



Identification of Plant Transcription Factor DNA-Binding Sites Using seq-DAP-seq

Stephanie Hutin, Romain Blanc-Mathieu, Philippe Rieu, François Parcy,
Xuelei Lai, Chloe Zubieta

► To cite this version:

Stephanie Hutin, Romain Blanc-Mathieu, Philippe Rieu, François Parcy, Xuelei Lai, et al.. Identification of Plant Transcription Factor DNA-Binding Sites Using seq-DAP-seq. Kerstin Kaufmann; Bernd Mueller-Roeber. Plant Gene Regulatory Networks Methods and Protocols. Series: Methods in Molecular Biology, 2698, Humana Press/Springer, pp.119-145, 2023, 978-1-4939-7124-4. 10.1007/978-1-0716-3354-0_9 . hal-04225070

HAL Id: hal-04225070

<https://hal.science/hal-04225070>

Submitted on 2 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Identification of plant transcription factor DNA-binding sites using seq-DAP-seq

Stephanie Hutin¹, Romain Blanc-Mathieu¹, Philippe Rieu¹, François Parcy¹, Xuelei Lai^{1,2} and
Chloe Zubieta¹

¹Univ. Grenoble Alpes, CNRS, CEA, INRAE, Laboratoire Physiologie Cellulaire et Végétale,
IRIG-DBSCI-LPCV, 17 Avenue des Martyrs, 38054, Grenoble, France

² Current address: National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan
Laboratory, Huazhong Agricultural University, Wuhan 430070, China

Running head: Seq-DAP-seq protocol

Abstract

The identification of genome-wide transcription factor (TF) binding sites is a critical step in deciphering gene and transcriptional regulatory networks. However, determining the genome-wide binding of specific TFs or TF complexes remains a technical challenge. DNA affinity purification sequencing (DAP-seq) and modifications such as sequential DAP-seq (seq-DAP-seq) are robust *in vitro* methods for mapping individual TF or TF complex binding sites in a genome-wide manner. DAP-seq protocols use a genomic DNA (gDNA) library from any target organism with or without amplification, allowing the determination of TF binding on naked or endogenously modified DNA, respectively. As a first step, the gDNA is fragmented to ~200bp, end-repaired and sequencing adaptors are added. This gDNA library can be used directly or an amplification step may be performed to remove endogenous DNA modifications such as cytosine methylation. DNA libraries are then incubated with an affinity-tagged TF or TF-complex immobilized on magnetic beads. The TF or TF complex of interest is generally produced using recombinant protein expression and purified prior to DNA affinity purification. After incubation of the DNA library with the immobilized TF of interest, multiple wash steps are performed to reduce non-specific DNA binding and the TF-DNA complexes eluted. The eluted DNA is PCR-amplified and sequenced using next-generation sequencing. The resulting sequence reads are mapped to the corresponding reference genome, identifying direct potential bound regions and binding sites of the TF or TF complex of interest. Predictive TF binding site (TFBS) models for the TF or TF complex of interest are generated from the bound regions using downstream bioinformatics analysis pipelines. Here, we present a detailed protocol outlining the steps required for seq-DAP-seq of a heterooligomeric TF complex (Figure 1) and briefly describe the downstream bioinformatics pipeline used to develop a robust TFBS model from sequencing data generated from a DAP-seq experiment. [Figure 1 near here]

Key Words (5-10)

DNA affinity purification sequencing (DAP-seq), sequential DNA affinity purification sequencing (seq-DAP-seq), transcription factors, transcription factor binding sites, DNA methylation, recombinant protein expression, sequential pull-down

1. Introduction

Proper gene regulation requires the finely orchestrated activity of thousands of transcription factors (TF) and TF complexes. Multicellular organisms dedicate 6-8% of their protein coding genes to the production of TFs and understanding the DNA-binding syntax of these factors is a fundamental challenge in biology *(1)*. *In vivo* techniques such as chromatin immunoprecipitation followed by polymerase chain reaction (ChIP-PCR) or deep sequencing ChIP-seq have been commonly used to identify TFBSs *in vivo*, allowing the interrogation of specific potential TFBS or the identification of TFBS genome-wide, respectively *(2, 3)*. New *in situ* techniques such as Cleavage Under Targets and Release Using Nuclease (CUT&RUN) *(4–6)* and Cleavage Under Targets and Tagmentation (CUT&Tag) *(7)* have further refined the identification of *in vivo* binding by investigating TF binding in the native chromatin environment without a cross-linking and DNA shearing step. CUT&RUN uses a protein A/G tethered micrococcal nuclease to release bound antibody-TF-DNA complexes, avoiding disruption of the chromatin. CUT&Tag uses a similar strategy but with a protein A/G-Tn5 transposase preloaded with sequencing adaptors, allowing for release of antibody-TF-DNA complexes and adaptor ligation in a single step. These techniques have certain drawbacks including the need for antibodies specific to the TF of interest or the use of a transgenic line expressing a tagged version of the TF of interest, potential poor signal to noise ratios due to low endogenous TF expression levels and artefacts due to target TF overexpression if a transgene is used. In addition, as TFs often form multiprotein complexes including with other TFs,

chromatin remodelers and scaffold proteins, identification of direct DNA binding and modeling of the TFBS of an individual TF or specific TF complex remains a challenge.

In vitro techniques such as protein binding microarrays (PBM) and systematic evolution of ligands by exponential enrichment (SELEX) and SELEX-seq use specific TFs or TFs complexes and either immobilized DNA or a random DNA library to identify TFBS (8, 9). In a PBM experiment, a TF or TF complex is expressed with an epitope tag such as FLAG-tag peptide (DYKDDDDK), Maltose Binding Protein (MBP) or Glutathione S-transferase (GST) for purification and identification of DNA binding events. The TF of interest is applied to a microarrays of immobilized dsDNA from a specific genome or random sequences, the microarray is washed to remove nonspecifically bound TFs and then a primary antibody is used to detect and quantify TF-DNA binding events (9). PBMs allow the modelling of TFBS of a specific TF or TF complex and the identification of putative binding sites in the genome but suffer from the use of short stretches of sequence of 10-12 base pairs that may miss longer TF recognition sites and require the use of a TFBS model to predict genomic binding sites. SELEX and SELEX-seq also use purified TF or TF complexes and query a random synthetic DNA library (8, 10–12). Successive rounds of DNA enrichment and amplification are performed allowing the generation of TFBS models or the identification of a single or limited number of high affinity binding motifs as a function of the number of rounds of selection performed. The TFBS models generated from SELEX experiments are not derived from genomic sequences and no direct mapping to the genome can be performed.

DAP-seq and seq-DAP-seq use purified TF or TF complexes and fragmented genomic DNA (gDNA) to overcome some of the drawbacks of the *in vivo* and *in vitro* techniques described above (13–17). Recombinant expression of the epitope-tagged TF or TF complex is performed

using a cell or cell-free expression system. One or more purification steps of the target TF or TF complex using multiple affinity tags may be required to ensure a single homogeneous species to be assayed. The purified TF complex of interest is immobilized and used to pull down specific binding sequences from a gDNA library pre-ligated with adaptors, the bound fragments are eluted, PCR amplified with barcode sequences and analyzed using next-generation sequencing, allowing for genome-wide binding-site mapping reflective of both local sequence context and DNA methylation status (13, 14). Alternatively, the gDNA library may be amplified prior to the DNA-binding step to remove any modifications such as methylation which may be present and may impact TF binding or site recognition (18). DAP-seq benefits from the use of purified specific TFs or TF complexes and the simplified *in vitro* system eliminates the complexity of different complexes containing the TF of interest that may form *in vivo* and alter DNA-binding specificity. In addition, by examining fragmented genomic DNA, the noise generated from non-specific binding is spread over the entire genome, thus eliminating the need of selection and amplification rounds required by SELEX-seq, for example. While DAP-seq and ampDAP-seq protocols have been well-described (13–15), we focus particular attention on the production and sequential pulldown steps required for working with heteromeric TF complexes (seq-DAP-seq) and provide the initial steps of an easy-to-follow data analysis pipeline.

Here we present a detailed schematic protocol designed for the seq-DAP-seq novice. The protocol includes (1) the preparation of a fragmented gDNA library from plant genomic DNA, ligation to Y-adaptors and optional PCR library amplification; (2) the production and purification of a heteromeric TF complex using multiple epitope tags and sequential pull downs; (3) DNA binding to the TF complex and preparation for next generation sequencing and (4) a

basic bioinformatics data analysis pipeline and on-line resources for processing DAP-seq sequencing data.

2. Materials

2.1 Equipment

1. Temperature controlled Covaris S220 ultrasonicator or similar
2. Qubit Fluorimeter or similar
3. Microvolume spectrophotometer
4. PCR machine/ thermocycler
5. Quantitative qPCR machine
6. Gel electrophoresis system
7. Blotting system
8. Refrigerated Microcentrifuge
9. Magnetic rack for 1.5 ml reaction tubes
10. Rotating wheel
11. Water bath at 37 °C
12. -20 °C freezer
13. -80 °C freezer
14. Ice bucket
15. Microtube 130 AFA (Covaris S220) or similar
16. Sterile 1.5 ml reaction tubes (Eppendorf or similar)
17. Sterile thin-walled PCR tubes

2.2 Kits

1. DNeasy Pro Kit (Qiagen)
2. End-It™ DNA repair kit (Lucigen) or similar
3. AMPure XP beads (Beckman Coulter) or similar
4. Miniprep kit (Qiagen) or similar
5. PCR clean-up kit (Qiagen) or similar
6. Qubit dsDNA HS quantification kit (ThermoFisher)
7. NEBNext Library Quant Kit for Illumina or similar
8. Gel Extraction Kit (Qiagen) or similar
9. TnT® SP6 High-Yield Wheat Germ Protein Expression System (Promega) or similar

2.3 Reagents and materials

1. Taq polymerase (NEB Taq DNA polymerase or similar)
2. Taq polymerase buffer (NEB ThermoPol or similar)
3. 25 mM MgCl₂ (in ddH₂O)
4. 1 mM dATP (stock solution in ddH₂O)
5. ddH₂O
6. T4 DNA ligase
7. 20 µg/µL Glycogen
8. 7.5 M NH₄CH₃CO₂ (ammonium acetate)
9. Phenol:chloroform:isoamyl alcohol (25:24:1)
10. 70% ethanol
11. 100% (vol/vol) ethanol
12. 1 M NaCH₃CO₂ (sodium acetate), pH 5.2
13. 100 mM dNTPs

14. Phusion High-Fidelity DNA Polymerase or similar
15. Agarose, molecular biology grade
16. SYBR Safe or similar
17. 100-bp DNA ladder
18. qPCR Master Mix (PCR PowerSYBER Green ThermoFisher or similar)
19. pTnT vectors or expression vectors or similar (user determined)
20. E. coli DH5 α or similar
21. SDS gel
22. Whatman paper
23. Whatman OPTITRAN BA-S 85 REINFORCED NC 0.45 μ m membrane
24. Milk powder
25. Tween 20
26. Plastic wrap
27. ECL Plus Western Blotting Detection system (Amersham)
28. 3xFLAG peptide (Merck Millipore)
29. anti-c-Myc magnetic beads (Thermo Scientific)
30. anti-FLAG magnetic beads (Merck Millipore)
31. anti-c-Myc (ThermoFisher)
32. anti-FLAG antibodies (Sigma-Aldrich)

2.4 Buffers

1. IP buffer: phosphate-buffered saline supplemented with 0.005% NP40 and protease inhibitors (Roche) and 1 mM TCEP
2. EB: 10 mM Tris-Cl, pH 8.5
3. 10x T4 DNA ligase buffer

4. 10x Taq polymerase buffer
5. 10x Phusion buffer
6. 1x SDS running buffer: 25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3
7. Laemmli loading dye: 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH approx. 6.8
8. Blotting buffer: 25 mM Tris pH 8, 190 mM glycine, 20% methanol
9. 1x PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, and 1.8 mM KH₂PO₄
10. Western Blot Stripping buffer: 2% SDS, 100 mM beta-mercaptoethanol, 50 mM Tris, pH 6.8

2.5 Primers

1. Y-shaped adaptor strand A 5'-CACGACGCTCTTCCGATCT-3' (stock at 60 μ M in EB)
2. Y-shaped adaptor strand B P- GATCGGAAGAGCACACGTCTG-3' (P represents a 5' phosphate group) (stock at 60 μ M in EB)
3. TruSeq universal primer (Illumina) (stock at 100 μ M in EB)
5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
GCTCTCCGATCT-3'
4. Indexed primer kit (NEBnext Indexed primers, or similar) (stock at 100 μ M in EB)
5. Indexed primer: All indexed primers follow the following sequence structure:
5'CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3', in which **XXXXXX** is the unique barcode sequence

5.1 Index

AAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

1

| | |
|--|----|
| 5.2 Index | 2 |
| CAAGCAGAAGACGGCATAACGAGAT ACATCGGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.3 Index | 3 |
| CAAGCAGAAGACGGCATAACGAGAT GCCTAAGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.4 Index | 4 |
| CAAGCAGAAGACGGCATAACGAGAT TGGTCAGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.5 Index | 5 |
| CAAGCAGAAGACGGCATAACGAGAT CACTGTGT GACTGGAGTTCA | |
| GACGTGTGCTCTTCCGATC | |
| 5.6 Index | 6 |
| CAAGCAGAAGACGGCATAACGAGAT ATTGGCGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.7 Index | 7 |
| CAAGCAGAAGACGGCATAACGAGAT GATCTGGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.8 Index | 8 |
| CAAGCAGAAGACGGCATAACGAGAT TCAAGTGT GACTGGAGTTCA | |
| GACGTGTGCTCTTCCGATC | |
| 5.9 Index | 9 |
| CAAGCAGAAGACGGCATAACGAGAT CTGATCGT GACTGGAGTTCA | |
| GACGTGTGCTCTTCCGATC | |
| 5.10 Index | 10 |
| CAAGCAGAAGACGGCATAACGAGATAAGCTAGTGACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.11 Index | 11 |
| CAAGCAGAAGACGGCATAACGAGAT GTAGCCGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |
| 5.12 Index | 12 |
| CAAGCAGAAGACGGCATAACGAGAT TACAAGGT GACTGGAGTTC | |
| AGACGTGTGCTCTTCCGATC | |

2.6 Bioinformatics requirements

1. Basic knowledge of Bash language is needed to run the pipeline.
2. Dependencies are: FastQC (19), NGmerge (20), Bowtie 2 (21), Samtools (22) Bedtools (23), MACS2 (24), MSPC, MEME Suite (25), Python and R scripts available on the GitHub page provided (https://github.com/Bioinfo-LPCV-RDF/DAPseq_TFanalysis) and optionally SRA Toolkit if DAP-seq sequence reads are downloaded from the NCBI.

3. Methods

3.1 DAP-seq and ampDAP-seq input library preparation

The same genomic DNA library can be used for several individual DAP-seq and seq-DAP-seq experiments, increasing assay throughput and limiting library-to-library bias. Illumina adaptor barcoding sequences may be chosen for multiplexing and are added at the final PCR amplification step. For ampDAP, libraries with different barcodes can be prepared in parallel but should not be combined until final multiplexing. This DNA extraction and library preparation protocol is for gDNA from *Arabidopsis thaliana* Col-0 but genomic DNA from the target plant of choice may be used. The libraries may be prepared and stored at -20°C or -80°C for later use.

1. Start the DNA library preparation with and without endogenous DNA modifications by extracting *Arabidopsis* Col-0 gDNA using a kit following the manufacture's manual or using standard CTAB protocols (26). Scale-up as required for the number of individually barcoded libraries that are desired. Each library preparation requires 5 µg of high quality gDNA. Perform library quantification using a Qubit kit. **Note 1**
2. Dilute 5 µg gDNA to a final volume of 125µl in EB into a microTUBE-130 AFA.
3. Fragment the DNA (Figure 2). Use the target fraction size of 200-300 bp on a temperature controlled ultrasonicator with the program target 200 bp. (I.e. peak power 175.0, Duty Factor 10.0%, Cycle/Burst 200, duration 180 s, sample temp 4-8 °C for DNA fragmentation.)

4. Transfer the sample to a clean sterile 1.5 ml tube and performed ethanol precipitation of the DNA.
5. Start an ethanol precipitation by adding 2.5 μ L glycogen (20 μ g/ μ L), 62.5 μ l (0.5x the sample volume) of 7.5 M NH_4OAc and top up the sample with 475 μ l (2.5x sample + ammonium acetate volume) 100% ethanol. Vortex to mix the sample. **Note 2**
6. Place the tubes at $-20\text{ }^{\circ}\text{C}$ overnight to precipitate the DNA from the sample or at $-80\text{ }^{\circ}\text{C}$ for at least 1 hour.

The protocol can be safely interrupted at this point

7. Centrifuge the sample in a tabletop centrifuge at $4\text{ }^{\circ}\text{C}$ for 30 min at $16,000 \times g$ to pellet the DNA.
8. Carefully remove the supernatant by pipetting without disturbing the DNA pellet. Carefully inspect the tube for the DNA pellet which is usually visible as a clear/white pellet and take care not to touch the pellet with the pipette tip.
9. Add 1 mL of 70% ethanol to wash the DNA pellet.
10. Centrifuge the sample for 2 min at $16,000 \times g$ at $4\text{ }^{\circ}\text{C}$. Carefully discard the supernatant by pipetting in order to remove as much of the remaining ethanol as possible.
11. Air dry the DNA at room temperature by opening the tube on the benchtop for 5–10 min. Allowing the pellet to dry for too long may make it more difficult to resuspend.
12. Resuspend the DNA pellets in 34 μ L of EB buffer by pipetting up and down 40x.
13. If the pellet is still visible, place the sample at $37\text{ }^{\circ}\text{C}$ in a water bath for 5 min to help resuspend the DNA and pipette up and down 40x.
14. End-repair the fragmented DNA (Figure 2). Transfer 34 μ l of the DNA to a clean PCR tube and add 16 μ l of End-It Repair Kit Master Mix (5 μ L $10\times$ End-It buffer, 5 μ L 2.5

mM dNTP (each) mix, 5 μ L 10 mM ATP and 1 μ L End-It enzyme mix (T4 polymerase., T4 kinase and Klenow fragment)) and mix by gently tapping the side of the tube. **Note 3 and Note 4**

15. Centrifuge at room temperature for 5 s at 3,000x g to collect any sample adhering to the sides of the tube.
16. Incubate at room temperature for 45 min and stop the reaction by heating to 70 °C for 10 min in a thermocycler. Transfer the sample to a clean 1.5 ml tube.
17. Perform an ethanol precipitation by adding 1 μ L glycogen (20 μ g/ μ L), 25 μ L (0.5 x the sample volume) of 7.5 M NH₄OAc and top up the sample with 189 μ L 100% ethanol. Vortex to mix the sample.
18. Repeat steps 6-11.
19. Resuspend the DNA pellet in 32 μ L of EB buffer by pipetting up and down 40x.
20. If the pellet is still visible, place the sample at 37 °C in a water bath for 5 min to help resuspend the DNA and pipette up and down 10x.

[Figure 3 near here]

21. Transfer 32 μ L sample to undergo an A-tailing reaction (Figure 3) into a clean PCR tube and add 5 μ L 10x ThermoPol Buffer, 10 μ L 1 mM ATP, 1 μ L Taq DNA polymerase and mix gently.
22. Centrifuge at room temperature for 5 s at 3,000g to collect any sample adhering to the walls of the tube.
23. Incubate at 72 °C for 30 min. in a thermocycler.
24. Transfer to a clean 1.5 ml reaction tube.
25. Repeat steps 17-18.
26. Resuspend DNA in 35 μ L of EB. Place the sample at 37 °C in a water bath for 5 min to aid resuspension.

27. Measure the DNA concentration using 2 μ L DNA and a Qubit with dsDNA Qubit Assay kit.

[Figure 4 near here]

28. To prepare the Y-shape adaptor, mix equimolar quantities of Y-shaped adaptor A and adaptor strand B (Figure 4) at 30 μ M concentration for each in a 1.5 ml reaction tube in EB.

29. Heat in a thermocycler or heating block to 95°C for 15 min.

30. Heat water to 95°C in a 500 ml beaker.

31. Place the tube with the adaptors to be ligated in a foam floating tube rack and place in beaker.

32. Let cool on benchtop until RT. **Note 5**

33. Transfer 30 μ L sample (**step 26**) to a clean PCR tube to be ligated to the Y-shape adaptor.

34. Add 10 μ L Ligation Master Mix (5 μ L T4 DNA Ligase buffer and 5 μ L 10x T4 DNA ligase buffer) and 10 μ L annealed Y-adaptors (**step 31**) in EB and mix gently by pipetting up and down 5 times.

35. Centrifuge at room temperature for 5 s at 3,000x g to collect any sample adhering to the sides of tube.

36. Incubate at room temperature for 3 h.

37. Perform Ethanol precipitation (steps 17-18).

38. Resuspend the DNA in 30 μ L of EB. Place the sample at 37 °C for 5 min to aid resuspension.

39. Measure the DNA concentration of 2 μ L of samples of the Y-adaptor ligated sample using a Qubit kit and compare to the concentration before the ligation (**step 27**). This is the DAP-seq input library.

40. Test 15 ng of the library before and after adaptor ligation with qPCRs. In a PCR tube, mix 15 ng of pre-ligated DNA (step 26) or DAP-seq input library with 1 μ L TruSeq Illumina universal primer (25 μ M), 1 μ L of any index primer (25 μ M) and top up to 10 μ L with nuclease free water. Add 10 μ L of 2x Syber Green Master Mix.
41. Load the samples into a real time qPCR machine and run the following program: 98 °C for 2 min, 98 °C for 30 s followed by 98 °C for 15 s 60 °C for 30 s 72 °C for 2 min (30 cycles), 72 °C for 10 min. Compare non-ligated and ligated samples using Absolute Quantification option to verify successful ligation.

Adaptor ligated gDNA may be used for a DAP-seq experiment at this point (DAP-seq input library) if step 39 is successful. **Note 6** and **Note 7**

3.1.1 ampDAP libraries without DNA modifications (Figure 5)- [Figure 5 near here]

1. Take 15 ng of gDNA sample after adaptor ligation (step 37).
2. Perform PCR amplification using TruSeq Illumina universal primer and the index primer with distinct barcode sequence (example set of 12 unique index primers are given in materials list). **Note 8**
3. In a PCR tube, mix 15 ng of template DNA with 10 μ L (5 \times) Phusion HF Buffer, 2.5 μ L dNTPs, 1 μ L TruSeq Illumina Universal primer (25 μ M stock), 1 μ L index primer with distinct index (25 μ M stock), 1 μ L and Phusion DNA Polymerase (2,000 U/ml) and top up to 50 μ L with ddH₂O. Perform for each indexed adaptor library desired.
4. Load the samples into a thermocycler and run the following program: 95 °C for 2 min, (1 cycle); 98 °C for 30 s, 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 2 min (12 cycles), 72 °C for 10 min.

5. Perform PCR clean-up using the Qiagen PCR clean-up kit according to the manufacturer's protocol.
6. Quantify library using Qubit Fluorimeter with dsDNA HS Qubit Assay kit according to the manufacturer's protocol.
7. Aliquot into 50 ng samples and freeze at -20 °C or -80 °C for long term storage. **Note 9**

The protocol can be safely interrupted at this point

3.2 TF expression, TF complex formation and sequential pull down –

1. Clone each gene encoding the TF of interest into a pTnT vector with appropriate epitope tag (*i.e.* 3xFLAG or 5xMyc) to obtain the plasmids of interest.
2. Plasmid DNA is amplified in *E. coli* DH5 α and purified using a Qiagen miniprep kit or similar according to the manufacturer's protocol (**Note 10**). At least 2 μ g of purified plasmid DNA is required per TnT reaction at a suitable concentration (~200 ng/ μ l). Equimolar amounts of plasmid for each TF to be expressed should be used in a single reaction mix. DNA concentration may be measured on a microvolume spectrometer and purity determined by the A260/280 ratio (ratio should be > 1.8).
3. TnT expression systems such as the TnT SP6 High-yield Wheat Germ Protein Expression System from Promega is used to express the targets proteins according to the manufacturer's protocol.
4. Place 30 μ l of TnT mix in a PCR tube. For single input reactions, 2-4 μ g of total plasmid is used in a 50 μ l final reaction volume. For double input reactions, 2 μ g plasmid of each construct is used in a 50 μ l reaction volume, top up to 50 μ l with ddH₂O. **Note 11**

5. Reactions are incubated for 2 h at 25 °C in a thermocycler to maintain a constant temperature.
6. Reserve 2 µl of reaction mix for western blots to confirm protein expression with the appropriate anti-FLAG or anti-Myc antibody. Commercially available antibodies anti-c-Myc and anti-FLAG antibodies at a dilution of 5000 x are sufficient to determine expression of the full-length protein.
 1. Prepare an SDS gel or use a commercially available one. The percentage of acrylamide should be chosen according to the size of the protein.
 2. Place the gel in an adequate running equipment in 1x SDS running buffer.
 3. Load the gel with your sample and a ladder.
 4. Run a SDS-gel at 180 V until the proteins are well separated. An indication are the dyes in the loading dye.
 5. Cut 4 pieces of Whatman paper slightly bigger than your gel (e.g. 9x9 cm) and a membrane in the size of the gel (8x8 cm).
 6. Wet all components with blotting buffer.
 7. Build the blot: In the cassette place on the – side a sponge tissue, then two Whatman papers followed by the membrane followed by the gel, then two Whatman paper. Pay attention, once the gel touches the membrane it should not me moved. Try to avoid air bubbles by wetting the components in between the stacks.
 8. Roll over the stack with a 10 mL pipet to expel air bubbles.
 9. Place the second tissue (sponge) and close the cassette.
 10. Place the cassette in the running chamber in the correct orientations – to – and + to +).
 11. Transfer over night to a membrane at 10 mA in a cold room.

12. Disassemble the blot. Handle the membrane with forceps.
13. Transfer the membrane into a blot box and add 5 % Milk powder/1xPBS/0.1% Tween 20 for 1h.
14. Dilute the primary antibody (1:200, 1:500, 1:1000) in 1 mL 5% milk powder/1x PBS/0.1% Tween 20.
15. Place the membrane onto a clean surface (e.g., clean bench top or glass surface).
16. Distribute the 1 mL of primary antibody solution onto the membrane taking care to fully cover the membrane and seal with a plastic wrap.
17. Incubate 1h at room temperature.
18. Place the membrane into a blot box and remove the excess antibody by washing the membrane 5 times, 5 min in 1x PBS/0.1% Tween 20.
19. Dilute the secondary antibody (1:10000) in 10 mL 5% milk powder/1x PBS/0.1% Tween 20 and incubate 1h. This solution can be reused and stored at 4 °C or -20 °C for long term storage.
20. Wash the membrane 5 times, 5 min in 1x PBS/0.1% Tween 20.
21. Incubate the membrane with ECL Plus Western Blotting Detection reagent for 1 min.
22. Expose the membrane in an imager or using ECL-sensitive film.
23. Heat the stripping buffer to 50°C in a water bath.
24. Incubate the membrane with the stripping buffer at 50°C for 15 min two times.
25. Wash the membrane with 1x PBS/0.1% Tween 20 several times for 5 min until the stripping buffer is completely removed .
26. Block the membrane and reuse to check for the second epitope tag.
27. Repeat steps 19-22 with the second antibody.

For recombinant protein expression in bacteria see **Note 12 and 13**.

7. Prepare the magnetic beads by resuspending them by inverting the storage tube 5x.
8. Pipette 100 μ l of magnetic beads (anti-FLAG or anti-Myc magnetic beads) in storage buffer into to a clean 1.5 ml reaction tube and add 200 μ l of IP buffer. **Note 14**
9. Place the tube on a magnetic rack and pipette off the buffer once the solution has cleared and all the beads have been drawn to the magnet.
10. Remove the tube from the magnetic rack and repeat the wash step 3x wash with 200 μ l of IP buffer by gently pipetting up and down 5x.
11. Add 100 μ l of IP buffer and transfer the magnetic beads in a clean tube and store on ice.

[Figure 6 near here]

12. In a clean 1.5 ml reaction tube, 50 μ l (Step 3.1) of the TnT reaction is combined with 130 μ l cold IP buffer and 20 μ l anti-FLAG magnetic beads in order to start the first pull down (Figure 6). **Note 15 and 16**
13. Gently mix by pipetting up and down 3x.
14. Incubate 1 h at 4 °C on a rotating wheel or by gently tapping the tube every 5 min to resuspend the beads.
15. Immobilize the anti-FLAG magnetic beads by placing the reaction tubes on a pre-chilled magnetic rack and carefully remove supernatant.
16. Wash three times with 100 μ l IP buffer, gently pipetting up and own each time to resuspend the beads.
17. Remove and discard supernatant.
18. Elute the FLAG-tagged proteins or protein complexes by addition of 175 μ l of IP buffer supplemented with 200 μ g/ml 3xFLAG peptide.

19. Immobilize the beads on the prechilled magnetic rack and carefully collect the supernatant.
20. Place the supernatant containing the eluted TF complexes in a clean 1.5 ml reaction tube and add 20 μ l of washed anti-c-Myc magnetic beads (step 11) (Figure 7). **[Figure 7 near here]**
21. Incubate 1 h at 4°C on a rotating wheel or by gently tapping the tube every 5 min to resuspend the beads.
22. Immobilize the beads on a magnetic stand and carefully pipette off the supernatant and discard it.
23. Wash the beads 3x with 100 μ l of IP buffer and resuspend in 50 μ l volume of IP buffer.

The purified TF or TF-associated complex, while still bound on magnetic beads are used for seq-DAP-seq.

3.3 Binding of DNA to immobilized TF complex- **[Figure 8 near here]**

1. Resuspend the beads in 100 μ l of IP buffer or optimized binding buffer.
2. Add 50 ng DAP-seq input library pre-ligated with Illumina Y-adaptor (DAP input library step 38) or ampDAP input library (step 47) (Figure 8).
3. Incubate at 4 °C for 90 min on a rotating wheel.
4. Place the samples on a prechilled magnetic rack and remove the supernatant by pipetting.
5. Wash the beads six times with 100 μ L IP buffer or optimized binding buffer by gently pipetting up and down 2x while keeping the tube on ice.

6. Transfer the buffer and beads to a new PCR tube and wash 2x with 100 μ l IP buffer or optimized binding buffer by pipetting gently up and down 2x.
7. Place the tube on the prechilled magnetic rack. Remove the supernatant and resuspend the beads in 30 μ l of EB and transfer to a PCR tube.
8. Heat the PCR tube in a thermocycler at 98 °C for 10 min.
9. Resuspend beads and transfer to a 1.5 ml reaction tube.
10. Place the tube on a magnetic stand and collect the supernatant (Figure 8). **Note 17 and 18**

The protocol can be safely interrupted at this point samples may be frozen at -20 and stored for later PCR amplification and sequencing.

3.4 DNA amplification, pooling and sequencing

1. In a PCR tube, mix 25 μ L of the supernatant with 10 μ L (5x) Phusion HF Buffer, 2.5 μ L dNTPs, 1 μ L TruSeq Illumina Universal primer (25 μ M stock in EB), 1 μ L index primer with distinct index (25 μ M stock in EB), 1 μ L and Phusion DNA Polymerase (2,000 U/ml), and top up to 50 μ L with ddH₂O.
2. Load the samples into a thermocycler and run the following program: 98 °C for 2 min, then 98 °C for 30 s, 98 °C for 15 s 60 °C for 30 s 72 °C for 2 min (20 cycles), 72 °C for 10 min.
3. Purify the PCR products using AMPure XP beads. An optional qPCR verification step can be performed if the TF complex of interest has known binding sites. **Note 19** and Figure 9

[Figure 9 near here]

4. Quantify DNA concentration using a Qubit fluorimeter and dsDNA HS assay kit. DNA concentration should be > 10 ng/ul.
5. Measure the molarity of the libraries using the NEBNext Library Quant Kit for Illumina following the manufacturer's instructions. **Note 20 and 21**
6. Combine 5 µl of each library with different barcodes in equal molarity in a 1.5ml tube. The rest of the reactions can be stored at -20 °C. Library concentration should be >10 ng/µl.
7. Quantify DNA concentration using a Qubit fluorimeter and dsDNA HS assay kit.
8. Run the sample on an Illumina sequencer such as MiSeq Instrument according to the manufacturer's specifications. Multiplexing of equimolar DNA libraries may be performed at this point.
9. Use paired-end sequencing of 150 cycles for 10-20 million reads per lane.

3.5 Bioinformatic analysis (Figure 10)- [Figure 10 near here]

Processing of DAP-seq sequence reads to identify bound regions and determine TFs or TF complex binding sites can be decomposed in five main steps (Figure 10). A bioinformatics pipeline encapsulated in a Conda environment is provided at https://github.com/Bioinfo-LPCV-RDF/DAPseq_TFanalysis. The pipeline provides an example DAP-seq sequencing dataset for BIM2, an *Arabidopsis thaliana* TF, with sequencing data from an ampDAP-seq experiment with paired end reads with two replicates and an input ampDAP DNA library as control. An example of a successful and unsuccessful seq-DAP-seq experiment are shown in Figure 11. Explanations and main commands for each step are described below. **Note 22**

[Figure 11 near here]

1. Check the reads quality by using FastQC (19). Note 23. Adapters are removed using NGMerge (20) with the `-a` option and reads are trimmed using Trimmomatic (27). Commands are given below to run the bioinformatics pipeline with example or user provided data. TF_1 and TF_2 refer to forward and reverse reads, respectively.

```
fastqc -o out_dir TF_1.fastq.gz TF_2.fastq.gz
```

```
NGMerge -a -1 TF_1.fastq.gz -2 TF_2.fastq.gz -o out_dir
```

```
trimmomatic PE TF_1_noAdpaters.fastq.gz -2 TF_2_noAdpaters.fastq.gz  
TF_1_paired.fq.gz TF_1_unpaired.fq.gz TF_2_paired.fq.gz TF_2_unpaired.fq.gz  
LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36
```

2. Map the reads to the reference genome sequence by using command *bowtie2* of the Bowtie 2 package, for all replicates and the control independently(21). Prior to mapping, the reference genome is first indexed with the command *bowtie2-build*. Resulting alignment files are in SAM format that are parsed to discard poorly mapped reads (more than three mismatches, mapping quality below 30 and suboptimal alignments) and to remove PCR duplicates before being converted to BAM file using Samtools commands. Example command line input is given below.

```
bowtie2-build genome.fasta genome_index
```

```
bowtie2 -x bowtie genome_index -1 TF_1_paired.fq -2 TF_2_paired.fq -S out.sam --  
dovetail
```

```
samtools view -Sh out.sam | grep -e "^@" -e 'XM:i:[012][^0-9]' | awk ' $1~/@/ || $5>30  
{print $0}' | grep -v "XS:i:" > out.filtered.sam
```

3. To prepare for peak calling, the BAM files are converted to BED files using *bedtools bamtoBed* command (23, 28, 29).

4. Call the peaks for each replicate separately by comparing their alignment file with that of the control input DNA using MACS2 (24). Resulting peaks files are then compared for consistency using MSPC (30) to keep ones that are observed in all (-c option set to 100%) or a majority (-c set to 50%) of replicates.

```
macs2 callpeak -t replicate.bed -c control.bed -B -f BEDPE -n out_dir -g genome_length
--call-summits -q 0.00001
```

```
mssp_mk -i (peak_rep1.narrowPeak peak_rep2.narrowPeak) -r Tec -w 1e-4 -s 1e-8 -c
100% -o outdir -d 1
```

Note 24

5. Analyze the bound regions to identify an enriched DNA sequence motif that likely corresponds to a transcription factor binding sites (TFBS). Several tools exist, the most popular being *meme-chip* from the MEME-suite package (25). **Note 25 and 26.**

```
meme-chip -oc out_dir -nmeme 600 -meme-maxsize 600000 -meme-minw 4 -meme-
maxw 8 -meme-nmotifs 3 -dreme-m 0 -noecho training_peaks.fas
```

6. Once a motif has been established, evaluate its ability to discriminate between bound regions (positive data) and unbound genomic regions or randomly shuffled positive sequences (negative test set). The letter-probability matrix (that gives the frequency of each of the four nucleotides at each position in the motif) generated by *meme-chip* in step 4 (located in meme/meme_out/meme_mini.txt file) can be used (providing some format conversion) by several tools to search DNA sequence for this motif. In our pipeline, script *meme2pfm.sh* operates this conversion.
7. Run the *score.py* script to scan the positive and negative set fasta sequences.
8. Run the *ROC.R* script to compute a ROC curve and AUC to determine the quality of the model. **Note 27.**

4. Notes

1. For large scale genomic DNA extraction, cost effective methods such as CTAB-based extraction (31), phenol:chloroform:isoamyl alcohol (25:24:1) or non-organic solvent extraction may be preferred (26, 32). We do not recommend using a microvolume spectrometer for gDNA library quantification as the measured concentration may be higher due to contaminating molecules that absorb at A260. If using a spectrophotometer, the A260/A280 ratio should be > 1.8 for a high quality gDNA library.
2. All ethanol precipitation steps can be replace by using Qiagen MinElute Reaction Cleanup kit. This minimizes the time for DNA purification/concentration to 10 mins. We have good results with kits for DNA purification/concentration steps.
3. DNA end repair and 3'dA tailing may be performed as one step using kits such as NEB End Repair/dA-Tailing Module. We have separated the steps to illustrate each step of the protocol.
4. End-It™ commercial kits have all the required components for creating blunt ended DNA premixed and reduce pipetting errors and we recommend using these. The concentrations of the various component are given in parentheses.
5. Steps 27-31 may be performed prior to or during End repair and A-tailing reactions to save time.
6. DNA quantification using a Qubit fluorimeter and dsDNA HS assay kit gives more accurate quantification than a microspectrophotometer.

7. If PCR amplification of the adaptor ligated library requires more than 15 cycles to reach saturation, this may indicate a problem with the library and we recommend using a new gDNA extraction. DNA loss during wash steps and residual ethanol are likely causes of poor qPCR results.

8. Care should be taken to avoid contamination of libraries or accidental pooling after bar code addition. A detailed example of single barcoding with one index primer and the universal primer is shown in Figure 12. **[Figure 12 near here]**

9. We have stored DAP and ampDAP libraries at -20°C for up to 1 year without an effect on downstream steps.

10. Miniprep DNA extractions may be pooled to obtain sufficient plasmid DNA for cell free protein expression. Alternatively, phenol:chloroform:isoamyl alcohol (25:24:1) DNA extractions may be performed. The quantity and quality of plasmid DNA obtained from phenol chloroform extraction is generally better than that obtained from commercial miniprep kits.

11. We have successfully produced up to four protein complexes in a single reaction mix. Performing the protein expressions in a single TnT reaction may be important for the generation of stable heteromeric TF complexes. In our hands, expression of proteins from multiple plasmids in a single reaction volume results in better protein yields and active TF complexes when compared to separate reactions and subsequent mixing of the protein partners. Stoichiometry of input plasmid should reflect final complex stoichiometry. However, in certain cases, plasmid concentrations may need to be optimized to produce active complexes,

particularly if the proteins have different stabilities. These proteins can be produced in excess to ensure active complex formation. We recommend testing by EMSA with different plasmid concentrations.

12. Standard expression and purification of tagged proteins from *E. coli* or other expression systems may be used to obtain TF or TF complexes of interest. Affinity purification using tags including 6x-His, GST or MBP may be used for purification from cultures of 10 ml. For DAP-seq applications 1 µg of purified protein is sufficient and is generally attainable from a 10 ml culture, although this may vary depending on the TF target. While protein tags (GST or MBP) and His tags are often used for recombinant protein purification on inexpensive resins, we recommend including a 3x FLAG tags or 5x Myc tags in the construct for protein immobilization and DNA binding experiments due to the high affinity binding of 5x Myc and 3x FLAG tagged proteins to anti-FLAG and anti-Myc magnetic beads. We do not recommend the use of His tags and Ni-NTA magnetic beads due to nonspecific interactions with the resin and leaching/ protein loss during the DNA binding and washing procedure.

13. Care should be taken with protein fusion tags. If the DNA-binding domain is N-terminal the tag should be located on the C-terminus and vice versa. Large tags may interfere with DNA-binding. Bulky protein tags such as MBP and GST may cause steric hindrance of the DNA-binding domain, resulting in poor DNA binding efficiency and care should be taken with tag location and/or inclusion of a spacer sequence between the tag and TF of interest.

14. The buffer is critical for successful seq-DAP-seq. If an optimized buffer from EMSA or other experiments is known, we recommend using this buffer. For most TF complexes IP or EB buffer may be used without dithiothreitol or beta-mercaptoethanol as these reducing agents may

negatively affect performance of anti-FLAG and anti-Myc magnetic beads. TCEP does not affect bead performance and may be used at concentrations up to 1 mM.

15. All buffers and reagents should be at 4 °C and kept on ice.

16. For sequential pull-down reactions, we prefer to use anti-FLAG magnetic beads and a FLAG peptide elution step followed by immobilization on anti-Myc magnetic beads due to the efficient elution of the proteins using 3x FLAG peptides and the high affinity and specificity of the anti-Myc magnetic beads.

17 Add 20 µl of 1x SDS-PAGE Protein Sample Buffer to the remaining beads for western blot verification of protein complex. Boil 20 µl of resuspended beads for 1 min, place on the magnetic stand and run 10 µl of supernatant on a 12% SDS PAGE for Western blot to control the immunoprecipitation.

18. These DNA-binding steps (1-9) should be performed in triplicate or quadruplicate for robust statistics.

19. AMPure beads are efficient for purification of DNA fragments and size selection. We have found that 1.2-1.5x volume of beads, 5 min. binding of DNA followed by 2 washes in 70% ethanol, 5-10 min. drying on the benchtop at room temperature and 30 µl elution (EB) give the best results. A slightly higher volume of beads is recommended by the manufacturer (1.8x).

20. Optional size selection may be performed as follows. We have found that these steps decrease the amount of DNA available for sequencing and are often not necessary.

1. Place up to 60 μ l of the pool into a new 1.5 ml reaction tube and add 12 μ l of 6x gel loading dye.
2. Load the sample across a large well of a 1% (wt/vol) agarose gel containing 0.005% (vol/vol) ethidium bromide or SybrSafe.
3. Load a 100-bp ladder in a separate lane.
4. Run the sample at 100 V for approximately 20 min.
5. Cut a ~200 to 400 bp DNA smear from the gel.
6. Extract the DNA using the Qiagen Gel Extraction Kit following the manufacturers manual.
7. Elution in 25 μ l of EB.

Fragment size of 200-400 bp is selected due to optimal read length for Illumina MiSeq Instrument and may vary depending on sequencing platform used. Longer reads up to 600 bp are possible but the genomic library should then be prepared with longer fragment size.

21. If possible positive and negative qPCR controls should be performed prior to deep sequencing (if known binding sites are available from ChIP data, for example). If there is no difference in amplification of the positive vs. negative control, this is an indication that the seq-DAP-seq protocol has failed. We have not had a case where the NGS gave useable data if the positive control qPCR verification failed. We recommend optimizing the binding buffer for seq-DAP-seq and repeating the protocol. We have found that amplicon sizes of 150 bp and 5-10 positive and 5-10 control regions is sufficient for library verification. 1 μ l of library diluted to 50-100 μ l is sufficient for qPCR verification. Enrichment of 4-fold or higher (CT cycle differences of at least 2 for positive and control qPCR) is a good indication that the experiment worked (Figure 9).

22. With the example data, the pipeline runs within an hour on a standard computer (6 CPU and 15 Gb memory).

23. In the FastQC output, one should pay attention to the quality score at each position along the reads. The quality score usually decreases towards the 3' end of the reads. After adaptor sequences have been trimmed off (NGmerge), one can trim low quality reads with Trimmomatic with option LEADING: 3 TRAILING: 3 SLIDINGWINDOW: 4:20 MINLEN: 36. This setting will remove leading/trailing low quality or N bases (below quality 3), scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 20 and drop reads which are less than 36 bases long after these steps. Quality score of 20 means that there is a probability of 1/100 that the base is incorrect. Trimmomatic has other options to trim reads with alternative strategies.

24. This step identifies genomics regions bound by the factor. In the alignment files, an excess of read is expected at bound regions. The difficulty lies in differentiating regions that collected lot of reads due to TF binding from the ones that accumulate lot of reads due to random/biased sequencing. Results of the peak calling should be checked using a genome browser to sense the quality of the peak calling. For this coverage files (BAM or bedgraph) peaks coordinates files (BED format) for each replicate can be loaded on a genome browser (IGB or IGV) for visual inspection. We have found this to be a simple way to qualitatively assess the peak calling step.

25. Since the tools used in this step are memory demanding it is preferable to run it on a subset of the data, for example on the 600 most significant peaks in the MSPC output as shown in the

example here. This constitutes the training set for motif detection and model construction. The remaining sequences will be used as a positive test set in the model validation step.

26. Meme-chip may be run several times varying the minimum and maximum motif width so as to tune the motif detection. One can start with wide range of width (4 to 20 bp) with options `–meme-minw` set to 4 and `–meme-maxw` set to 20. Then visualize the result to spot the most significant motif(s) and re-run with a size range that correspond to this motif to trim off edges with low information content values (below ~ 0.25). An important option is the `-meme-pal` option that will look for palindromic motif only.

27. A ROC curve close to the diagonal with an AUC near 0.5 means that the model randomly collects sequences, whereas a curves well above the diagonal ($\text{AUC} \gg 0.5$) means that the models correctly identifies most sequences from the positive set.

5. References

1. Babu MM, Luscombe NM, Aravind L, et al (2004) Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology* 14:283–291
2. Robertson G, Hirst M, Bainbridge M, et al (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 4:651–657
3. Mukhopadhyay A, Deplancke B, Walhout AJM, et al (2008) Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in *Caenorhabditis elegans*. *Nat Protoc* 3:698–709
4. Meers MP, Bryson TD, Henikoff JG, et al (2019) Improved CUT&RUN chromatin profiling tools. *Elife* 8:e46314

5. Skene PJ and Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6:e21856
6. Skene PJ, Henikoff JG, and Henikoff S (2018) Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* 13:1006–1019
7. Kaya-Okur HS, Wu SJ, Codomo CA, et al (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10:1930
8. Riley TR, Slattery M, Abe N, et al (2014) SELEX-seq, a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods Mol Biol* 1196:255–278
9. Bulyk ML (2007) Protein binding microarrays for the characterization of DNA-protein interactions. *Adv Biochem Eng Biotechnol* 104:65–85
10. Tuerk C and Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510
11. Slattery M, Riley T, Liu P, et al (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *147:1270–82*
12. Ellington AD and Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *346:818–22*
13. O'Malley RC, Huang SC, Song L, et al (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* 165:1280–1292
14. Bartlett A, O'Malley RC, Huang SC, et al (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* 12:1659–1672
15. Li M, Lin W, Hinckley W, et al (2022), Uncovering the “ZIP code” for bZIP dimers reveals novel motifs, regulatory rules and one billion years of cis-element evolution, <https://www.biorxiv.org/content/10.1101/2022.04.17.488518v1>

16. Lai X, Vega-Léon R, Hugouvieux V, et al (2021) The intervening domain is required for DNA-binding and functional identity of plant MADS transcription factors. *Nat Commun* 12:4760
17. Lai X, Stigliani A, Lucas J, et al (2020) Genome-wide binding of SEPALLATA3 and AGAMOUS complexes determined by sequential DNA-affinity purification sequencing. *Nucleic Acids Research* 48:9637–9648
18. Vanyushin BF and Ashapkin VV (2011) DNA methylation in higher plants: Past, present and future. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1809:360–368
19. Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
20. Gaspar JM (2018) NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. 19:536
21. Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. 9:357–9
22. Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
23. Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. 47:11 12 1-34
24. Gaspar JM (2018) Improved peak-calling with MACS2.
25. Bailey TL, Boden M, Buske FA, et al (2009) MEME SUITE: tools for motif discovery and searching. 37:W202-8
26. Healey A, Furtado A, Cooper T, et al (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* 10:21

27. Bolger AM Lohse, M& Usadel, B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *bioRxiv* 170
28. Quinlan AR and Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *26*:841–2
29. Dale RK, Pedersen BS, and Quinlan AR (2011) Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *27*:3423–4
30. Jalili V, Matteucci M, Masseroli M, et al (2018) Using combined evidence from replicates to evaluate ChIP-seq peaks. *34*:2338
31. Wilson P, Ganguly D, Hou X, et al CTAB genomic DNA extraction from Arabidopsis leaf material, <https://www.protocols.io/view/ctab-genomic-dna-extraction-from-arabidopsis-leaf-quidwue>
32. Anderson CB, Franzmayr BK, Hong SW, et al (2018) Protocol: a versatile, inexpensive, high-throughput plant genomic DNA extraction method suitable for genotyping-by-sequencing. *Plant Methods* 14:75

Figure captions

Figure 1. Overview of seq-DAP-seq protocol. Simplified schematic depicting the main steps for a seq-DAP-seq experiments. A genomic DNA library is prepared, fragmented and sequencing adaptors are added. Optionally, the gDNA library may be PCR amplified and barcodes added to create ampDAP input libraries. Tagged TFs are expressed and purified using one or more pulldown steps and magnetic beads. TF complexes are immobilized on magnetic beads and the DAP or ampDAP input library is added. The TF-DNA complexes are washed to decrease the amount of non-specific or poorly bound DNA. The proteins are then denatured to

release the DNA fragments. The DNA fragments are PCR-amplified, barcoded if necessary and sequenced using paired-end sequencing.

Figure 2. Schematic of DNA end repair reaction. 5' phosphate is shown in red.

Figure 3. Schematic of A-tailing reaction. Added 3' A is labelled.

Figure 4. Y-adaptor annealing and ligation reaction. The Y-adaptor has an overhanging T complementary to the 3' A from the A-tailing reaction step 21.

Figure 5. AmpDAP library barcoding. Barcodes are added using a unique indexing primer.

Figure 6. Protein binding to magnetic beads. For a two-protein transcriptional complex, sequential pull downs are performed to isolate a single species (green and yellow circles). In the first pulldown step, the TF complex is immobilized on magnetic beads via an epitope tag (i.e. 3xFLAG tag with anti-FLAG magnetic beads) and eluted.

Figure 7. Second pull down to isolate the hetero TF complex. A second pull down using a different epitope tag is used to purify the hetero TF complex.

Figure 8. DNA binding and amplification. The DAP or ampDAP input library is mixed with the TF complex of interest, the tagged TF complex is immobilized on a magnetic rack and washed to remove non-specific DNA or poorly bound DNA. The TF-DNA complexes are resuspended in IP buffer, the proteins denatured by heating and the supernatant containing the

enriched DNA library collected. The DNA is amplified, barcoded (if necessary) and ready for sequencing.

Figure 9. Example of qPCR amplification quality control, green: positive sets; red: negative sets.

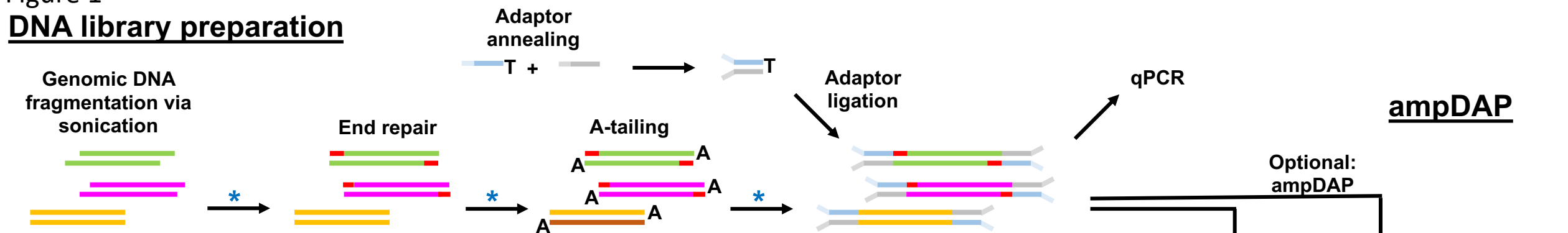
Figure 10. Overview of the bioinformatic analysis pipeline for DAP-seq sequencing reads. The individual steps are color-coded and numbered and the required programs given for each step in white.

Figure 11. Example of IGB snapshot from ampDAP-seq experiments. In black are successfully mapped reads and peaks for triplicates. In blue is an example of a failed ampDAP-seq experiment with no identifiable peaks.

Figure 12. Detailed example of barcoding and DNA amplification. Primers are color-coded and labelled.

Figure 1

DNA library preparation



TF complex pull downs

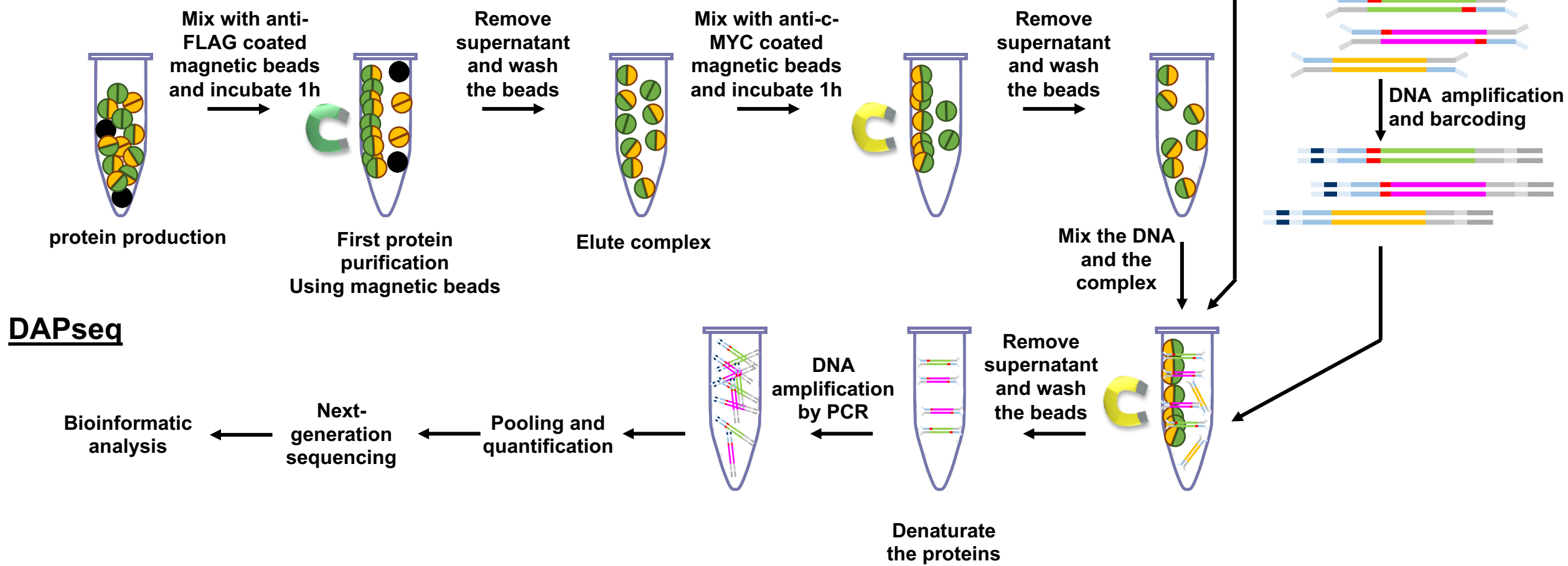


Figure 2

Genomic DNA fragmentation via
sonication

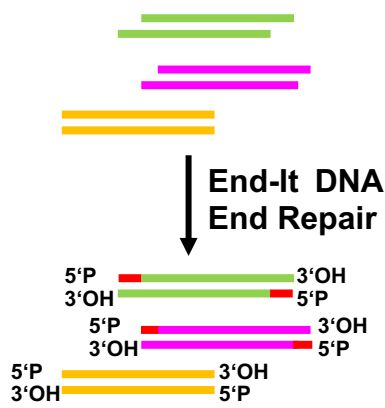


Figure 3

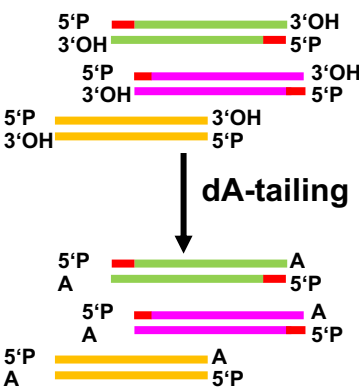


Figure 4

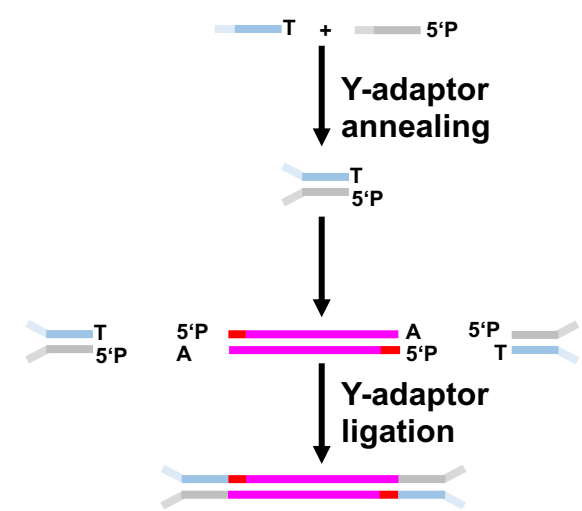


Figure 5

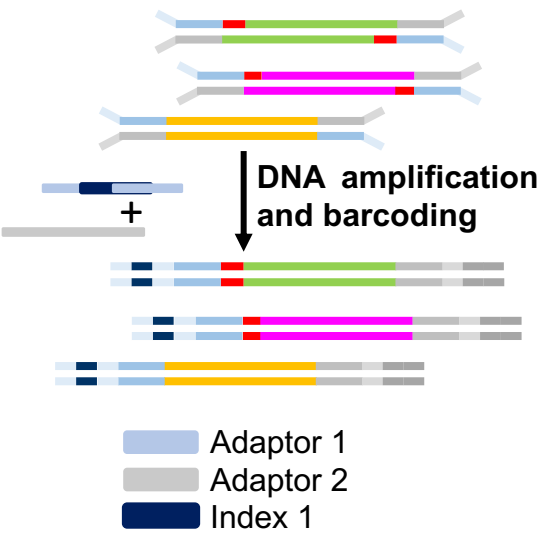


Figure 6

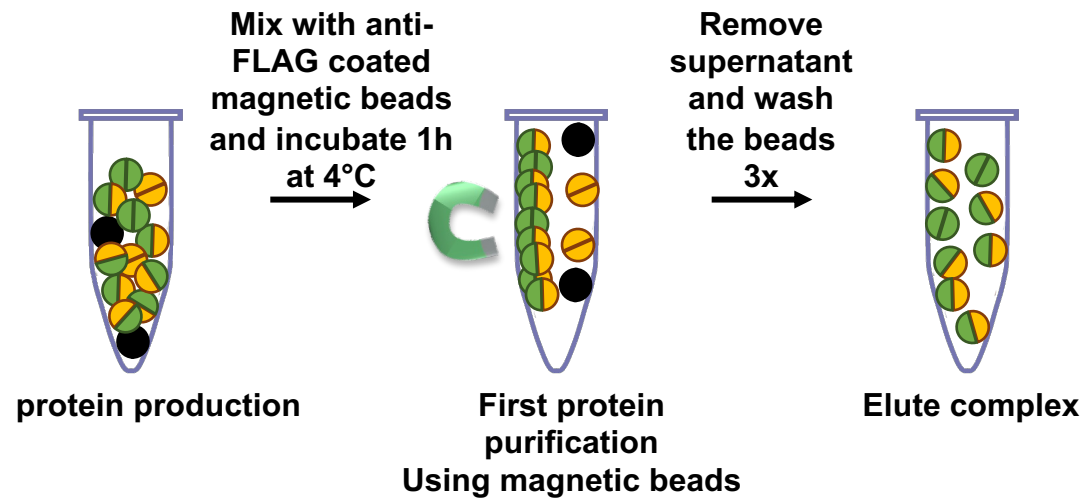


Figure 7

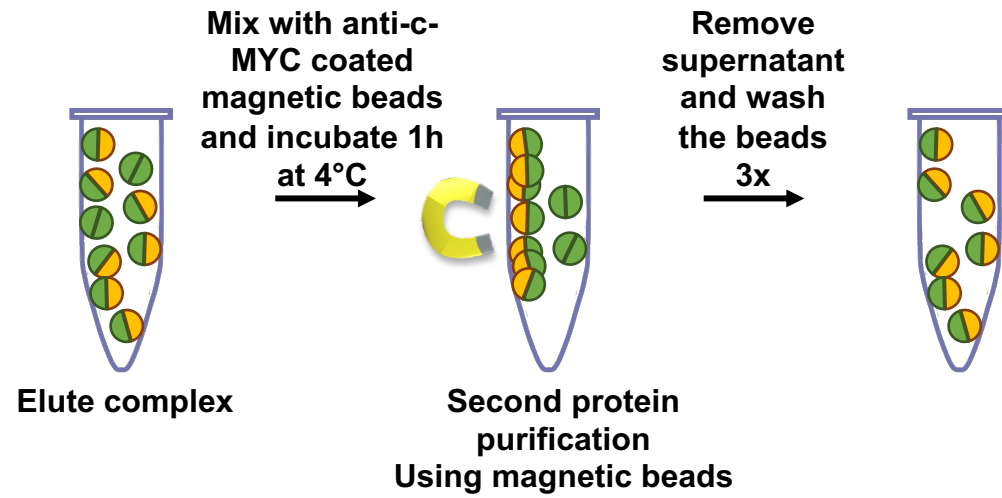


Figure 8

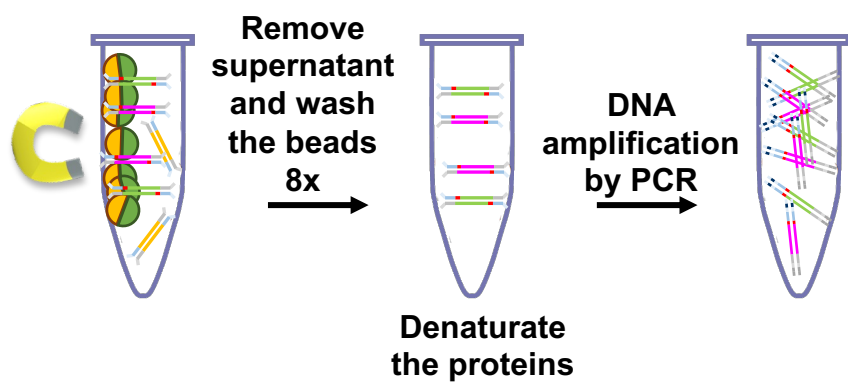


Figure 9

