, RT-qPCR analysis of HA-CDF2 mRNA levels in cotyledons and hypocotyls of cdf2-1 or cdf2-1 pif4-2 mutants. All values are normalized to APA1 levels. Data are presented as mean values ± SEM of three biological replicates. Statistical significance was determined by pairwise one-sided t-test (P = 0.4964 and 0.3869 in cotyledon and hypocotyl, respectively). Significant P < 0.05. NS, not significant. d, Western blotting analysis of the abundance of HA-CDF2 protein in cotyledons and hypocotyls in cdf2-1 or cdf2-1 pif4-2 mutants. Actin was used as the loading control. Western blots represent one of three independent biological replicates. Tissues were harvested at ZT-0.5 to ZT-1 from 7-day-old SD-grown seedlings in c and d.
Extended Data Fig. 3 | High reproducibility among the three biological replicates of ChIP-seq experiments with CDF2::HA-CDF2 in cdf2-1 and cdf2-1 pif4-2 backgrounds. High reproducibility among the three biological replicates of ChIP-seq experiments with CDF2::HA-CDF2 in cdf2-1 and cdf2-1 pif4-2 backgrounds. a, Pearson correlation between the read coverage along the genome for input and chromatin-immunoprecipitated (IP) samples. b, CDF2-binding profile for the CONSTANS (CO) gene. The panels display two biological replicates for the ChIP-seq control in Col-0 and three biological replicates for ChIP-seq of CDF2 to the CO locus, visualized with the Integrated Genome Browser (IGB). c, CDF2- and PIF4-binding profiles to YUCCA8 and d to CIRCADIAN CLOCK ASSOCIATED1 (CCA1). Each set of panels displays two biological replicates for the ChIP-seq control in Col-0, three biological replicates for ChIP-seq of CDF2 in cdf2-1 and cdf2-1 pif4-2 backgrounds, and a single biological replicate for ChIP-seq of PIF4, visualized with the IGB.
Extended Data Fig. 4 | Re-analysis of PIF4 ChIP seq data. Re-analysis of PIF4 ChIP seq data. a, Positional distribution of PIF4-binding peaks on Arabidopsis genes. The observed distribution (black), the 95% confidence interval (shaded blue) and mean (dashed line) of 1,000 random peak sets are shown. b, Histogram showing the positional distribution of G-boxes relative to the peak centre (E-value = 3.5e-39). c, Histogram showing the number of PIF4 peaks with different numbers of G-box motifs. d, The distance between consecutive G-box motifs in a single PIF4 peak. The density of the distance between consecutive G-boxes in the promoters of all non-PIF targets was used as a control. e, Overlapping gene set between PIF4 and CDF2 peaks. Note that in a limited number of cases, multiple peaks in a single sample (merged peaks) overlap with a single peak in the other sample. f, Overlapping gene set between PIF4 peaks and CDF2 peaks with significantly reduced binding affinity in the pif4 mutant. Similar to in e, multiple peaks in a single sample can overlap with a single peak in the other sample. g, GO-term enrichment based on the gene set between PIF4 peaks and CDF2 peaks with significantly reduced binding affinity in the pif4 mutant. Enrichment tests were performed with topGO for each GO term separately using the Fisher Exact test. Raw p-values were adjusted using the Benjamini-Hochberg procedure. GO terms with adjusted p-values < 0.05 were considered significantly enriched.
Extended Data Fig. 5 | Transcriptome gene-expression profiling in experiments performed with different genetic backgrounds and tissues.

Transcriptome gene-expression profiling in experiments performed with different genetic backgrounds and tissues. **a.** Twenty cotyledons and fifty hypocotyls were sampled for RNA-seq. **b.** The two major principal components (PC) from a PC analysis of the log2(FPKM+1)-transformed gene expression levels in the different samples. **c** and **d.** The number of significantly up- and downregulated genes in the cotyledons (c) and hypocotyls (d) of *pif4* and *cdfq* mutants relative to the wild type. **e** and **f.** The number of genes showing the same and opposite directions of regulation in the cotyledon (e) and hypocotyls (f) of *pif4* and *cdfq* mutants relative to the wild type. A hypergeometric test showed that the probability of the overlap between the differentially expressed genes (DEGs) in *pif4* and *cdfq* for cotyledons and hypocotyls was 4.882719e-74 and 0, respectively, indicating that the difference in the expression level of genes in these transcriptomes is higher than expected by chance. **g.** Overlapping gene set between CDF2 targets and DEGs in the *cdfq* mutant relative to the wild type. **h.** Overlapping gene set between PIF4 targets and differentially expressed genes in the *pif4* mutant relative to the wild type.
Extended Data Fig. 6 | The predicted CDF2 and PIF4 protein structure. The predicted CDF2 and PIF4 protein structure. a, Predicted aligned error of CDF2. The colour at position (x, y) indicates AlphaFold’s expected position error at residue x, when the predicted and true structures are aligned on residue y. b, Overall view of predicted CDF2 protein with model confidence. c, Prediction of intrinsically disordered tendency of CDF2 (http://www.pondr.com/). d, Predicted aligned error of PIF4. The colour at position (x, y) indicates AlphaFold’s expected position error at residue x, when the predicted and true structures are aligned on residue y. e, Overall view of predicted PIF4 protein with model confidence. f, Prediction of the intrinsically disordered tendency of PIF4.
**Extended Data Fig. 7** | The putative interaction between a metal ion and the four cysteine residues stabilizes the structure of the CDF2^DOF domain in an appropriate conformation for DNA binding. The putative interaction between a metal ion and the four cysteine residues stabilizes the structure of the CDF2^DOF domain in an appropriate conformation for DNA binding. 

**a**, Modelled structure of the CDF2^DOF domain with the four cysteine residues (C140, C143, C165 and C168) highlighted. Removal of the metal ion by adding a divalent metal chelator destabilizes the structure. 

**b**, Size-exclusion chromatography analysis of CDF2^DOF WT protein. The CDF2^DOF domain is fused with an MBP tag at the N terminus. The x-axis and y-axis indicate the elution volume and the protein absorption at 280 nm, respectively. 

**c**, Gel-shift analysis of the interactions between CDF2^DOF WT, Mu1 proteins and DNA in various concentrations of EDTA. Size-exclusion chromatography analysis in **b** and EMSA assays in **c** were performed twice with similar results.
Extended Data Fig. 8 | Gel-shift analysis of the interactions between PIF4^KSL^ WT protein and different length of DNA probes. Gel-shift analysis of the interactions between PIF4^KSL^ WT protein and different length of DNA probes (as described in Fig. 4f). Several different sizes of complexes of PIF4^KSL^ bound to the longer DNA fragment, `b` could be reduced using shorter DNA fragments (fragment `c`) containing only one G-box. EMSA assays were performed twice with similar results.
Extended Data Fig. 9 | A multiple sequence alignment for PIFs and MYC2 and MYC3 proteins. A multiple sequence alignment for PIFs and MYC2 and MYC3 proteins (https://esprid.ibcp.fr/ESPript/ESPript/). Black bar, residues of predicted PIF4bHLH domain. Light blue bar, residues of MYC2 bHLH domain. Yellow triangle, residues involved in MYC2bHLH dimer formation.
Extended Data Fig. 10 | Size-exclusion chromatography analysis of CDF2\textsuperscript{Dof} and PIF4\textsuperscript{HLH} WT proteins. Size-exclusion chromatography analysis of CDF2\textsuperscript{Dof} and PIF4\textsuperscript{HLH} WT proteins. a and b, Size-exclusion chromatography and gel analysis of CDF2\textsuperscript{Dof} WT protein. c and d, Size-exclusion chromatography and gel analysis of PIF4\textsuperscript{HLH} WT protein. e and f, Size-exclusion chromatography and gel analysis of CDF2\textsuperscript{Dof} and PIF4\textsuperscript{HLH} WT proteins. The CDF2\textsuperscript{Dof} domain or PIF4\textsuperscript{HLH} was fused with an MBP tag at the N terminus. The x-axes and y-axes for a, c and e indicate the elution volume and the protein absorption at 280 nm, respectively. Size-exclusion chromatography analysis in b, d and f were performed twice with similar results.
 Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|-----|-----------|
| ☑   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☑   | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☑   | The statistical test(s) used AND whether they are one- or two-sided |
| ☑   | Only common tests should be described solely by name; describe more complex techniques in the Methods section. |
| ☑   | A description of all covariates tested |
| ☑   | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☑   | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☑   | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted |
| ☑   | Give P values as exact values whenever suitable. |
| ☑   | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☑   | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☑   | Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated |

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: MorphoGraphX;
AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/);
Protein Data Bank (PDB, https://www.rcsb.org/);
Pymol (https://pymol.org/2/)

Data analysis: Image J, MorphoGraphX and R.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw data for RNA-seq and ChIP-seq will be released on 08. 2022.
Raw data are for RNA-seq series RNA-seq series PRJNA747146 and ChIP-seq series PRJNA747820.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/re-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
- Analysis of Arabidopsis hypocotyl length. Statistical analysis was performed via Tukey’s least significant difference (LSD) test \((P \leq 0.05)\). \(n = 28\) hypocotyls.
- Analysis of epidermal cells in non-dividing cell files of Arabidopsis hypocotyls. Significant differences among genotypes were observed \((p < 0.05)\), using ANOVA followed by Tukey’s pairwise multiple comparison; \(n = 4\) hypocotyls.

**Data exclusions**
- No data were excluded during the analysis.

**Replication**
- At least three independent experiments were done to verify the reproducibility of the findings. And all the replicates were successful.

**Randomization**
- Seedlings were grown randomly in growth chamber and seedlings were harvested randomly for data collection.

**Blinding**
- The investigators were blinded to group allocation during data collection and/or analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Antibodies            |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology and archaeology |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |
| ☒   | Clinical data         |
| ☒   | Dual use research of concern |

### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChiP-seq              |
| ☒   | Flow cytometry        |
| ☒   | MRI-based neuroimaging |

### Antibodies

**Antibodies used**
- anti-HA antibody (ab9110, Abcam); anti-Myc antibody (22765, Cell Signaling Technology); anti-HA antibody (12013819001, Roche) or anti-Actin antibody (sc-47778, Santa Cruz); anti-Myc antibody (HRP-conjected, 20405, Cell Signaling Technology).

**Validation**
- Anti-HA (HRP) and anti-Myc (HRP-conjected, 20405, CST) antibodies were used at 1:5000-fold dilution for western blot for in vitro pull down experiments.
- Anti-HA (HRP) and anti-AC-T (HRP) antibodies were used at 1:2,000 and 1:4,000-fold dilutions for western blot for diurnal protein accumulation detection.

### ChiP-seq

**Data deposition**
- Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

**Data access links**

May remain private before publication.

Access for the referees is provided to all the raw peak files through the following links:

CDF2::HA-CDF2_cdf2_pif4_rep-1.narrowPeak http://84.22.105.30/natureplantshegao/CDF2::HA-CDF2_cdf2_pif4_rep-1.narrowPeak
**Files in database submission**

The files that will be in the database submission correspond to the MACS2 peak calling files. They are named according to the Chip sample:

- CDF2::HA-CDF2_cdf2_pif4_rep-1.narrowPeak
- CDF2::HA-CDF2_cdf2_pif4_rep-2.narrowPeak
- CDF2::HA-CDF2_cdf2_pif4_rep-3.narrowPeak
- CDF2::HA-CDF2_cdf2_pif4_rep-1.narrowPeak
- CDF2::HA-CDF2_cdf2_pif4_rep-2.narrowPeak
- CDF2::HA-CDF2_cdf2_pif4_rep-3.narrowPeak

**Methodology**

**Replicates**

- ChIP-seq experiments were done with three independent biological repeats.

| Sample          | TotalReads | UniquelyMapped | MaxReadLength | layout       |
|-----------------|------------|----------------|---------------|--------------|
| CDF2::HA-CDF2_cdf2_rep-1 | 25469658  | 18530900       | 151           | single       |
| CDF2::HA-CDF2_input_rep-1 | 22914897  | 16662023       | 151           | single       |
| CDF2::HA-CDF2_cdf2_rep-2 | 24122923  | 18232832       | 151           | single       |
| CDF2::HA-CDF2_input_rep-2 | 22315072  | 17087284       | 151           | single       |
| CDF2::HA-CDF2_cdf2_rep-3 | 24150078  | 17576376       | 151           | single       |
| CDF2::HA-CDF2_input_rep-3 | 21005583  | 15278099       | 151           | single       |
| CDF2::HA-CDF2_cdf2_pif4_rep-1 | 26571112  | 20705627       | 151           | single       |
| CDF2::HA-CDF2_input_pif4_rep-1 | 23928821  | 19346629       | 151           | single       |
| CDF2::HA-CDF2_cdf2_pif4_rep-2 | 27397745  | 21268854       | 151           | single       |
| CDF2::HA-CDF2_input_pif4_rep-2 | 27132603  | 21539254       | 151           | single       |
| CDF2::HA-CDF2_cdf2_pif4_rep-3 | 22604576  | 17600545       | 151           | single       |
| CDF2::HA-CDF2_input_pif4_rep-3 | 27377627  | 21983733       | 151           | single       |

**Sequencing depth**

| Sample          | TotalReads | UniquelyMapped | MaxReadLength | layout       |
|-----------------|------------|----------------|---------------|--------------|
| CDF2::HA-CDF2_cdf2_rep-1 | 25469658  | 18530900       | 151           | single       |
| CDF2::HA-CDF2_input_rep-1 | 22914897  | 16662023       | 151           | single       |
| CDF2::HA-CDF2_cdf2_rep-2 | 24122923  | 18232832       | 151           | single       |
| CDF2::HA-CDF2_input_rep-2 | 22315072  | 17087284       | 151           | single       |
| CDF2::HA-CDF2_cdf2_rep-3 | 24150078  | 17576376       | 151           | single       |
| CDF2::HA-CDF2_input_rep-3 | 21005583  | 15278099       | 151           | single       |
| CDF2::HA-CDF2_cdf2_pif4_rep-1 | 26571112  | 20705627       | 151           | single       |
| CDF2::HA-CDF2_input_pif4_rep-1 | 23928821  | 19346629       | 151           | single       |
| CDF2::HA-CDF2_cdf2_pif4_rep-2 | 27397745  | 21268854       | 151           | single       |
| CDF2::HA-CDF2_input_pif4_rep-2 | 27132603  | 21539254       | 151           | single       |
| CDF2::HA-CDF2_cdf2_pif4_rep-3 | 22604576  | 17600545       | 151           | single       |
| CDF2::HA-CDF2_input_pif4_rep-3 | 27377627  | 21983733       | 151           | single       |

**Antibodies**

- anti-HA antibody (ab9110, Abcam).

**Peak calling parameters**

- Bowtie2 Genome index for Arabidopsis thaliana was constructed using bowtie2-build [Athaliana.fasta] [genome] (rest of parameters was left to their default values). Reads were mapped using the following command: bowtie2 -p 4 -x genome -U [sample-fastaq-file] | samtools view -Shb > [sample.bam]. Resulting bam files were read-position-based sorted using samtools sort -o [sorted.bam] [sample.bam]. Finally reads with mapping quality lower than 30 were removed by executing: samtools view -q 30 -h -b [sorted.bam] > [final.bam], macs2 callpeak -t [ChIP.bam] -c [input.bam] -m 2 20 -q 0.01 -n [output] -g 12066 -B --SPMR. Finally diffbind was used and only merged peaks were considered that were present in all three replicates.

**Data quality**

As described in the Peak calling parameters. Stringent removal of alignments of mapping quality < 30 increased the probability that the reads were uniquely mapped. Next, due to the large number of initial peaks we used FDR cutoff of 0.01. Finally, further stringent filtering was performed by requiring a peak to be observed in all three replicates. The following summary was generated for the peak sets:

- Chip totalPeaks enrichedSplus
- CDF2::HA-CDF2_cdf2_rep-1 15102 2800
- CDF2::HA-CDF2_cdf2_rep-2 14361 2771
- CDF2::HA-CDF2_cdf2_rep-3 11542 1401
- CDF2::HA-CDF2_cdf2_pif4_rep-1 17306 284
- CDF2::HA-CDF2_cdf2_pif4_rep-2 13275 2165
- CDF2::HA-CDF2_cdf2_pif4_rep-3 10655 1927

**Software**

- Raw peaks were called for each ChIP-seq/input sample pair using the MACS2 software. Given that one of our main goals in the study was to determine differential binding, raw peaks were further processed using the diffbind R package.