Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

FLC genomic occupancy as a biological benchmark for plant ChIP-seq: a complete workflow from crosslinking to motif validation

Mohsen Hajheidari

m.hajheidari@uni-koeln.de

University of Cologne

Maryam Kheradmand Prouch

Ferdowsi University of Mashhad

Anna Ossowski

University of Cologne

Kerstin Becker

University of Cologne

Juliette Meaux

University of Cologne

Method Article

Keywords: ChIP-seq, Arabidopsis thaliana, FLOWERING LOCUS C, chromatin immunoprecipitation, transcription factor, RFP-Trap, MADS-box, motif enrichment

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11390817/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is widely used to identify genomic regions associated with a protein of interest, including transcription factors and other chromatin-associated proteins. However, the reliability of ChIP-seq experiments is questionable, because they often do not recover expected transcription factor-DNA binding sites. A robust workflow is essential for reliably detecting protein-associated genomic regions, particularly in plant tissues.

Results

Here we present a complete and robust plant ChIP-seq workflow from tissue harvest to annotated peaks, generating high signal-to-noise ratio, developed for approximately 2–3 g of 7-day-old *Arabidopsis thaliana* seedlings expressing a tagged protein of interest. The wet-lab procedure comprises formaldehyde crosslinking under vacuum infiltration, isolation of nuclei, detergent-assisted chromatin extraction, sonication, and affinity enrichment using RFP-Trap Magnetic Agarose beads. A particularly critical, yet often overlooked, parameter for achieving a reproducible high signal-to-noise ratio is the optimization of the balance between the amount of input chromatin containing the tagged protein and the volume of affinity beads used for immunoprecipitation. Then, affinity enrichment is followed by stringent sequential washing, reversal of crosslinks and column-based purification of ChIP DNA, with input chromatin and mock IP samples processed in parallel as controls. Purified DNA is converted into dual-indexed Illumina libraries using the TruSeq DNA Nano kit, quantified by qPCR and sequenced in paired-end mode (2×151 bp) on an Illumina NovaSeq X Plus. Reads are then quality-trimmed with fastp, aligned to the TAIR10 reference genome with Bowtie2, and peaks are called against matched input controls with MACS3 and annotated with ChIPseeker (https://github.com/mhajheidari/FLC-ChIP-seq-analysis). We benchmarked the workflow using the *Arabidopsis* MADS-box transcription factor FLOWERING LOCUS C (FLC), for which the workflow identified 1183 reproducible peaks called in all four biological replicates, narrowing to 957 peaks after mock IP-based specificity filtering. We identified peaks located at or near all previously reported FLC direct target genes, as well as at several additional developmentally important genes not previously reported as potential FLC targets, including the FLC locus itself. De novo motif discovery and central-enrichment analysis (CentriMo) confirmed strong, significant enrichment of the known FLC binding motif at the called peaks ($E = 3.1\times10^{-40}$), matching the published FLC position weight matrix (TOMTOM $E = 1.3\times10^{-23}$).

Conclusions

The protocol provides detailed buffer formulations, incubation conditions, computational parameters, and practical handling recommendations, and can be readily adapted to different plant lines expressing the tagged protein of interest. The FLC motif benchmark demonstrates the robustness of the workflow

and its ability to reliably capture biologically relevant and specific transcription factor-DNA interactions, addressing a key challenge in approaches for identifying protein-DNA interactions, including ChIP-seq. Overall, this workflow provides a practical and reproducible framework for conducting and validating ChIP-seq experiments in plants.

Background

Transcription factors (TFs) are central regulators of cellular identity and function. By recognizing specific DNA sequences and integrating developmental, environmental, and intracellular signals, TFs establish and maintain cell-type-specific gene-expression programs. Thus, identifying TF-DNA binding sites is essential for understanding transcriptional regulation and the molecular mechanisms underlying developmental and physiological processes. Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) enables genome-wide identification of genomic regions associated with a protein of interest *in vivo* and has become a widely used approach for characterizing transcription factor binding landscapes. However, ChIP-seq experiments do not necessarily recover all previously reported or biologically expected TF-DNA binding sites, and the number and strength of detected binding regions can vary substantially between experiments and biological systems [1–3]. This incomplete or variable recovery can result from both biological and technical factors, including the abundance and stability of the TF-DNA interaction, tissue- or developmental-stage-specific TF occupancy, chromatin accessibility, crosslinking efficiency, antibody affinity and specificity, chromatin fragmentation, immunoprecipitation efficiency, and the amount and quality of starting material. These considerations are particularly relevant to plant ChIP-seq experiments, where efficient tissue disruption, preservation of nuclei, chromatin extraction, and recovery of protein-associated chromatin can present additional technical challenges. Consequently, a robust and reproducible ChIP-seq workflow is essential for reliably detecting TF-associated genomic regions.

To demonstrate the performance and reproducibility of the ChIP-seq procedure, we selected the *Arabidopsis thaliana* MADS-box transcription factor FLOWERING LOCUS C (FLC) as a representative chromatin-associated protein. FLC is a well-characterized transcriptional repressor and central regulator of flowering time, with an extensively characterized genomic binding landscape and several established direct target genes [4–6], providing a biological reference for evaluating ChIP-seq performance.

Reported direct targets of FLC include genes involved in flowering and other developmental processes, such as FLOWERING LOCUS T (FT), SHORT VEGETATIVE PHASE (SVP), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 15 (SPL15), TARGET OF FLC AND SVP 1 (TFS1), TEMPRANILLO 1 (TEM1), SCHLAFMÜTZE (SMZ), TARGET OF EAT 3 (TOE3), APETALA1 (AP1), JASMONATE-ZIM-DOMAIN 6 (JAZ6), and JASMONATE-ZIM-DOMAIN 9 (JAZ9) [4–6]. We used these established targets, together with the published genomic FLC binding motif, as a biological benchmark for assessing the ability of the workflow to recover FLC-associated genomic regions and to specifically enrich for its known DNA-binding motif. Thus, FLC ChIP-seq provides

a functional benchmark that complements technical quality metrics and serves as a reference for evaluating the performance of this protocol in *Arabidopsis* seedlings.

The protocol described here uses formaldehyde-mediated crosslinking to stabilize protein-DNA interactions, followed by mechanical disruption and differential centrifugation to enrich nuclei. Chromatin is extracted using SDS-containing buffer and fragmented by sonication. Following dilution of the chromatin, an initial pre-clearing step is performed using Binding Control Magnetic Agarose Beads. The protein of interest is subsequently enriched using RFP-Trap Magnetic Agarose [7] when the target protein carries an mCherry or similar tag such as mRFP, mRFPruby, mPlum, tagRFP, mKate2, mOrange, PA-mCherry, mScarlet. Following immunoprecipitation, the beads are subjected to sequential low-salt, high-salt and LiCl washes before chromatin elution. Crosslinks are reversed by prolonged incubation at 65°C, followed by RNase A and proteinase K treatment and DNA purification using a silica membrane column. Purified ChIP, input and mock IP DNA is then converted into dual-indexed Illumina sequencing libraries and sequenced in paired-end mode. Finally, we provide a complete bioinformatic workflow, including motif-enrichment analysis, to quantitatively evaluate peak specificity and assess whether the identified genomic regions are enriched for the expected DNA-binding motif. Together, these experimental and computational steps provide an integrated framework for evaluating ChIP-seq performance and reproducibility. The complete analysis pipeline is available on GitHub (https://github.com/mhajheidari/FLC-ChIP-seq-analysis).

Methods

Plant material and growth conditions

Arabidopsis thaliana lines used in this study were in the Columbia (Col-0) background. The *flc-2 FRI-Sf2 FLCg-mCherry* line (genomic FLC-mCherry fusion expressed from the native FLC promoter) was kindly provided by Mike Howard (John Innes Centre, United Kingdom). Seeds were germinated on half-strength Murashige and Skoog (MS) medium supplemented with 0.5% sucrose. Seedlings were grown under a 16 h light/8 h dark photoperiod at a light intensity of 120 µmol m^{-2} s^{-1}, with temperatures maintained at 20°C during the light period and 18°C during the dark period. Approximately 2-3 g of 7-day-old seedlings was used as starting material for each sample. The amount of starting material and chromatin input may require optimization depending on the abundance of the target protein and the plant material.

Experimental design and controls

At minimum, the experimental design should include an input chromatin control for each biological sample, appropriate negative/control plant material where applicable, and biological replicates. The present protocol reserves input DNA from each chromatin preparation before immunoprecipitation; input samples are subsequently subjected to the same crosslink reversal and DNA purification procedures as the ChIP samples.

The reported ChIP-seq dataset for this study consists of eight biological replicates of FLC::FLCg-mCherry (n = 8; A1-A8), each with its own matched input control (Ai1-Ai8). Prior to immunoprecipitation, the protocol pools pairs of biological replicates according to DNA concentration to obtain approximately 30 µg DNA for one pooled immunoprecipitation sample [A1+A2, A3+A4, A5+A6 and A7+A8 (with inputs pooled correspondingly)], yielding four pooled ChIP samples (P1-P4) and four pooled input samples (Pi1-Pi4).

In addition to ChIP samples and their matched input controls, four starting biological mock IP samples were prepared as negative controls for the immunoprecipitation. Following the same pairwise-pooling design applied to the FLC::FLCg-mCherry and input samples, every two mock IP starting replicates were combined according to DNA concentration to obtain approximately 30 µg DNA per final immunoprecipitation, reducing the number of mock IP biological replicates from four to two. The mock IP samples (RFP-Trap pulldown performed on chromatin lacking the FLCg-mCherry transgene) were treated identically to the FLC::FLCg-mCherry samples throughout the entire procedure, from crosslinking through DNA purification, and were sequenced and used as a secondary specificity filter applied downstream of peak calling (see “Library preparation”, “Sequencing” and “Bioinformatic analysis” below).

The choice of mock IP control affects how closely nonspecific background is modeled. A fluorescent-protein line expressed under the native promoter of the protein of interest (here, the FLC promoter), most closely resembles the FLC::FLCg-mCherry samples in expression pattern, level and timing, and is therefore expected to provide the most stringent estimate of nonspecific RFP-Trap-associated signal. A fluorescent-protein line under a constitutive promoter (e.g. CaMV 35S) is a practical alternative when a native-promoter line is not available, although differences in expression level and pattern may make it a less precise match. Wild-type (untagged) plant material is also commonly used as a mock IP control in ChIP-seq studies [8], and controls for nonspecific antibody/bead binding, though it does not account for potential effects of the fluorescent-protein tag itself. Mock IP controls, unlike DNA input, also correct for nonspecific binding of the affinity reagent and bead matrix to chromatin, but typically yield less DNA and are consequently more susceptible to technical noise than input controls [9]; the two control types are therefore complementary rather than interchangeable, motivating their combined use in this protocol.

Reagents and materials

Reagents

· Paraformaldehyde (Merck, cat. No. P6148)

· KOH (Carl Roth, cat. No. 4330.1)

· Glycine (Carl Roth, cat. No. 0079.4)

· HEPES (Carl Roth, cat. No. 6763.1)

· DTT (Sigma-Aldrich, cat. No. D9779)

· PMSF (Thermo Fisher Scientific, cat. No. 36978)

· Plant protease inhibitor cocktail (Sigma Aldrich, cat. No. P9599)

· RNase A (Thermo Fisher Scientific, cat. No. EN0531)

· Proteinase K (New England Biolabs, cat. No. P8107S)

· Tris (Carl Roth, cat. No. 4855.5)

· EDTA (Merck, cat. No. 324503)

· $MgCl_2$ (Carl Roth, cat. No. 2189.1)

· NaCl (Carl Roth, cat. No. 3957.1)

· Sucrose (Carl Roth, cat. No. 9097.1)

· Glycerol (Thermo Fisher Scientific, cat. No. J61059.K2)

· Triton X-100 (Carl Roth, cat. No. 3051.3)

· SDS (Carl Roth, cat. No. CN30.2)

· IGEPAL CA-630 (Sigma-Aldrich, cat. No. I8896)

· Tween-20 (Sigma Aldrich, cat. No. 11332465001)

· Ethanol, absolute (Sigma-Aldrich, cat. No. 322221-1L-M)

· Sodium deoxycholate (Thermo Fisher Scientific, cat. No. 218591000)

· LiCl (Thermo Fisher Scientific, cat. No.AM9480)

Affinity reagents

· Binding Control Magnetic Agarose Beads (ChromoTek, cat. No. bmab,)

· RFP-Trap Magnetic Agarose (ChromoTek, cat. No. rtma)

DNA purification

· ChIP DNA Clean & Concentrator kit (Zymo Research, D5205)

Library preparation and sequencing

· Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. No. Q32851)

· TruSeq DNA Nano Library Prep Kit (Illumina, cat. No. 20015965)

· IDT for Illumina TruSeq DNA UD Indexes (Illumina, cat. No. 20020590)

· D1000 ScreenTape (Agilent, cat. No. 5067-5582) and D1000 Reagents (Agilent, cat. No. 5067-5583)

· Collibri Library Quantification Kit (Thermo Fisher Scientific, cat. No. A38524100)

· NovaSeq X Series 10B reagent kit 300 cycles (Illumina)

Consumables

· 50-mL centrifuge tubes

· 14-mL centrifuge tubes

· 2-mL microcentrifuge tubes

· 1.5-mL microcentrifuge tubes

· Sonication tubes compatible with the Bioruptor

· 0.45- or 0.22-µm syringe filters

· Magnetic rack

· Laboratory tissue

· 96-well PCR plates and adhesive plate seals

· Filter pipette tips

Equipment

· Nylon Net Filter/Miracloth 60 (Millipore, Ref: NY600900)

· Nylon Net Filter/Miracloth 20 (Millipore, Ref: NY200900)

· 0.22-µm PES (polyethersulfone) syringe filter (VWR)

· Thermomixer comfort 1.5 ml and 2 ml (Eppendorf)

· Glass vacuum desiccator 12.5"diam x 10"H, 14"H with lid (Glaswerk Wertheim)

· Vacuum pump (KNF LABOPORT, N 816.3)

· Refrigerated mini centrifuge 1.5ml/2ml (Sigma Aldrich, Hettich MIKRO 200R)

· Microcentrifuge 1.5ml/2ml (Eppendorf 5424)

· Refrigerated centrifuge 50 ml/15 ml (Thermo Fisher Scientific, Thermo Heraeus Multifuge 1S-R)

· Revolver rotator (neoLabLine, D-6049)

· Bioruptor Pico sonication system with Bioruptor cooling (Diagenode)

· Qubit fluorometer (Invitrogen, Q3.0)

· Thermal cycler with heated lid, 96-well (Applied Biosystems GeneAmp® PCR System 9700)

· Magnetic stand for 96-well plates (Dyna Mag ™ Side Skirted Life Technologies)

· Microplate centrifuge (Eppendorf, Centrifuge 5810R**)**

· TapeStation system (Agilent Technologies, 4200 TapeStation)

· QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific)

· NovaSeq X Plus sequencing system (Illumina)

Buffers and solutions

Crosslinking buffer

Prepare fresh 1% (w/v) formaldehyde in 1× PBS. For 250 mL, add 20 mL of 1× PBS to a small glass bottle, 2.6 g paraformaldehyde pellets, 98.5 µL of 1 N KOH or 27.7 µL of 20% KOH, stir at 65°C in a fume hood until the paraformaldehyde is completely dissolved (approximately 30 min), adjust the final volume to 250 mL with 1× PBS, filter through a 0.45- or 0.22-µm syringe filter; cool to 4°C before use.

Note: Prepare the crosslinking solution freshly and perform paraformaldehyde solution in a suitable chemical fume hood.

Nuclear Isolation Buffer [NIB; Buffer M1 (1x)]

50 mM HEPES-NaOH (pH 7.4), 5 mM $MgCl_2$, 25 mM NaCl, 2 mM EDTA, 2 mM DTT, 5% (0.146 M) sucrose, 30% glycerol and 0.25% Triton X-100. Immediately before use, add DTT (2 mM), 6µl/ml of 400mM PMSF

and plant protease inhibitor cocktail at 10 µL/ml.

Nuclear Wash Buffer [Buffer M2, (3x)]

50 mM HEPES-NaOH (pH 7.4), 20 mM $MgCl_2$, 100 mM NaCl, 40% sucrose, 40% glycerol and 0.25% Triton X-100. Immediately before use, add DTT (2 mM), 6µl/ml of 400mM PMSF, and plant protease inhibitor cocktail (10 µL/mL) to 1x buffer (The concentration of buffer M2 is three times higher than working concentration and needs to be diluted before use to prepare 1x buffer).

Nuclear Wash Buffer without Triton X-100 (Buffer M3)

Prepare Buffer M3 using the Nuclear Wash Buffer formulation above (1x buffer M2) but omit Triton X-100.

Dilution, wash, elution and phosphate buffers

· **Dilution Buffer I:** 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 150 mM NaCl

· **Dilution Buffer II:** 16.7 mM Tris-HCl pH 7.5, 1.2 mM EDTA, 167 mM NaCl, 1.1% Triton X-100 (filter sterilize)

· **Wash Buffer 1 (low-salt):** 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.5, 150 mM NaCl (filter sterilize)

· **Wash Buffer 2 (high-salt):** 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.5, 500 mM NaCl (filter sterilize)

· **Wash Buffer 3 (LiCl wash):** 0.25 M LiCl, 1% IGEPAL, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 7.5 (filter sterilize)

· **Wash Buffer 4:** 10 mM Tris-HCl pH 7.5, 0.5 mM EDTA

· **Resuspension buffer:** 10 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.5% SDS (filter sterilize)

· **Elution buffer (1% SDS buffer):** 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS (filter sterilize)

Phosphate buffer (10×)

10× concentrations: 1.35 M NaCl, 27 mM KCl, 80 mM Na_2HPO_4 and 15 mM KH_2PO_4.

Adjust to pH 7.4 with HCl. Dilute to 1× before use.

Procedure

Crosslinking of plant tissue

Timing: approximately 45-60 min, excluding tissue harvesting and freezing

1. Harvest 2-3 g of 7-day-old seedlings and place the tissue in a glass beaker containing 25 mL ice-cold 1× PBS without formaldehyde. Keep the glass beaker on ice.

2. Remove the PBS and add 50 mL ice-cold 1% formaldehyde crosslinking buffer. Ensure that the tissue is completely submerged.

3. Vacuum-infiltrate the tissue for 3 min. Slowly release the vacuum and ensure that the floating tissue becomes submerged and apply the vacuum for an additional 6 min using two successive 3-min periods.

4. Release the vacuum and incubate the tissue in crosslinking buffer for an additional 2 min.

5. Remove the crosslinking buffer and add 50 mL of 125 mM glycine prepared in 1× PBS. Mix the tissue and apply vacuum for 5 min to facilitate infiltration of the glycine solution.

6. Slowly release the vacuum and remove the glycine solution. Wash the tissue three times with cold water.

7. Remove excess liquid using laboratory tissue. Snap-freeze the tissue in liquid nitrogen. Store at −80°C or proceed immediately to nuclei isolation.

Note: Maintain the tissue at low temperature before and after crosslinking. Excessive crosslinking can reduce chromatin accessibility and immunoprecipitation efficiency; therefore, the specified infiltration and incubation times should be maintained consistently between samples.

Nuclei isolation

Timing: approximately 1.5-2 h

8. Grind the frozen plant tissue thoroughly in liquid nitrogen using mortar and pestle.

9. Transfer approximately 5 mL of tissue powder to a 50-mL centrifuge tube.

10. Add 35 mL NIB (Buffer M1). Homogenize using a rotator mixer for 10 min in a cold room.

11. Filter the homogenate through 60-µm and 20-µm nylon meshes into a new 50-mL tube.

12. Adjust the volume to 35 mL with NIB. Centrifuge at 1800 × g for 20 min at 4°C and carefully discard the supernatant.

13. Resuspend the pellet in 20 mL Nuclear Wash Buffer (M2) and centrifuge at 1800 × g for 10 min at 4°C. Repeat the resuspension and centrifugation step using M2 buffer and discard the supernatant.

14. Resuspend the pellet in 10 mL Nuclear Wash Buffer lacking Triton X-100 (M3). Centrifuge at 2000 × g for 20 min at 4°C.

Note: Resuspend nuclear pellets gently to minimize mechanical disruption of nuclei. Keep samples cold throughout nuclei isolation.

Chromatin extraction and sonication

Timing: approximately 2-3 h

15. Resuspend the nuclear pellet in 500 µL TE buffer containing 0.5% SDS (Resuspension buffer) and transfer the suspension to a 1.5-mL microcentrifuge tube. Incubate for 45 min at 4°C with gentle agitation using a rotator mixer (Revolver rotator).

16. Transfer the chromatin suspension into three sonication tubes, using approximately 250 µL per tube and a maximum of 300 µL per tube.

17. Sonicate using a Bioruptor with 11 cycles of 30 s ON and 30 s OFF at 4°C.

18. Pool the contents of the three sonication tubes corresponding to each sample into a 2-mL microcentrifuge tube.

19. Centrifuge at maximum speed for 15 min at 4°C and transfer the supernatant to a clean tube.

20. Snap-freeze the sonicated chromatin in liquid nitrogen and store at −80°C if it is not processed immediately (sonicated chromatin can be stored at −80°C for several months).

21. Reserve 10 µL of each chromatin preparation as input DNA control and store at −20°C. Keep 5 µL for assessment of DNA fragmentation by agarose gel electrophoresis.

Note: Sonication efficiency should be evaluated during protocol establishment and whenever the sonicator, tube type, sample volume or tissue amount is changed.

Expected result: Sonication should produce a reproducible distribution of 200 to 600 bp DNA fragments.

Pre-clearing of chromatin

Timing: approximately 1.5 h

22. Resuspend the Binding Control Magnetic Agarose Beads (bmab) and transfer 25 µL bead slurry to 500 µL ice-cold Dilution Buffer I.

23. Place the tube on a magnetic rack and allow the beads to collect until the supernatant becomes clear.

24. Remove and discard the supernatant. Repeat the wash two more times.

25. Dilute approximately 300 µL sonicated chromatin with 1550 µL Dilution Buffer II, corresponding to an approximately fivefold dilution. Add the diluted chromatin to the equilibrated bmab from step 24.

26. Incubate for 1 h at 4°C with rotation. Magnetically separate the beads until the supernatant is clear and then transfer the cleared chromatin to a clean tube. Reserve 50 µL of diluted lysate for immunoblot analysis and also 2 µL to measure DNA concentration with Qubit.

Note: Avoid transferring magnetic beads into the immunoprecipitation reaction.

Preparation of RFP-Trap Magnetic Agarose beads

Timing: approximately 10-15 min

27. Resuspend the RFP-Trap Magnetic Agarose thoroughly. Transfer 8.5 µL bead slurry to 500 µL ice-cold Dilution Buffer I.

28. Magnetically separate the beads until the supernatant is clear. Remove the supernatant and repeat the wash twice.

mCherry immunoprecipitation

Timing: 4 h to overnight

29. Normalize chromatin preparations according to DNA concentration; approximately 30 µg DNA is used per final biological replicate.

30. Combine two starting biological replicates according to their DNA concentration measured at step 26 to obtain approximately 30 µg DNA. Combine the chromatin samples in a 14-mL centrifuge tube. If the volume is below 6 mL, adjust the volume to 6 mL using Dilution Buffer II.

31. Add 1.5 mL of the 6 mL chromatin suspension to each of four 2 mL tubes containing 8.5 µL bead slurry of equilibrated RFP-Trap Magnetic Agarose from step 28.

32. Incubate by end-over-end tumbling for 4 h or overnight at 4°C. Magnetically separate the beads by allowing them to collect for approximately 1 min. Reserve 50 µL of supernatant for immunoblot analysis. Discard the remaining supernatant.

Note: To ensure reproducibility, the DNA normalization and pooling procedure must be identical across all biological replicates. Maintaining an appropriate balance between the amount of input chromatin (and therefore the abundance of the tagged protein) and the volume of magnetic agarose beads is critical for achieving a high signal-to-noise ratio. The optimal ratio depends largely on target protein abundance and the bead binding capacity (for example, RFP-Trap Magnetic Agarose beads bind approximately 22.5 µg of RFP per 25 µL of slurry). A practical approach to establish this balance is to perform a pilot depletion assay. Take a sample of the pre-cleared chromatin lysate to serve as starting baseline. Incubate this sample with the RFP-Trap magnetic agarose beads, perform magnetic separation, and collect the resulting post-IP supernatant (flow-through). Compare the amount of remaining tagged protein in the supernatant against the starting baseline to track depletion efficiency by running both fractions on an immunoblot. The lowest bead volume that achieves complete depletion of the target protein can then be selected. Combining this immunoblot assessment with the measured DNA concentration of the initial pre-cleared chromatin allows to calculate a standardized bead-to-chromatin ratio to apply consistently across all biological replicates.

Washing of immunoprecipitated chromatin

Timing: approximately 45-60 min

33. Add 1 mL Wash Buffer 1 to the beads and gently resuspend. Incubate for 5 min at 4°C with rotation. Magnetically separate the beads and discard the supernatant.

34. Add 500 µL Wash Buffer 1, resuspend, combine two tubes belonging to the same biological replicate and incubate for 5 min at 4°C with rotation; separate and discard the supernatant.

35. Add 500 µL Wash Buffer 2, resuspend, combine two tubes belonging to the same biological replicate and incubate for 5 min at 4°C with rotation (one tube remains per final biological replicate), magnetically separate the beads, and discard the supernatant. Repeat the Wash Buffer 2 with 1 ml buffer once more.

36. Add 1 ml Wash Buffer 3, incubate for 5 min at 4°C with rotation, magnetically separate the beads, and discard the supernatant. Repeat washing with Wash Buffer 3 once more.

37. Add 1 mL Wash Buffer 4, resuspend, magnetically separate the beads, and remove the supernatant completely.

Note: Avoid disturbing the bead pellet when removing wash solutions. Incomplete removal of wash buffers can dilute the elution buffer and affect downstream DNA purification.

Elution of immunoprecipitated chromatin

Timing: approximately 30 min

38. Add 250 µL Elution Buffer (1% SDS buffer) to each bead preparation. Incubate for 30 min at 65°C with shaking at 1300 rpm.

39. Place the tubes on a magnetic rack and allow the beads to collect completely and transfer the cleared eluate to a clean tube.

Reversal of crosslinks

Timing: 6 h to overnight

40. Add 10 µL of 5 M NaCl to each ChIP eluate. Incubate at 65°C for at least 6 h or overnight to reverse formaldehyde crosslinks.

41. Process the input DNA controls in parallel: adjust the volume to 250 µL with Elution Buffer, add 10 µL of 5 M NaCl, and incubate at 65°C for at least 6 h or overnight.

Note: Input samples should undergo the same crosslink-reversal conditions as the ChIP samples. The corresponding input samples of two combined samples in step 30 should have been combined in similar proportion like the samples in step 30 before starting the step 41.

RNase A and proteinase K treatment

Timing: approximately 2.5 h

42. Add 1 µL RNase A (10 mg/mL) to each sample and incubate at 37°C for 15 min.

43. Add 10 µL of 1 M Tris-HCl, pH 6.5 and 3 µL proteinase K (10 mg/mL) and incubate at 45°C for 2 h.

Purification of ChIP DNA

Timing: approximately 30-45 min

44. Add five volumes of ChIP DNA Binding Buffer from ChIP DNA Clean & Concentrator kit to each sample volume (e.g. 250 µL Binding Buffer to 50 µL sample) and mix briefly.

45. Transfer the mixture to a Zymo-Spin column placed in a collection tube.

46. Centrifuge at ≥ 10,000 × g for 30 s and discard the flow-through; repeat until the complete sample has passed through the column.

47. Add 200 µL Zymo DNA Wash Buffer to the column and centrifuge at ≥ 10,000 × g for 30 s. Repeat the wash once.

48. Transfer the column to a clean 1.5-mL microcentrifuge tube and add 15 µL Zymo kit Elution Buffer directly to the center of the column matrix and centrifuge at ≥ 10,000 × g for 30 s. Retain the eluate containing purified ChIP DNA.

DNA quality control

Measure the concentration of purified ChIP and input DNA using a Qubit fluorometer, and assess chromatin fragmentation using agarose gel electrophoresis or TapeStation platform.

Library preparation

Timing: approximately 4-5 h

49. Use up to 100 ng DNA in 60 µL Resuspension Buffer (RSB) as starting material for library preparation with the TruSeq DNA Nano Library Preparation Kit (Illumina).

50. Repair DNA ends: add 40 µL End Repair Mix to each sample, mix, and incubate on a thermal cycler at 30°C for 30 min.

51. Select the library insert size roughly 250-650 bp, centered around 300-450 bp: Add 160 µL diluted SPB to each sample to remove large DNA fragments, incubate at room temperature for 5 min, and transfer 250 µL of the cleared supernatant to a new plate on a magnetic stand. To remove small DNA fragments, add 90 µL undiluted SPB to the supernatant, incubate at room temperature for 5 min, wash the pelleted beads twice with 200 µL 80% ethanol, air-dry, and elute in 20 µL RSB.

Note: Samples can be paused at this stage; seal the plate and store at -25°C to -15°C for up to 7 days if not continuing immediately.

52. Adenylate 3' ends: add 12.5 µL A-Tailing Mix to each sample, mix, and incubate on a thermal cycler at 37°C for 30 min followed by 70°C for 5 min.

53. Ligate adapters: add Ligation Mix and a unique dual-index DNA adapter (IDT for Illumina TruSeq DNA UD Indexes) to each sample, mix, and incubate at 30°C for 10 min; stop the ligation reaction with Stop Ligation Buffer.

54. Purify the ligated libraries with two rounds of SPB bead clean-up, washing with 200 µL 80% ethanol and eluting in RSB.

Note: Samples can be paused at this stage; seal the plate and store at -25°C to -15°C for up to 7 days if not continuing immediately.

55. Amplify libraries by PCR using PCR Primer Cocktail and Enhanced PCR Mix for 12 cycles (98°C for 20 s, 60°C for 15 s, 72°C for 30 s per cycle, preceded by 95°C for 3 min and is followed by 72°C for 5 min); when no ChIP DNA is detectable by Qubit before library preparation, use 15 cycles instead. Check the unpurified library concentration on an Agilent TapeStation (D1000 ScreenTape) after the initial cycling, and run 3 additional cycles if the concentration is below 2 nM.

56. Purify the amplified libraries with SPB bead clean-up, washing with 200 µL 80% ethanol and eluting in 30-32.5 µL Elution Buffer or RSB. A second round of clean-up was performed for libraries with prominent adapter-dimer peaks.

Sequencing

Library size distribution was assessed on an Agilent TapeStation (D1000 ScreenTape), and libraries with similar fragment-size profiles were pooled to a final concentration of approximately 10 nM or 2-4 ng/µL, with Tween-20 added to a final concentration of 0.1%. Pooled libraries were quantified using the Collibri Library Quantification Kit (Thermo Fisher Scientific) on a QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific). Equimolar, pooled libraries were sequenced on an Illumina NovaSeq X Plus platform in paired-end mode (2×151 bp), aiming for 20 million clusters per sample.

Bioinformatic analysis

Adapter trimming and quality filtering were performed with fastp [10] prior to alignment. Reads were aligned to the *Arabidopsis thaliana* TAIR10 reference genome (Arabidopsis_thaliana.TAIR10.dna.toplevel.fa; annotation TAIR10.52 GTF; [11]) with Bowtie2 ([12]; paired-end, --very-sensitive preset, --no-unal --no-discordant --no-mixed); alignments were sorted and indexed with samtools [13], with samtools flagstat used for mapping QC. No separate PCR-duplicate removal step (e.g. Picard MarkDuplicates or samtools markdup) was applied; duplicate filtering was left to MACS3's default internal duplicate-filtering model (--keep-dup 1, i.e. at most one read retained per exact genomic position) during peak calling.

Peaks were called with MACS3 callpeak ([14]; format BAMPE, genome size 1.2×10^8, q-value cutoff 0.05), with each ChIP sample called against its matched input control. For ChIP samples, a peak was defined as reproducible if it was called ($q < 0.05$) in all 4 of 4 biological replicates (P1-P4 vs Pi1-Pi4). Reproducible regions were computed in R using the GenomicRanges and rtracklayer packages, in two steps. First, the peak calls from all four biological replicates were combined, and any peaks that overlapped one another were merged into a single, non-redundant set of candidate regions. Second, each candidate region was retained only if it independently overlapped a called peak in every one of the four replicates; the retained q-value for each region was the maximum (least significant) q-value among its overlapping per-replicate peaks. The resulting intersection set (n = 1183 peaks) was used for downstream annotation.

As a downstream check for nonspecific RFP-Trap binding not captured by the input-based background model above, peaks were called for the two sequenced mock IP replicates against a pooled Pi1-Pi4 input (samtools merge), using MACS3 with the same parameters as the primary peak calling [14]. Each of the 1183 peaks was flagged if it overlapped a called peak in at least one of the two mock IP replicates; 226 peaks (19.1%) were flagged and excluded, yielding a mock-filtered set of 957 peaks. This filtering step was applied after, and independently of, the input-based reproducibility criterion above, and did not alter the original ChIP reproducible peaks set.

The full set of 957 reproducible peaks was annotated with the ChIPseeker R package [15] against a TxDb built from the TAIR10.52 GTF, flanking gene information computed within 5000 bp (Supplementary Table 1). For motif enrichment analysis, the 957 peaks were resized to a fixed 200-bp window centered on each peak midpoint; genomic sequence was extracted from the TAIR10 reference with bedtools getfasta [16]. De novo motif discovery was performed with STREME [17] (widths 8-30 bp), run both in its default differential-enrichment mode and in central-distance mode (--objfun cd), and with MEME [18]. All tools scored sequences against an *Arabidopsis* promoter background (A = 0.3323, T = 0.3323, C = 0.1677, G = 0.1677) matched to the background used to build the reference FLC motif. Central positional enrichment of all discovered motifs, together with the published FLC binding motif (PlantTFDB AT5G10140/MP00079; [19]), was tested with CentriMo [20]. De novo motifs were compared with the published FLC motif using TOMTOM [21], and peaks were additionally scanned directly with the published FLC position weight matrix using FIMO [22] at a threshold of $p < 1\times10^{-3}$.

Signal-to-noise metrics: Fraction of reads in peaks (FRiP) was calculated for each ChIP, input and mock IP library as the percentage of properly paired, primary alignments that overlap the 957 reproducible, mock-IP-filtered peaks. Median fold enrichment over input was taken from the MACS3 narrowPeak output for peaks overlapping the 957-peak set. Strand cross-correlation, NSC, RSC and the ENCODE quality tag were computed for each ChIP replicate with phantompeakqualtools (run_spp.R, SPP v1.14) [2, 23].

Expected results

Following successful completion of the ChIP protocol, purified DNA should be obtained from immunoprecipitated, input samples and mock IPs, sufficiently pure and concentrated for downstream ChIP-seq library preparation. The most important intermediate quality-control point is chromatin fragmentation following sonication; consistent fragmentation across biological replicates is essential for reproducible ChIP-seq experiments. Successful RFP-based immunoprecipitation should result in enrichment of chromatin associated with the mCherry-tagged protein relative to appropriate controls. The optimal balance between the amount of input chromatin, and therefore of the tagged protein is essential for high signal-to-noise ratio. After library preparation, final libraries should show a single, broad fragment-size distribution on the TapeStation without a prominent adapter-dimer peak, and should reach a concentration of at least 2 nM for pooling.

Troubleshooting

Poor chromatin fragmentation or excessively large DNA fragments. These usually result from inefficient or insufficient sonication, for example because of incorrect instrument settings, an unsuitable sample volume or tube type, or overly concentrated chromatin. Verify the Bioruptor settings, sample volume, tube type and chromatin concentration, and reassess the fragment distribution. Increase the number of sonication cycles only stepwise and after establishing the effect on chromatin integrity.

Tagged protein remains in the post-IP supernatant. Incomplete depletion indicates that the beads are saturated relative to the amount of tagged protein, that the incubation was too short, or that less bead slurry was dispensed than intended.

Very low ChIP DNA yield. Low yield can result from low abundance of the tagged protein, incomplete capture by the beads, or insufficient input chromatin. Confirm expression of the mCherry fusion by microscopy or by immunoblot of the pre-cleared lysate. If more input chromatin is required, scale the bead volume proportionally so that the bead-to-chromatin ratio remains unchanged. Verify DNA quantification and bead handling.

High background or weak enrichment relative to input and mock IP controls. An excess of beads relative to the tagged protein increases the unoccupied binding surface and nonspecific capture of chromatin; insufficient washing or incomplete pre-clearing has a similar effect. A high proportion of ChIP peaks that are also called in the mock IP controls is a typical sign of this problem. Reduce the bead volume to the lowest volume that still achieves complete depletion of the tagged protein, verify bead equilibration and pre-clearing (steps 24–26) and the wash conditions, and ensure complete removal of wash buffers between washes.

Variable ChIP yield between replicates. Variation usually reflects unequal chromatin input, unequal chromatin fragmentation, an inconsistent bead-to-chromatin ratio, or differences in tagged-protein abundance between chromatin preparations. Quantify DNA in the pre-cleared chromatin (step 26), normalize and pool chromatin identically for all replicates (steps 29–30), and apply the same bead-to-chromatin ratio to every replicate. Compare tagged-protein levels between preparations by immunoblot of the pre-cleared lysates.

DNA degradation. Degradation is usually caused by excessive tissue handling or poor sample preservation. Keep tissue and chromatin cold and minimize processing time.

Poor downstream library preparation. This can result from insufficient DNA or an inappropriate fragment distribution. Quantify the DNA and assess the fragment size before library preparation.

Adapter dimers in the final library. Adapter dimers form when the DNA input is low relative to the amount of adapter. Perform an additional SPB bead clean-up (step 56) and re-check the library on the TapeStation before pooling.

Low final library concentration. This is usually caused by very low ChIP DNA input. Use 15 PCR cycles when ChIP DNA is not detectable by Qubit and add 3 cycles if the library is below 2 nM (step 55). Avoid unnecessary cycles, which increase PCR duplicates.

Safety considerations

Formaldehyde and paraformaldehyde are hazardous chemicals and should be handled in an appropriate chemical fume hood using suitable personal protective equipment and institutional waste-disposal procedures. Paraformaldehyde dissolution should not be performed on an open laboratory bench. Liquid nitrogen should be handled using appropriate cryogenic personal protective equipment. SDS, PMSF and other chemical reagents used in this protocol should be handled according to institutional safety procedures and the relevant safety data sheets.

Results and discussion

We evaluated the workflow using 7-day-old seedlings of the flc-2 FRI-Sf2 FLCg-mCherry line, in which a genomic FLC-mCherry fusion is expressed from the native FLC promoter. Rather than relying on a single quality metric, performance was assessed at four complementary levels: (i) reproducibility of peak calling across four biological replicates; (ii) specificity relative to both input chromatin and mock IP controls; (iii) recovery of previously reported FLC direct target genes; and (iv) enrichment and central positioning of the known FLC DNA-binding motif within the called peaks. Previously published plant ChIP methods have been validated mainly by ChIP-qPCR at a small number of reference loci [24, 25], by peak overlap between biological replicates [25], by signal-to-noise metrics such as the FRiP [26], or by comparing histone-mark distributions with gene expression [27]. Combining replicate-, control- and sequence-based criteria provides an assessment of ChIP-seq performance that does not depend on any single locus or metric.

Peak identification results

Applying the reproducibility criterion described above (a peak called in all 4 of 4 biological replicates) identified 1183 reproducible peaks. After removing peaks that also overlapped a peak called in the mock IP controls, a mock IP-filtered, specificity-validated set of 957 peaks remained and was used for downstream motif enrichment analysis (see below).

To further evaluate whether the workflow recapitulates the expected FLC genomic binding landscape, the mock IP-filtered and specificity-validated peak set (n = 957) was cross-referenced with the AGI locus identifiers of 11 previously reported FLC direct target genes [4–6]. Ten of the 11 target genes were associated with, or located near, a reproducible and specificity-validated FLC binding peak (Fig. 1). No reproducible peak was detected at or near SMZ (AT3G54990) when requiring reproducibility across all four biological replicates. However, relaxing the reproducibility criterion to peaks detected in at least

three of four biological replicates additionally recovered a peak 2.3 kb upstream of the SMZ transcription start site. Visual inspection of the aligned read coverage confirmed consistent, peak-like enrichment at this locus across all four ChIP biological replicates (Fig. 1). In one replicate, however, the signal fell just below the MACS3 significance threshold and was therefore not called as a significant peak, explaining its absence from the strict 4-of-4 reproducible peak set. In addition to these previously characterized targets, we identified reproducible FLC binding peaks near several genes that, to our knowledge, have not previously been reported as direct FLC targets. Notably, a reproducible FLC binding peak was detected at the FLC promoter itself, suggesting the possibility of negative autoregulatory feedback (Fig. 1). Together, the recovery of previously characterized FLC target loci and the motif enrichment observed below support the ability of the workflow to capture the expected genomic binding landscape of FLC.

Signal-to-noise ratio of the FLC ChIP-seq libraries

To assess signal-to-noise ratio, we used two complementary metrics recommended by the ENCODE consortium [2], the fraction of reads in peaks (FRiP), which depends on peak calling, and strand cross-correlation, which is computed directly from the read alignments.

In the four FLC ChIP replicates, 2.6–3.3% of properly paired, primary alignments overlapped the 957 reproducible, mock-IP-filtered FLC peaks, exceeding the ENCODE minimum of 1% for transcription-factor ChIP-seq [2] (Fig. 2A). The corresponding values were approximately 0.5% for the matched inputs and approximately 0.25% for the mock IP controls, so the ChIP samples contained roughly 5–7 times more reads in FLC peaks than the inputs. For the peaks called in each replicate that overlapped the 957-peak set, the median MACS3 fold enrichment over input was approximately 4.6–6.4. Because the 957 peaks were defined by excluding regions called in the mock IPs, the low mock-IP FRiP partly reflects this filtering, and the comparison with the inputs is the more conservative one. FRiP values of a few percent are typical for successful transcription-factor ChIP-seq, because transcription factors occupy only a small fraction of the genome [2].

Strand cross-correlation analysis, which also does not depend on peak calling, showed a distinct fragment-length peak at 245–255 bp in all four ChIP replicates (Fig. 2B-E), consistent with the fragment size distribution after sonication. The relative strand coefficient (RSC; 1.72–1.82) was well above the ENCODE threshold of 0.8, and all replicates received the highest ENCODE quality tag (Qtag = 2) [2]. The normalized strand coefficient (NSC; 1.049–1.061) was close to the ENCODE guideline value of 1.05. Low NSC values are common for transcription factors with a limited number of binding sites, because only a small fraction of reads contributes to the fragment-length peak [2]; the RSC, which compares this peak with the read-length (phantom) peak, is therefore the more informative measure in this case. Together, the FRiP values, fold-enrichment estimates, and cross-correlation profiles indicate that the workflow generated reproducible, high-quality ChIP-seq libraries with substantial enrichment of FLC-associated chromatin over both input and mock-IP controls.

Motif enrichment results, computational validation against the known FLC binding site

De novo motif discovery on 957 peaks independently recovered the same AT-rich core in both STREME objective functions (default differential-enrichment and central-distance modes) and in classic fixed-width MEME. This de novo motif matched the published FLC binding motif (PlantTFDB AT5G10140/MP00079; [19]) by TOMTOM comparison ($E = 1.3\times10^{-23}$; Fig. 3A) and closely resembles the previously reported FLC CArG-box-type binding site [4–6]. CentriMo confirmed strong, significant central enrichment (i.e. positional bias toward the peak midpoint, consistent with direct binding) for the de novo motif ($E = 3.1\times10^{-40}$), for the published FLC PWM itself ($E = 1.3\times10^{-38}$), and for the corresponding motifs recovered by MEME ($E = 3.2\times10^{-35}$) and STREME central-distance mode ($E = 7.1\times10^{-29}$; Fig. 3B). A direct FIMO scan of the peak set with the published FLC PWM ($p < 1\times10^{-3}$) returned 1642 significant matches, with matched sequences closely resembling the reference consensus (e.g. AATTTCCTAAAATAGAAAGAC, ACAAACCAAAAATGGAAAGAA).

Together, these results indicate that the ChIP-seq peaks called for samples harboring FLC-tagged protein, following both input-based peak calling and mock IP-based specificity filtering, are significantly and specifically enriched for the known FLC binding motif, providing computational validation of the immunoprecipitation and peak-calling steps of this protocol. This motif-based validation complements the robust wet-lab workflow described above and offers a generalizable strategy for identification of binding profile of any tagged chromatin-associated protein, including transcription factors.

Conclusions

We present a complete and benchmarked plant ChIP-seq workflow spanning tissue crosslinking, library preparation and sequencing through to peak annotation and motif-based validation, using the Arabidopsis MADS-box transcription factor FLC as a biological reference. The workflow combines input-based primary peak calling with a secondary specificity filter based on mock IP controls. This additional filtering step improved motif enrichment and is therefore recommended when mock IP material is available.

Unbiased de novo motif discovery independently recovered the previously characterized CArG-box-type FLC binding motif, which was significantly and specifically enriched at peak centres. This provides quantitative computational support for the specificity of the identified ChIP-seq peaks and complements the experimental validation provided by the wet-lab protocol.

The 957 FLC binding sites identified in this study are almost twice as numerous as the 505 sites reported in a previous antiserum-based study [4]. Nevertheless, the recovery of the same FLC binding motif, together with its strong enrichment at peak centres, supports the biological relevance of the additional sites identified here. These results suggest that the workflow may capture a broader set of

FLC binding sites and potentially additional FLC target genes than previously reported; these candidate sites and targets should be further evaluated experimentally.

Overall, this integrated wet-lab and computational workflow provides a robust framework for plant ChIP-seq analysis and can be readily adapted to other tagged chromatin-associated proteins in *Arabidopsis* and, in principle, to other plant species.

Abbreviations

ChIP-seq
Chromatin immunoprecipitation followed by sequencing
TF
Transcription factor
FLC
FLOWERING LOCUS C
RFP
Red fluorescent protein
PBS
Phosphate-buffered saline
EDTA
Ethylenediaminetetraacetic acid
SDS
Sodium dodecyl sulfate
PMSF
Phenylmethylsulfonyl fluoride
DTT
Dithiothreitol
TAIR10
The *Arabidopsis* Information Resource genome assembly, version 10
MACS3
Model-based Analysis of ChIP-Seq, version 3
PWM
Position weight matrix
QC
Quality control
qPCR
Quantitative PCR
RSB
Resuspension Buffer
SPB
Sample Purification Beads

UD
Unique dual (indexes)
FIMO
Find Individual Motif Occurrences
STREME
Sensitive, Thorough, Rapid, Enriched Motif Elicitation
MEME
Multiple Em for Motif Elicitation
CentriMo
Central Motif enrichment analysis
TOMTOM
Motif comparison tool (MEME Suite)

Declarations

Ethics approval and consent to participate

Not applicable. This study involved only *Arabidopsis thaliana* plant material and did not involve human participants, human data, or animals. AI was used for grammatical editing of the text.

Consent for publication

Not applicable.

Availability of data and materials

The data have been deposited in the European Nucleotide Archive under accession number PRJEB128133. The scripts used for the ChIP-seq data analysis have been deposited in GitHub and are publicly available at https://github.com/mhajheidari/FLC-ChIP-seq-analysis.

Competing interests

The authors declare that they have no competing interests.

Funding

We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding under Germany´s Excellence Strategy – EXC-2048/1 – project ID 390686111 as well as under the Research Consortium TRR341 “Plant Ecological Genetics”.

Authors' contributions

M.H. and J.d.M. designed the study. M.H. and M.K.P. performed the ChIP-seq experiments. A.O. and K.B. performed library preparation and sequencing. M.H. conducted the computational analyses. M.H. wrote the manuscript draft with the input from J.d.M.

Acknowledgements

The authors thank Mike Howard (John Innes Centre, United Kingdom) for kindly providing the flc-2 FRI-Sf2 FLCg-mCherry line.

References

1. Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316(5830):1497–502. https://doi.org/10.1126/science.1141319.
2. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012;22(9):1813–31. https://doi.org/10.1101/gr.136184.111.
3. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009;10(10):669–80. https://doi.org/10.1038/nrg2641.
4. Deng W, Ying H, Helliwell CA, et al. FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. Proc Natl Acad Sci U S A. 2011;108(16):6680–5. https://doi.org/10.1073/pnas.1103175108.
5. Mateos JL, Madrigal P, Tsuda K, et al. Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in *Arabidopsis*. Genome Biol. 2015;16:31. https://doi.org/10.1186/s13059-015-0597-1.
6. Richter R, Kinoshita A, Vincent C, et al. Floral regulators FLC and SOC1 directly regulate expression of the B3-type transcription factor TARGET OF FLC AND SVP 1 at the *Arabidopsis* shoot apex via antagonistic chromatin modifications. PLoS Genet. 2019;15(4):e1008065. https://doi.org/10.1371/journal.pgen.1008065.
7. Rothbauer U, Zolghadr K, Muyldermans S, et al. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. Mol Cell Proteom. 2008;7(2):282–9. https://doi.org/10.1074/mcp.M700342-MCP200.
8. Song L, Koga Y, Ecker JR. Profiling of transcription factor binding events by chromatin immunoprecipitation sequencing (ChIP-seq). Curr Protoc Plant Biol. 2016;1(2):293–306. https://doi.org/10.1002/cppb.20014.
9. Xu J, Kudron MM, Victorsen A, et al. To mock or not: a comprehensive comparison of mock IP and DNA input for ChIP-seq. Nucleic Acids Res. 2021;49(3):e17. https://doi.org/10.1093/nar/gkaa1155.

10. Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560.
11. Lamesch P, Berardini TZ, Li D, et al. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 2012;40(D1):D1202–10. https://doi.org/10.1093/nar/gkr1090.
12. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9. https://doi.org/10.1038/nmeth.1923.
13. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2):giab008. https://doi.org/10.1093/gigascience/giab008.
14. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.
15. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31(14):2382–3. https://doi.org/10.1093/bioinformatics/btv145.
16. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033.
17. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics. 2021;37(18):2834–40. https://doi.org/10.1093/bioinformatics/btab203.
18. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28–36.
19. Jin J, Tian F, Yang DC, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45(D1):D1040–5. https://doi.org/10.1093/nar/gkw982.
20. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res. 2012;40(17):e128. https://doi.org/10.1093/nar/gks433.
21. Gupta S, Stamatoyannopoulos JA, Bailey TL, et al. Quantifying similarity between motifs. Genome Biol. 2007;8(2):R24. https://doi.org/10.1186/gb-2007-8-2-r24.
22. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017–8. https://doi.org/10.1093/bioinformatics/btr064.
23. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol. 2008;26(12):1351–9. https://doi.org/10.1038/nbt.1508.
24. Lee JH, Jin S, Kim SY, et al. A fast, efficient chromatin immunoprecipitation method for studying protein-DNA binding in *Arabidopsis* mesophyll protoplasts. Plant Methods. 2017;13:42. https://doi.org/10.1186/s13007-017-0192-4.
25. Huang X, Pan Q, Lin Y, et al. A native chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in strawberry fruits. Plant Methods. 2020;16:10. https://doi.org/10.1186/s13007-020-0556-z.

26. Tao X, Feng S, Zhao T, et al. Efficient chromatin profiling of H3K4me3 modification in cotton using CUT&Tag. Plant Methods. 2020;16:120. https://doi.org/10.1186/s13007-020-00664-8.
27. Canton M, Farinati S, Forestan C, et al. An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in peach reproductive tissues. Plant Methods. 2022;18:43. https://doi.org/10.1186/s13007-022-00876-0.

Figures

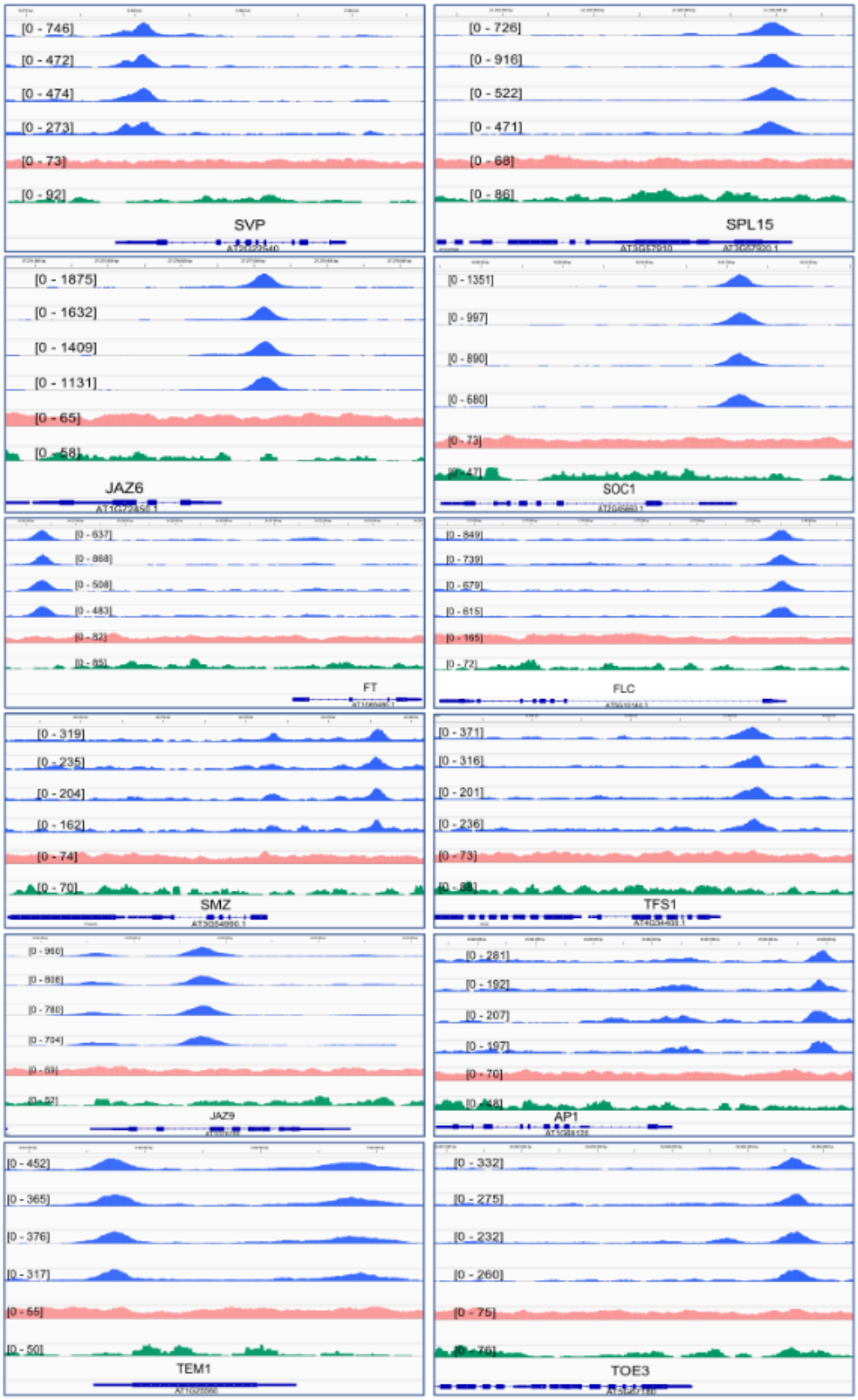


Figure 1

ChIP-seq coverage profiles at FLC direct target loci and at the FLC locus itself. Genome browser (IGV) views showing aligned read coverage across 12 genomic regions: SVP (AT2G22540), SPL15 (AT3G57920), JAZ6 (AT1G72450), SOC1 (AT2G45660), FT (AT1G65480), SMZ (AT3G54990), TFS1 (AT4G34400), JAZ9 (AT1G70700), AP1 (AT1G69120), TEM1 (AT1G25560), TOE3 (AT5G67180), and FLC (AT5G10140). In each panel, the top four tracks (blue) show ChIP-seq coverage for the four pooled

FLC::FLCg-mCherry biological replicates (P1-P4), the fifth track (pink) shows coverage for the Pi1 input control, and the sixth track (green) shows coverage for one of the mock IP controls as representative. Gene models (TAIR10 annotation) are shown below each panel, with genomic coordinates indicated at the top of each track group. Signal ranges (read count) are auto-scaled per track and indicated in brackets at the top left of each track. A clear, reproducible enrichment of ChIP-seq signal is observed across all four biological replicates at each locus, relative to the flat, background-level signal in the input and mock IP controls, consistent with direct FLC binding at these sites.

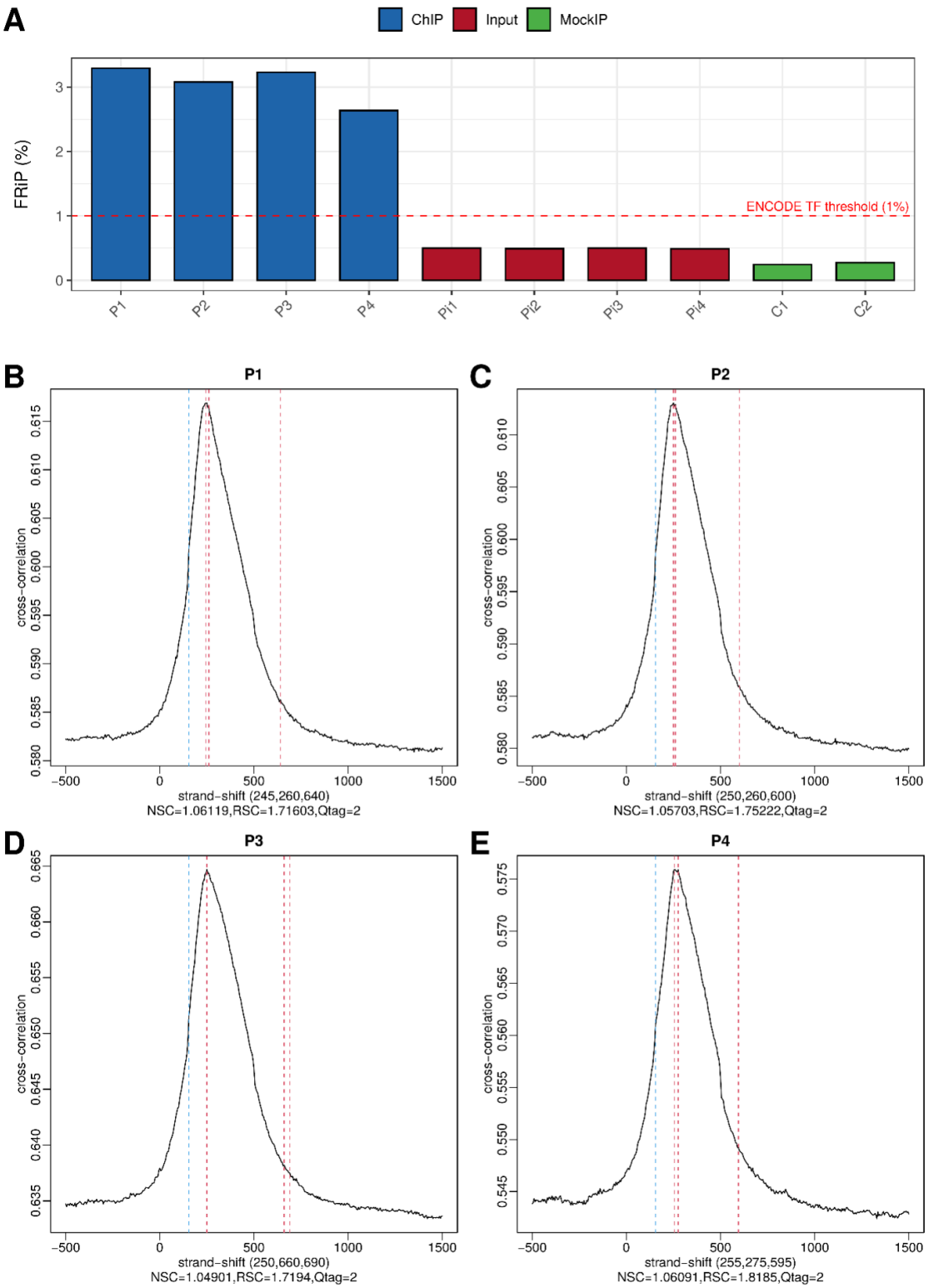


Figure 2

FLC ChIP-seq shows strong enrichment over input and mock IP controls. (A) Fraction of reads in peaks (FRiP) for the four FLC ChIP replicates (P1-P4, blue), their matched input samples (Pi1-Pi4, red) and the two mock IP controls (C1, C2, green). FRiP is the percentage of properly paired, primary alignments that overlap the 957 reproducible, mock-IP-filtered FLC peaks. The dashed red line marks the ENCODE-recommended minimum FRiP of 1% for transcription-factor ChIP-seq. (B-E) Strand cross-correlation profiles for ChIP replicates P1 (B), P2 (C), P3 (D) and P4 (E), generated with phantompeakqualtools (run_spp.R, SPP v1.14). Cross-correlation between forward- and reverse-strand read densities is plotted against strand shift (bp). Red dashed lines mark the three top-ranked candidate strand shifts, which are listed in brackets below each plot; the first value is the estimated fragment length. The dashed line near 151 bp marks the read length (phantom peak). The normalized strand coefficient (NSC), relative strand coefficient (RSC) and ENCODE quality tag (Qtag; scale −2 to +2) are given below each panel.

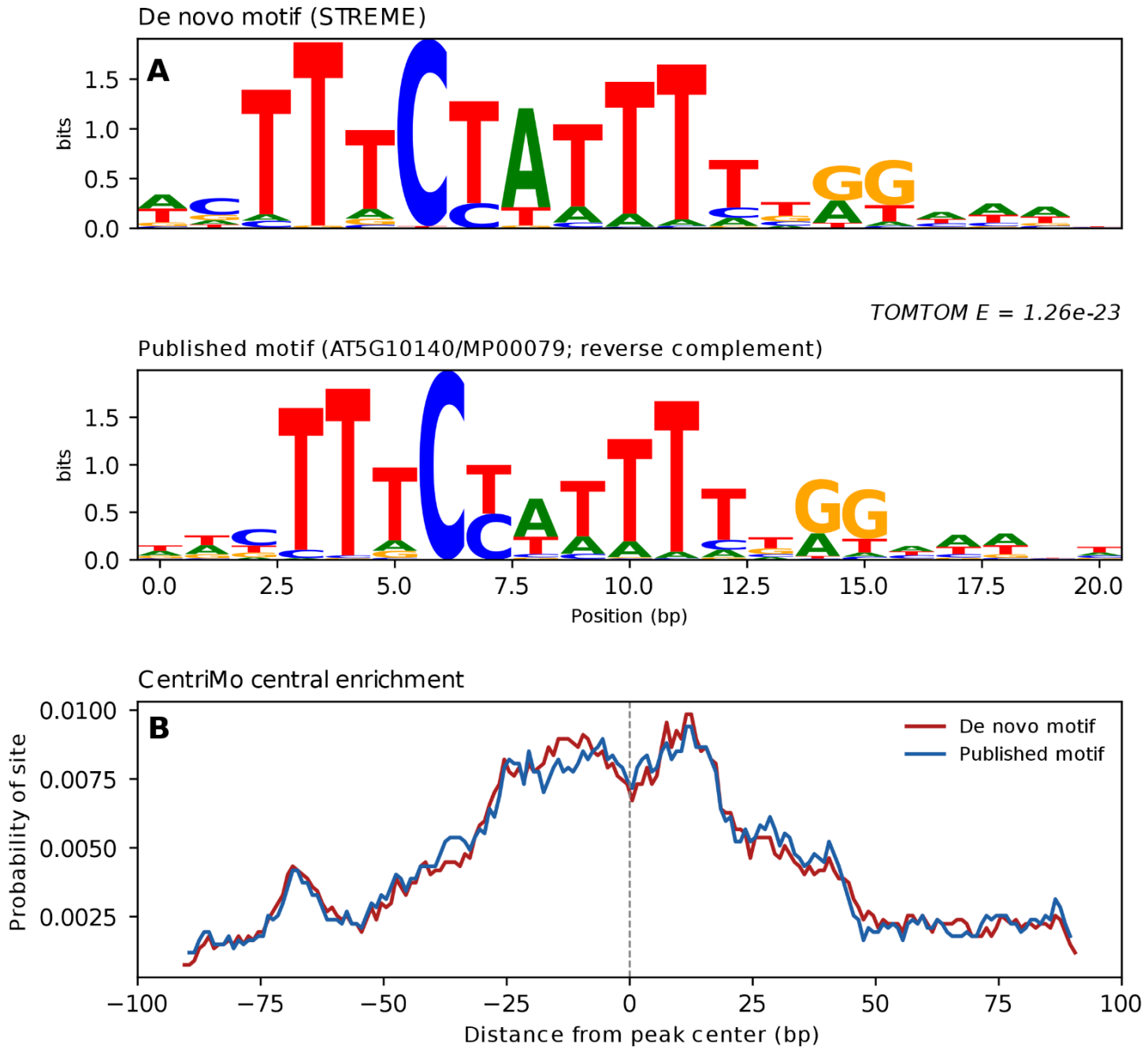


Figure 3

Computational validation of the ChIP-seq peaks against the known FLC binding motif. (A) Sequence logo of the de novo motif recovered by STREME from the 957 mock-filtered ChIP samples peaks (top), aligned against the published FLC binding motif (PlantTFDB AT5G10140/MP00079; shown as its reverse complement to match the de novo motif's orientation), with the TOMTOM match E-value. (B) CentriMo central-enrichment profiles for the de novo motif and the published FLC motif, showing the probability of a motif match as a function of distance from the ChIP-seq peak center.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.csv