Chapter 7

Identification of Brassinosteroid Target Genes by Chromatin Immunoprecipitation Followed by High-Throughput Sequencing (ChIP-seq) and RNA-Sequencing

Trevor Nolan, Sanzhen Liu, Hongqing Guo, Lei Li, Patrick Schnable, and Yanhai Yin

Abstract

Brassinosteroids (BRs) play important roles in many growth and developmental processes. BRs signal to regulate BR-INSENSITIVE1-ETHYL METHANESULFONATE-SUPPRESSOR1 (BES1) and BRASSINAZOLE-RESISTANT1 (BZR1) transcription factors (TFs), which, in turn, regulate several hundreds of transcription factors (termed BES1/BZR1-targeted TFs or BTFs) and thousands of genes to mediate various BR responses. Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq) with BES1/BZR1 and BTFs is an important approach to identify BR target genes. In combination with RNA-sequencing experiments, these genomic methods have become powerful tools to detect BR target genes and reveal transcriptional networks underlying BR-regulated processes.

Key words Transcription factor, Target genes, ChIP-seq, RNA-seq, Gene expression

1 Introduction

Brassinosteroids (BRs) are an important group of plant steroid hormones that regulate numerous processes, including growth and development as well as both biotic and abiotic responses [1, 2]. BRs are perceived by the plasma membrane localized receptor BRASSINOSTEROID INSENSITIVE1 (BRI1) and coreceptor BRI1-ASSOCIATED KINASE1 (BAK1) and several other signaling components; ultimately, BR signaling activates BR-INSENSITIVE1-ETHYL METHANESULFONATE-SUPPRESSOR1/BRASSINAZOLE-RESISTANT1 (BES1/BZR1) family transcription factors (TFs) [3–6]. BES1 and BZR1 bind to the promoters of over 6600 target genes [7, 8], including several hundred BES1/BZR1-targeted TFs (BTFs), directing a transcriptional network that controls the expression of approximately 5000 BR-regulated genes under various growth stages and

Eugenia Russinova and Ana I. Caño-Delgado (eds.), *Brassinosteroids: Methods and Protocols*, Methods in Molecular Biology, vol. 1564, DOI 10.1007/978-1-4939-6813-8_7,

conditions. Understanding how BES1 and BTFs control BR-regulated gene expression requires information regarding the genes bound and regulated by specific TFs in the signaling network.

RNA sequencing (RNA-seq) and chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) are now established as powerful techniques to determine genes that are transcriptionally regulated and directly bound by a particular TF, respectively [9, 10]. To study BR-regulated genes by RNA-seq, plants are grown and treated with or without the most active BR, brassinolide (BL). RNA is extracted and submitted for library preparation and sequencing by means of the Illumina platform. Genes regulated by a TF can be determined by performing RNA-seq on loss-of-function mutants for a TF (or its family members) and/or gain-of-function overexpression lines. Similarly, ChIP-seq with antibodies generated against a TF or with tagged TF lines (e.g., MYC, FLAG, or green fluorescent protein [GFP]) can reveal the direct targets of that TF. ChIP is accomplished by cross-linking DNA–protein complexes in vivo with formaldehyde, performing nuclear isolation followed by sonication to sheer the chromatin, and, subsequently, immunoprecipitating the DNA–TF complex with specific antibodies to the tag or TF. The DNA can then be reverse cross-linked, purified, and validated by ChIP-quantitative polymerase chain reaction (qPCR) of known target genes before library preparation and next-generation sequencing. The high-throughput sequencing data can be analyzed by bioinformatics and statistic tools. In this chapter, the RNA-seq and ChIP-seq protocols to identify BR target genes are described and an overview of the methods for data analyses is presented as well.

In combination, RNA-seq and ChIP-seq can be used to determine the genes a particular TF binds to and regulates. In the case of BTFs, these studies provide additional insight into how BR signaling is carried out and fine-tuned, but this methodology is generally applicable to any signaling pathway or process of interest. Genome-wide ChIP experiments in combination with RNA-seq are used to identify the targets for both BES1 and BZR1 [7, 8, 11]. When combined, approximately 27 % or 1765/6629 genes bound by BES1 and/or BZR1 are also regulated by BRs (Fig. 1).

The genes that are both bound and regulated by a TF have been considered the gold standard in ChIP studies, but recent work has demonstrated a hit-and-run model that occurs for at least some TFs [12, 13]. In this situation, genes are transiently bound by a TF and are activated, with dissociation of the TF, meaning that genes that are regulated, but not stably bound, could still be true targets of a TF. In any case, combining ChIP-seq and RNA-seq can help determine both the class of target genes that are stably bound and regulated by a TF and the genes that are regulated but

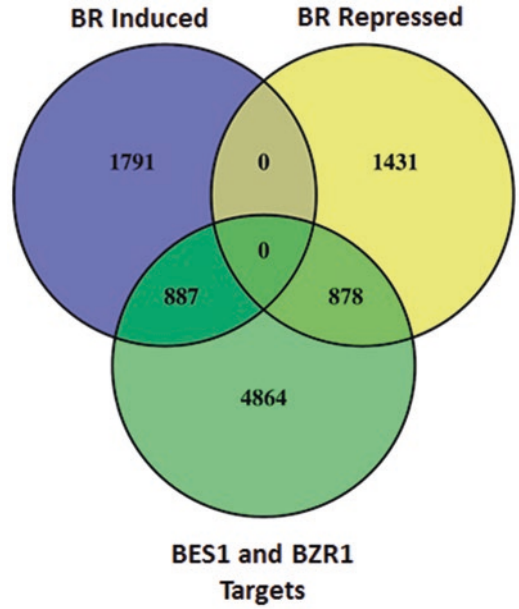

Fig. 1 Comparison of BR-regulated genes from RNA-seq and BES1 and BZR1 target genes identified by ChIP [7, 8, 11].

not bound by a TF. The latter either represents indirect regulation by other TFs or transiently bound targets, which can be distinguished by additional experiments. The protocols presented are based on published work and our own experimentation [14–16].

2 Materials

2.1 Plant Materials

1. *Arabidopsis thaliana* (L.) Heynh. seeds, including wild type (Columbia-0 accession), overexpression lines, and mutant lines for the TF of interest.
2. Plant growth trays and pots.
3. Soil.
4. Plant growth chamber with 16 h light–8 h dark regime and 20–22 °C temperature.

2.2 ChIP Solutions, Kits, and Reagents

1. 1 % (v/v) formaldehyde.
2. 2 M glycine.
3. Nuclear extraction buffer A: 10 mM Tris–HCl (pH 8.0), 0.4 M sucrose, 5 mM β-mercaptoethanol (add before use), 1 mM phenylmethylsulfonylfluoride (PMSF) (add before use), protease inhibitor cocktail (add before use).
4. Nuclear extraction buffer B: 10 mM Tris–HCl (pH 8.0), 0.25 M sucrose, 10 mM $MgCl_2$, 1 % (v/v) Triton X-100,

5 mM β-mercaptoethanol (add before use), 1 mM PMSF (add before use), protease inhibitor cocktail (add before use).

5. Nuclear Extraction Buffer C: 10 mM Tris–HCl (pH 8.0), 1.7 M sucrose, 2 mM $MgCl_2$, 0.15 % (v/v) Triton X-100, 5 mM β-mercaptoethanol (add before use), 1 mM PMSF (add before use), protease inhibitor cocktail (add before use).
6. Nuclear Lysis Buffer: 50 mM Tris–HCl (pH 8.0), 10 mM ethylenediaminetetraacetic acid (EDTA), 0.5 % (v/v) sodium dodecyl sulfate (SDS), 5 mM β-mercaptoethanol (add before use), 1 mM PMSF (add before use), protease inhibitor cocktail (add before use).
7. ChIP dilution buffer: 20 mM Tris–HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, 1 % (v/v) Triton X-100, 5 mM β-mercaptoethanol (add before use), 1 mM PMSF (add before use), protease inhibitor cocktail (add before use).
8. 5 M NaCl.
9. Protease K.
10. Antibodies: control immunoglobulin G (IgG) and antibody specifically recognizing the TF of interest or a tag (e.g., MYC, FLAG, or GFP) fused to the TF.
11. Dynabeads for protein A or protein G.
12. Low-salt wash buffer: 20 mM Tris–HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, 1 % (v/v) Triton X-100, 0.005 % (w/v) SDS.
13. High-salt wash buffer: 20 mM Tris–HCl (pH 8.0), 2 mM EDTA, 500 mM NaCl, 1 % (v/v) Triton X-100, 0.005 % (w/v) SDS.
14. TE buffer: 10 mM Tris–HCl (pH 8.0), 1 mM EDTA.
15. Chelating resin, such as Chelex 100 Resin (Bio-Rad).
16. ChIP elution buffer (freshly prepared): 0.1 M $NaHCO_3$, 1 % (w/v) SDS.
17. Glycogen.
18. SYBR Green PCR Master Mix.
19. ChIP DNA Clean & Concentrator Kit.
20. Illumina TrueSeq ChIP Sample Preparation Kit or Bio Scientific NEXTflex ChIP-seq Kit.

2.3 RNA-seq Solutions, Kits, and Reagents

1. 1 mM brassinolide: prepare stock in dimethylsulfoxide (DMSO) and store at −20 °C.
2. DMSO.
3. TRIzol.
4. Chloroform.

5. Isopropanol.
6. 75 % (v/v) ethanol made with RNase-free water.
7. RNase-free water.
8. RNase-Free DNase Set (Qiagen), including buffer RDD and DNase I stock solution used in DNase treatment.
9. Plant RNeasy Mini kit (Qiagen), including spin columns, 2-mL collection tubes, and buffers RLT and RPE used in RNA cleanup.

2.4 Equipment and Other Materials

1. 50-mL conical centrifuge tubes.
2. Miracloth.
3. Benchtop centrifuge equipped with a rotor capable of spinning 50-mL tubes.
4. Water bath sonicator.
5. Probe sonicator.
6. 1.5-mL low retention siliconized microcentrifuge tubes.
7. Magnetic stand.
8. Microcentrifuge.
9. Mx4000 Real-time PCR machine.
10. Mortar and pestle.
11. NanoDrop ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific) or 2100 Bioanalyzer (Agilent).
12. Vacuum pump.
13. Tube rotator.

3 Methods

3.1 ChIP-sequencing (ChIP-seq)

3.1.1 Preparation and Cross-Linking of Plant Materials

1. Grow healthy 4- to 5-week-old plants under long-day conditions (16 h light–8 h dark) at 20–22 °C. Include wild-type control plants as well as loss-of-function (T-DNA insertion or RNA interference [RNAi] mutant) and/or overexpression lines (*see* **Note 1**). Plants should be randomly distributed in growth flats to avoid differences due to environment.
2. Collect 4 g of plant leaf tissue for each genotype in a 50-mL tube. Rinse the plants with distilled water to remove residual soil. Add 37 mL of 1 % (v/v) formaldehyde to submerge the plant tissue. Apply vacuum for 15 min (release/reapply vacuum once at 7.5 min to help infiltration of fixing solution to plant tissue). For large amount of plants, *see* **Note 2**.
3. Add 2.5 mL of 2 M glycine to stop cross-linking and apply vacuum for 10 min.

4. Wash the fixed tissue with distilled water (three times). Dry the tissue gently with paper towels. Freeze the tissue in liquid nitrogen wrapped in aluminum foil. The tissue can be processed immediately or stored at −80 °C.

3.1.2 Chromatin Isolation and Sonication

1. Grind the tissue to a fine powder (*see* **Note 3**) with a medium-size mortar and pestle. Add 40 mL of freshly prepared nuclear extraction buffer A to the powder. Mix with a spatula and vortex at high-speed for 10 s.
2. Filter the mixture through two layers of Miracloth into a new 50-mL tube (*see* **Note 4**).
3. Centrifuge at 2000 × *g* for 20 min in a benchtop centrifuge.
4. Remove the supernatant and resuspend the pellet with wide-mouth pipette tips (can cut the regular tips with scissors) in 1.2 mL freshly prepared Nuclear extraction buffer B. Centrifuge in a microfuge for 10 min at 12,000 × *g* at 4 °C.
5. Remove the supernatant and resuspend the pellet in 0.4 mL freshly prepared nuclear extraction buffer C. Centrifuge in a microfuge for 60 min at 12,000 × *g* at 4 °C.
6. Remove the supernatant and resuspend the nuclear pellet with 0.2 mL freshly prepared nuclear lysis buffer by pipetting. Add 0.4 mL freshly prepared ChIP dilution buffer. Mix well and keep on ice.
7. Sonicate the chromatin with a water bath sonicator in icy water for 30 cycles with 30 s-on and 30 s-off cycle. Replace the icy water after 15 min to prevent overheating of the chromatin samples.
8. Centrifuge at maximum speed for 10 min. Transfer 0.5 mL supernatant to a new tube. The chromatin can be stored at −80 °C at this stage.
9. Take 10 μL sonicated chromatin. Add 15 μL of ChIP dilution buffer and 1 μL 5 M NaCl, treated with 1 μL protease K at 50 °C for 30 min. Reverse cross-linking at 65 °C for 5 h and check the DNA size with 2 % (w/v) agarose gel (Fig. 2). For typical sonication results and use of alternative probe sonication, *see* **Note 5**.

3.1.3 Chromatin Immunoprecipitation and Washing

1. Dilute 0.5 mL chromatin solution ten times by adding 4.5 mL of ChIP dilution buffer. Use 0.8 mL to 1 mL for each immunoprecipitation reaction. From this step on, use 1.5-mL siliconized tubes.
2. Add 2–5 μg of IgG or specific antibody to each tube containing the diluted chromatin solution. Rotate at 4 °C overnight. For three genotypes, six samples need to be processed for each biological repeat.

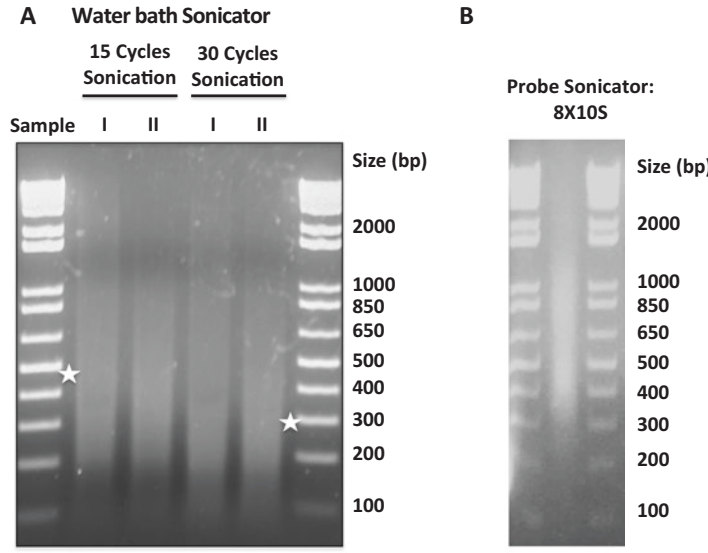

Fig. 2 Comparison of sonication methods for ChIP. (**a**) Sonication of chromatin for 15 or 30 cycles in a water bath sonicator. After 30 cycles, the average size of the chromatin fragments shifts to a decreased size (indicated with stars). (**b**) Sonication with a probe sonicator (power setting 2, 8 × 10 s, 1 min cooling).

3. For six samples, dilute 0.2 mL protein A Dynabeads (*see* **Note 6**) with 1 mL of ChIP dilution buffer in the 1.5-mL tubes by placing the bead-containing tubes in a magnetic stand for 20 s. Carefully remove the supernatant. Add again 1 mL of ChIP dilution buffer. Mix by inverting and precipitate the beads again. Add ChIP dilution buffer to 0.2 mL. The Dynabeads should be scaled up, when handling several biological repeats at the same time.
4. Add 30 μL of Dynabeads to each ChIP sample. Rotate at 4 °C in a tube rotator for 4 h.
5. Precipitate the Dynabeads with the ChIP complex as described above. Remove the supernatant and wash the beads with 1 mL of cold low-salt washing buffer.
6. Wash the Dynabeads with 1 mL of cold high-salt washing buffer (*see* **Note 7**).
7. Wash the beads twice with 1 mL of cold TE buffer and carefully remove any residual TE.

3.1.4 Elution of DNA and qPCR Analysis (See Note 8)

1. Add 0.1 mL 10 % (w/v) freshly prepared Chelex resin (1 g in 10 mL sterile water) to the Dynabeads (Subheading 3.1.3, **step 7**). Mix by vortexing.
2. Boil the mixture for 10 min to reverse cross-link the DNA–protein complexes. Mix the beads two to three times during the boiling process. Cool to room temperature.

3. Add 1 μL of protease K (10 μg/μL) to each tube and incubate at 50 °C in a water bath for 30 min. Mix three to four times during the incubation.
4. Boil the mixture again for 10 min to inactivate protease K.
5. Centrifuge 5 min at 12,000 × *g* at room temperature in a microcentrifuge. Collect the supernatant into a new tube as the first elution.
6. Add 0.1 mL of TE to the beads. Vortex and centrifuge as above. Collect 0.1 mL of supernatant and add to the first elution.
7. To perform qPCR with known target genes of a TF, for each gene, design a minimum of three pairs of primers, at different positions of the promoter region, with two pairs near potential TF-binding sites and one 3–5 kb upstream of the transcription start site as a negative control.
8. For each sample, use 2–4 μL of the purified DNA in each PCR reaction with 2–3 PCR reactions (technical repeats) per sample.
9. Run the PCR reactions as follows: 1 cycle at 95 °C for 10 min to activate the Taq polymerase, 40 cycles at 95 °C for 30 s, at 55 °C for 1 min, and at 72 °C for 30 s to amplify the DNA and collect fluorescence data, finally 1 cycle at 95 °C for 1 min and 55 °C for 30 s to obtain a melting curve for quality control of the primer pairs and PCR reactions.
10. Set the threshold to obtain a threshold cycle (Ct) for each reaction. The threshold is where significant and specific amplification occurs. Use the housekeeping gene ubiquitin 5 (UBQ5) as a reference gene to normalize the abundance of target genes. The normalized target gene abundance is calculated by delta threshold cycle (ΔCt).

    ΔCt1 = Ct (target gene treated) − Ct (UBQ5 treated).

    ΔCt2 = Ct (target gene control) − Ct (UBQ5 control).

    ΔΔCt = ΔCt1 − ΔCt2.

    the normalized abundance of the target gene = $2^{\Delta\Delta Ct}$.
11. Compare the difference in enrichment of the promoter fragments by comparing $2^{\Delta\Delta Ct}$: the higher the $2^{\Delta\Delta Ct}$, the more enrichment of the specific promoter fragment.

3.1.5 Elution of DNA for ChIP-seq (See Note 8)

1. Elute DNA from Dynabeads (Subheading 3.1.3, **step 7**) with 0.25 mL of elution buffer at 65 °C. Repeat the elution by adding a second aliquot of 0.25 mL of elution buffer.
2. Combine the two elutions. Add 20 μL 5 M of NaCl and reverse cross-linking by incubating at 65 °C overnight.

3. Add 1 μL of protease K and incubate at 50 °C for 1 h.
4. Purify DNA with the ChIP DNA Clean and Concentrator Kit. For an alternative purification method, *see* **Note 9**.
5. Amplify the ChIP products with the Illumina TrueSeq ChIP Sample Preparation Kit or Bio Scientific NEXTflex ChIP-seq Kit according to the manufacturer's protocols. The sequencing is done with an Illumina Hi-seq 2500 with 50-bp reads.

3.2 RNA Sequencing (RNA-seq)

3.2.1 Plant Growth and BL Treatment

1. Grow plants as described for ChIP-seq (*see* Subheading 3.1.1).
2. Treat plants with 1 μM of BL prepared in water or with the same amount of DMSO (mock treatment) by gently spraying (approximately 50 mL for one tray of plants).
3. After 2.5 h, flash-freeze tissue in liquid nitrogen, pooling whole rosettes of at least three to four randomly chosen plants per sample. Harvest enough samples for at least three biological replicates for each genotype and treatment.

3.2.2 RNA Extraction

1. Grind tissues to a fine powder using liquid nitrogen, a mortar, and pestle. Take care that the tissue stays frozen during the grinding process by adding more liquid nitrogen as needed.
2. Using a prechilled spatula (in liquid nitrogen), transfer 100 mg of the tissue powder to a precooled 1.5-mL tube. Extra tissue powder can be stored separately at −80 °C for backup or future processing (*see* **Note 10**).
3. Add 1 mL of TRIzol reagent to the 1.5-mL tube containing the frozen tissue powder. Shake vigorously until the mixture is well resuspended (*see* **Note 11**).
4. Incubate the sample at room temperature for 5 min.
5. Add 0.2 mL of chloroform and shake the tube by hand for 15 s.
6. Incubate the tube for 2–3 min at room temperature, then centrifuge at 12,000 × *g* for 15 min at 4 °C in a microcentrifuge.
7. Remove the upper aqueous phase and place it in a new 1.5-mL tube, taking care not to disturb the interphase.
8. Add 0.5 mL of 100 % (v/v) isopropanol to the newly transferred aqueous phase and incubate at room temperature for 10 min.
9. Centrifuge at 12,000 × *g* for 10 min at 4 °C in a microcentrifuge.
10. Carefully remove the supernatant from the tube and wash the pellet by adding 1 mL of 75 % (v/v) ethanol.
11. Vortex the sample briefly and centrifuge 7500 × *g* for 5 min at 4 °C.

12. Remove the supernatant and allow the pellet to air dry for 5–10 min. Be careful not to overdry the pellet, because it makes it difficult to resuspend it completely.
13. Resuspend RNA in 87.5 μL of RNase-free water by heating for 10–15 min at 55 °C.

3.2.3 DNase Digestion

1. Perform DNAse digestion with RNase-Free DNase Set by adding 10 μL buffer RDD and 2.5 μL DNase I stock solution to each tube.
2. Incubate the tubes at room temperature for 10 min.

3.2.4 RNA Clean Up using RNeasy Mini Kit

1. Add 0.35 mL of buffer RLT to each tube. Mix well, add 0.25 mL 100 % (v/v) ethanol, and mix again.
2. Load the sample into RNeasy Mini spin column placed in a 2-mL collection tube. Centrifuge at approximately 10,000 × *g* for 15 s. Discard the flow through.
3. Add 0.5 mL of buffer RPE to the column. Centrifuge at approximately 8000 × *g* for 15 s. Discard the flow through.
4. Add another 0.5 mL of buffer RPE to the column. Centrifuge at approximately 8000 × *g* for 2 min. Discard the flow through.
5. Place the column in a new 2-mL collection tube and centrifuge at maximum speed for 1 min to remove any residual buffer RPE.
6. Transfer the column to a new 1.5-mL collection tube and elute the RNA by adding 30–50 μL of RNase-free water directly onto the spin column membrane. Centrifuge 1 min at approximately 8000 × *g*.
7. Assess the quality and concentration of the RNA with a bioanalyzer (NanoDrop or Agilent) (*see* **Note 12**).

3.2.5 Library Preparation

After RNA isolation and quality control, samples can be used for library preparation. The type of library preparation used will depend on the sequencing platform and conditions. Several protocols or kits are available. Alternatively, RNA can be submitted directly to a core facility or sequencing provider for library preparation and sequencing.

3.3 Bioinformatics Analyses of RNA-seq and ChIP-seq Data

RNA-seq and ChIP-seq analyses share similar procedures for data quality assessment, sequence trimming, alignment, and read counting. The sections overview the protocol for sequencing data for the RNA-seq and ChIP-seq experiments generated by the Illumina technology.

3.3.1 Data Quality Assessment

RNA-seq and ChIP-seq generate massive amounts of sequencing data that are generally delivered in FASTQ format (en.wikipedia.org/wiki/FASTQ_format). The quality of raw sequencing data

can be assessed with FastQC (bioinformatics.babraham.ac.uk/projects/fastqc) that provides graphs and summary tables to overview data amount, quality, and adaptor contamination.

3.3.2 Quality and Adaptor Trimming

Raw data generated from a sequencing platform contain low-quality bases, typically enriched at the 3′-ends of each sequence, and adaptor sequences. The precise removal of low-quality sequences and adaptor contamination is important for downstream analyses. Many software tools have been developed for these purposes [17]. Trimmomatic [18] is a commonly used as trimming software that includes two modes for adaptor trimming: simple and palindrome mode. The former mode is universal for data from any sequencing platform, whereas the latter is designed for paired-end sequences. Palindrome mode has considerable advantages in sensitivity and specificity. After processing, the trimmed data are still in FASTQ format for subsequent procedures.

3.3.3 Alignment of Reads to a Reference Genome

For both RNA-seq and ChIP-seq, trimmed sequencing reads are typically aligned to a reference genome. For RNA-seq, reads are largely generated from mature mRNAs that are subjected to splicing. Therefore intron-tolerant aligners, such as GSNAP [19], STAR [20], and Tophat [21], should be used for alignments. For ChIP-seq, reads are from genomic DNAs. Either BWA [22] or Bowtie (or Bowtie2) [23, 24] can be applied. Both aligners are fairly fast and accurate. Owing to genome repetitiveness, reads can align to multiple genomic locations. Also some reads are aligned with too many mismatches, potential signaling misalignments. To reduce misalignments, postalignment filtering is used. A simple filtering is to set a minimal alignment quality score.

Sequence Alignment/Map format (SAM) is a generic alignment format employed in all next-generation sequencing programs to store alignments [25]. As SAM is a readable text file that is not storage efficient, it is usually converted to a BAM format that contains the same information, but is in a binary format, greatly enhancing storage efficiency. The conversion is implemented in SAMtools [25] that also provides an indexing function for BAM to improve the efficiency of further alignment access.

3.3.4 Alignment Visualization

Integrative Genomics Viewer (IGV) [26] is a flexible and user-friendly tool to enable visualization of diverse genomic data, including BAM alignment data. Multiple alignment data sets from different samples or different experiments (e.g., RNA-seq and ChIP-seq) can be displayed in the same panel. Other genomic data, such as genome annotation and variants, can be provided as well for visual exploration of the relationship among different genomic features.

3.3.5 Read Depth (Counts) at Genomic Regions

Read depth of genomic regions is determined based on the overlap between coordinates of reads mapped in the reference genome and genomic locations of target regions. Existing tools are available to compute read depth per genomic region. Two recommended software packages are HTSeq [27] and bedtools (bedtools.readthedocs.org). HTSeq is tailored for counting RNA-seq reads. It is flexible to count reads in defined genic regions, such as exons, genes, or coding regions, provided in GFF or GTF genome annotation files. It also offers options to deal with the cases in which gene models in a reference genome overlap. Bedtools is a powerful tool to cross-check multiple files in diverse data formats, such as BAM, BED, and GFF/GTF, enabling the counting of reads in defined genomic regions and calculating sequence coverage of genomic regions. Genomic regions can be genes that are used in RNA-seq or defined windows (or bins) used in ChIP-seq. Once count data are generated, normalization and statistical tests will be applied to identify differentiation between biological groups at genomic regions of biological interest.

3.3.6 Data Normalization and Statistical Tests for RNA-seq

The goal of RNA-seq is to identify genes that are differentially expressed (DE) among samples that differ in some known manner, genotype, tissue type, treatment, etc. To this end, read counts are compared for given genes across samples. Differences in sequencing depths, technical variation, and biological variation are all sources for variation observed in read counts per unit (for instance, a gene) among samples. For RNA-seq, a good experimental design is required to estimate different variation sources and declare the statistical significance of the biological factor of interest. To account for variation in sequencing depth, a normalization method, i.e., reads per kilobase per million reads (RPKM), has been widely used [28]. Similar to RPKM, fragments per kilobase of exon per million mapped reads (FPKM) was used later to avoid double counting of paired-end reads [29]. For RNA-seq, several statistical approaches for the identification of differential expression were developed in recent years, including DESeq [30, 31], edgeR [32], Cuffdiff [33], and baySeq [34]. Each of these software packages has its own algorithms for normalization and statistical modeling of gene expression. For all the software packages, a statistical test is performed for a gene or an isoform, resulting in an associated P value. When multiple-hypothesis tests are conducted, the same P value cutoff (for example, 0.05) for each test is not enough to guarantee the same level of false positives as that of a single-hypothesis test. To cope with this multiple testing problem, a false discovery rate (FDR) has been widely used [35]. As a further development the FDR, the Q value was introduced as an adjusted P value for the FDR determination [36]. The FDR estimates the expected fraction of incorrect rejections among all the rejections.

After a DE set of genes has been found, a test for enrichment of DE genes in certain families or functional groups can be performed. For example, the software GOseq [37] is used for a gene ontology (GO) enrichment analysis. *P* values for overrepresentation of DE genes for each GO category can be generated by comparison to a null distribution of each GO category from repeated samplings of the same number of DE genes. Although the multiple-testing issue exists, the FDR for this analysis should be applied with caution due to the complicated relationship among GO terms breaking the assumption of independence of each test in most FDR algorithms. During sampling, genes are weighted with their read counts from all samples to account for the potential bias in the differential power to detect DE among genes with different numbers of read counts. In addition, other pathway visualization tools, such as MapMan [38] or Kyoto Encyclopedia of Genes and Genomes (KEGG; genome.jp/kegg), can be used to better understand which pathways and functional modules are subjected to significant expression regulation.

3.3.7 Data Normalization and Statistical Tests for ChIP-seq

The goal of ChIP-seq is to identify the genomic regions to which an immunoprecipitated (IP) protein (such as a TF) is bound. The fixed chromatin is fragmented by sonication and immunoprecipitated. Small genomic DNA fragments are purified, amplified, and subjected to high-throughput sequencing. CHANCE [39] evaluates IP strength by separately plotting curves for both the IP sample(s) and the control(s), showing the empirical cumulative percentage of reads covered by the windows/bins that are ordered increasingly by their aligned reads. Potential quality problems, such as weak IP enrichment and insufficient sequencing depth, can be identified with outputs from CHANCE.

By analyzing read depth at windows/bins in both IP samples and controls, genomic regions with significant read enrichment are considered to be IP peaks that are potential protein-binding regions. Sharp and broad peaks are two major types of peak signals. Sharp peaks, often associated with protein–DNA binding domains, can be detected by various peak callers, such as MACS [40], CisGenome [41], SPP [42], CSAR [43], ZINBA [44], HPeak [45], and BayesPeak [46]. The detection of broad peaks resulting, for example, from signals associated with histone modifications, requires different algorithms, including SICER [47], CCAT [48], and RSEG [49], or sharp peak callers, such as MACS, adjusted with the appropriate parameters. Similar to RNA-seq, a FDR [35] is needed to account for multiple testing.

To produce reproducible results, biological replicates of a ChIP-seq experiment are needed. To measure the reproducibility between replicates, irreproducible discovery rate (IDR) analysis can be applied to peaks identified from two replicates [50, 51].

When biological replication is applied, RNA-seq software packages, such as DESeq and edgeR, can also be adapted to deal with biological replicates and identify statistically significant signals in ChIP samples versus their controls by means of the ChIP-seq read counts [52].

4 Notes

1. Most protocols use seedlings for ChIP experiments, but we found that BRs affect more genes at the adult stage, so mature plants are used in this protocol for both ChIP-seq and RNA-seq experiments. The use of appropriate controls in critical for ChIP experiments. When an antibody produced against a TF is used, the ideal control would be a loss-of-function mutant or an RNAi line for that TF. Alternatively, when a tagged line is used, it is important to include a wild-type control. Whenever possible, the tagged line should be in a mutant background to avoid competition with the endogenous TF. In this case, the mutant is used as a control.
2. Generally, 4 g of tissue is enough for most TFs studied. However for low-abundance or tissue-specific TFs, additional plant material should be used. A protocol for scaling up ChIP experiments while retaining low background has recently been described [14]. For large amount of plants (for instance, 20 g of tissue), a 500-mL flask with 185 mL 1 % (v/v) formaldehyde should be used and vacuum period should be increased to 20 min. To stop cross-linking in the next step, use 12.5 mL 2 M glycine.
3. We found that complete grinding is critical for good nuclear yield. For 4 g of tissue, we usually grind for approximately 15 min with plenty of liquid nitrogen, to be added three to four times in the process; for large amounts of tissue, such as 20 g, grinding should be done for 30 min.
4. Do not use pipette tips to facilitate flowing of the mixture through the filter. You can swirl the tube with Miracloth gently if needed, but do not filter more than 40 mL with one set of Miracloth.
5. Sonication is a critical step and needs to be optimized for each experiment and laboratory conditions. Parameters include the type of sonicator used, power setting, number of cycles, and SDS concentrations. Several settings should be tested and the results checked by DNA gel electrophoresis. We found that a water bath sonicator works better than a probe sonicator under our experimental conditions, but it is possible to obtain good results with both types of sonicators (Fig. 2).

6. Depending on the type of primary antibody, protein A or protein G beads may be used. New England Biolabs provides an excellent reference for choosing between these (https://www.neb.com/tools-and-resources/selection-charts/affinity-of-protein-ag-for-igg-types-from-different-species). In our experience, magnetic beads, such as Dynabeads, produce low background and a high signal-to-noise ratio, but traditional protein A agarose beads could be substituted and often result in a higher total signal.
7. For antibodies with low affinities, the salt concentration can be reduced from 500 mM to 250 mM or the high-salt wash could even be completely omitted.
8. This method uses Chelex resin and boiling to reverse cross-linking. The protocol is quicker (saves 1 day) than elution with the 1 % (w/v) SDS and 0.1 M $NaHCO_3$. However, we found that the boiling process denatures small DNA fragments that do not affect the qPCR analysis. Nevertheless, for ChIP-seq in which the ChIP product will be ligated to adaptors, the elution should be done with SDS and $NaHCO_3$ and reverse cross-linking at 65 °C.
9. The DNA can also be purified by phenol–chloroform extraction followed by ethanol precipitation. Glycogen should be added to facilitate DNA precipitation.
10. The plant material is sufficient for multiple RNA extractions and the ground tissue powder can be kept as backup at −80 °C and used later if samples do not meet the quality control standards.
11. All RNA extraction steps should be carried out with filter tips. Take care to avoid RNase contamination. We routinely clean all benches and supplies to be used for RNA extraction with the decontamination detergent RNase Away before starting the experiment.
12. Samples should have an optical density $(OD)_{260/280}$ and $OD_{260/230}$ greater than 1.8. The amount of RNA needed will depend on the library preparation method used, but samples should generally contain at least 400 ng RNA at a concentration of approximately 10 ng/μL.

Acknowledgments

We thank Iowa State University DNA Facility for allowing us to use the Diagenode Bioruptor and Mike Baker for information on ChIP-seq product amplification. The work is supported by grant from NSF (IOS-1257631) and by the Plant Science Institute at Iowa State University.

References

1. Clouse SD (2011) Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development. Plant Cell 23:1219–1230
2. Li J, Chory J (1999) Brassinosteroid actions in plants. J Exp Bot 50:275–282
3. Guo H, Li L, Aluru M et al (2013) Mechanisms and networks for brassinosteroid regulated gene expression. Curr Opin Plant Biol 16: 545–553
4. Zhu J-Y, Sae-Seaw J, Wang Z-Y (2013) Brassinosteroid signalling. Development 140: 1615–1620
5. Jiang J, Zhang C, Wang X (2013) Ligand perception, activation, and early signaling of plant steroid receptor brassinosteroid insensitive 1. J Integr Plant Biol 55:1198–1211
6. Vriet C, Russinova E, Reuzeau C (2013) From squalene to brassinolide: the steroid metabolic and signaling pathways across the plant kingdom. Mol Plant 6:1738–1757
7. Yu X, Li L, Zola J et al (2011) A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in *Arabidopsis thaliana*. Plant J 65: 634–646
8. Sun Y, Fan X-Y, Cao D-M et al (2010) Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis*. Dev Cell 19:765–777
9. Mundade R, Ozer HG, Wei H et al (2014) Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. Cell Cycle 13:2847–2852
10. Soon WW, Hariharan M, Snyder MP (2013) High-throughput sequencing for biology and medicine. Mol Syst Biol 9:640
11. Wang X, Chen J, Xie Z et al (2014) Histone lysine methyltransferase SDG8 is involved in brassinosteroid-regulated gene expression in *Arabidopsis thaliana*. Mol Plant 7:1303–1315
12. Li Y, Mukherjee I, Thum KE et al (2015) The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants. Genome Biol 16:79
13. Para A, Li Y, Marshall-Colón A et al (2014) Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in *Arabidopsis*. Proc Natl Acad Sci USA 111:10371–10376
14. Lau OS, Bergmann DC (2015) MOBE-ChIP: a large-scale chromatin immunoprecipitation assay for cell type-specific studies. Plant J 84: 443–450
15. Nelson JD, Denisenko O, Sova P et al (2006) Fast chromatin immunoprecipitation assay. Nucleic Acids Res 34:e2
16. Gendrel AV, Lippman Z, Yordan C et al (2002) Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. Science 297:1871–1873
17. Chen C, Khaleel SS, Huang H et al (2014) Software for pre-processing Illumina next-generation sequencing short read sequences. Source Code Biol Med 9:8
18. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
19. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881
20. Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
21. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111
22. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
23. Langmead B (2010) Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics 32 Unit 11.7:1–14
24. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
25. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
26. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26
27. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169
28. Mortazavi A, Williams BA, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5:621–628
29. Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515
30. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion

for RNA-seq data with DESeq2. Genome Biol 15:550
31. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106
32. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140
33. Trapnell C, Hendrickson DG, Sauvageau M et al (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46–53
34. Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422
35. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B-Stat Methodol 57:289–300
36. Storey JD (2002) A direct approach to false discovery rates. J Roy Stat Soc Ser B-Stat Methodol 64:479–498
37. Young MD, Wakefield MJ, Smyth GK et al (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11:R14
38. Thimm O, Bläsing O, Gibon Y et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939
39. Diaz A, Nellore A, Song JS (2012) CHANCE: comprehensive software for quality control and validation of ChIP-seq data. Genome Biol 13:R98
40. Zhang Y, Liu T, Meyer CA et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137
41. Ji H, Jiang H, Ma W et al (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat Biotechnol 26:1293–1300
42. Kharchenko PV, Tolstorukov MY, Park PJ (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat Biotechnol 26:1351–1359
43. Muiño JM, Kaufmann K, van Ham RCHJ et al (2011) ChIP-seq analysis in R (CSAR): an R package for the statistical detection of protein-bound genomic regions. Plant Methods 7:11
44. Rashid NU, Giresi PG, Ibrahim JG et al (2011) ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol 12:R67
45. Qin Z, Yu J, Shen J et al (2010) HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data. BMC Bioinformatics 11:369
46. Spyrou C, Stark R, Lynch AG et al (2009) BayesPeak: Bayesian analysis of ChIP-seq data. BMC Bioinformatics 10:299
47. Zang C, Schones DE, Zeng C et al (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25:1952–1958
48. Xu H, Handoko L, Wei X et al (2010) A signal-noise model for significance analysis of ChIP-seq with negative control. Bioinformatics 26:1199–1204
49. Song Q, Smith AD (2011) Identifying dispersed epigenomic domains from ChIP-Seq data. Bioinformatics 27:870–871
50. Bailey T, Krajewski P, Ladunga I et al (2013) Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput Biol 9:e1003326. doi:10.1371/journal.pcbi.1003326
51. Li Q, Brown JB, Huang H et al (2011) Measuring reproducibility of high-throughput experiments. Ann Appl Stat 5:1752–1779
52. Furey TS (2012) ChIP-seq and Beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 13:840–852