

An optimized ChIP-Seq framework for profiling histone modifications in *Chromochloris zofingiensis*

Daniela Strenkert¹  | Matthew Mingay²  | Stefan Schmollinger¹  |
 Cindy Chen²  | Ronan C. O'Malley²  | Sabeeha S. Merchant^{3,4,5} 

¹California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA, USA

²DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

³Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

⁴Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley, CA, USA

⁵Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Correspondence

Sabeeha S. Merchant, Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA.
 Email: sabeeha@berkeley.edu

Funding information

U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Grant/Award Numbers: DE-SC0018301, DE-AC02-05CH11231

Abstract

The eukaryotic green alga *Chromochloris zofingiensis* is a reference organism for studying carbon partitioning and a promising candidate for the production of biofuel precursors. Recent transcriptome profiling transformed our understanding of its biology and generally algal biology, but epigenetic regulation remains understudied and represents a fundamental gap in our understanding of algal gene expression. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) is a powerful tool for the discovery of such mechanisms, by identifying genome-wide histone modification patterns and transcription factor-binding sites alike. Here, we established a ChIP-Seq framework for *Chr. zofingiensis* yielding over 20 million high-quality reads per sample. The most critical steps in a ChIP experiment were optimized, including DNA shearing to obtain an average DNA fragment size of 250 bp and assessment of the recommended formaldehyde concentration for optimal DNA-protein cross-linking. We used this ChIP-Seq framework to generate a genome-wide map of the H3K4me3 distribution pattern and to integrate these data with matching RNA-Seq data. In line with observations from other organisms, H3K4me3 marks predominantly transcription start sites of genes. Our H3K4me3 ChIP-Seq data will pave the way for improved genome structural annotation in the emerging reference alga *Chr. zofingiensis*.

KEYWORDS

ChIP-Sequencing, *Chromochloris*, epigenetics, formaldehyde crosslinking, green algae, histone lysine methylation

1 | INTRODUCTION

Chromatin is the higher order structure in which eukaryotic DNA is organized, facilitating condensed packaging of genomic DNA into the nucleus. Within chromatin, nucleosomes are considered to be the major structural subunit, in which approximately 146 bp of DNA are wrapped around dimers of core histones H2A, H2B, H3, and H4

(Luger, 2003). In addition to its packaging function, histones are hotspots for a variety of different posttranslational modifications (PTMs) that directly affect gene expression and thus contribute to the role of chromatin towards gene regulation (Li et al., 2007; Workman & Kingston, 1998). Modified N-terminal tails of histones emerge from the nucleosome core and are thus accessible to chromatin effectors, enzymes that convert them into regulatory function. Each organism's

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Plant Direct* published by American Society of Plant Biologists and the Society for Experimental Biology and John Wiley & Sons Ltd.

specific histone code consists of various different histone modifications (acetylation, monomethylation/dimethylation/trimethylation, phosphorylation, sumoylation, or ubiquitination) that can be combined to form distinct chromatin states, although the number of possible combinations is typically limited. In Arabidopsis, for instance, the combinatorial complexity is reduced to just nine distinct chromatin states (Sequeira-Mendes et al., 2014). Accordingly, the histone code provides an additional regulatory layer extending the information content of the genetic code (Luger, 2003). Previous studies have demonstrated that histone H3 lysine 4 trimethylation plays an important role during active transcription; for reviews, see Martin and Zhang (2005); Ruthenburg et al. (2007).

In *Saccharomyces cerevisiae*, trimethylation of lysine 4 at histone H3 (H3K4me3) is a signature motif marking transcription start sites (TSSs) of genes that are either actively transcribed or that are poised for transcription (Bernstein et al., 2002; Ng et al., 2003; Santos-Rosa et al., 2002; Schneider et al., 2004). In addition, H3K4me3 was implied in mediating epigenetic memory (Ng et al., 2003). Notably, enrichment of H3K4me3 at TSSs seems to be a highly conserved feature in many taxa. In human cell lines, mouse, and fly, TSSs of actively transcribed genes are also marked with H3K4me3 (Barski et al., 2007; Bernstein et al., 2005; Guenther et al., 2007; Schübeler et al., 2004). The same is true in land plants: In Arabidopsis, rice, and maize, trimethylation at lysine 4 of histone H3 occurred predominantly at gene promoters (He et al., 2010; Oh et al., 2008; Wang et al., 2009; Zhang et al., 2009; Zong et al., 2013).

Chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-Seq) is the method of choice for studying histone modification dynamics. ChIP is based on the recovery of DNA by immunoprecipitation using a specific antibody against the DNA-bound protein of interest, such as a modified histone. ChIP'ed DNA may be quantified using (q)PCR, microarrays or deep sequencing. Detailed and optimized ChIP protocols have been published for Tetrahymena (Dedon et al., 1991), Drosophila (Orlando et al., 1997), yeast (Hecht & Grunstein, 1999), mammalian cell lines (Das et al., 2004), Chlamydomonas (Strenkert et al., 2011), and land plants like Arabidopsis (Bowler et al., 2004), maize (Haring et al., 2007), and tomato (Ricardi et al., 2010). Here, we describe an optimized ChIP protocol for the green alga *Chromochloris*. As illustrated by ChIP-Seq with an antibody targeting H3K4me3, we show that our optimized protocol is suited for ChIP-Seq analyses of histone modifications with proper resolution and high coverage yielding excellent signal to noise during subsequent data analyses.

2 | MATERIALS AND METHODS

2.1 | Cultures and growth conditions

Chromochloris zofingiensis SAG 211-14 culture was grown in TAP medium with KNO₃ replacing NH₄Cl as nitrogen source and with modified trace elements from (Kropat et al., 2011) instead of Hutner's trace elements. Cells were grown at 90 μmol photons m⁻² s⁻¹,

140 rpm to a cell density of 2 × 10⁶ cells per milliliter for all experiments.

2.2 | Cross-reactivity and specificity of antibodies used for ChIP

All antibodies used in this study were evaluated on total cell lysate of *Chromochloris* SAG 211-14 cultures; 2 × 10⁷ cells were collected by centrifugation at 4°C, 1650 g and 2 min. The cell pellet was resuspended in 500-μl MilliQ water and 2× protein lysis buffer was added (125-mM Tris-Cl, pH 6.8, 20% glycerol, 4% SDS, 10% β-mercaptoethanol and 0.005% bromophenol blue). Samples were sonicated for 5 s to break the cell wall (Sonic Dismembrator System with a 1/2 in. probe, Model 505; 117 V, 50/60 Hz [Fisher Scientific Company, No. FB505110], settings: 1 s ON/1 s OFF, amplitude 50%). Twenty-microliter cell lysate corresponding to 4 × 10⁵ cells was loaded on each lane. Proteins were separated by 15% SDS-PAGE and transferred to nitrocellulose membranes (Amersham Protran 0.1 NC). The membrane was blocked for 30 min with 3% dried nonfat milk in phosphate-buffered saline (PBS) solution containing 0.1% (w/v) Tween 20 and then incubated in primary antiserum. The PBS solution was used as the diluent for both primary and secondary antibodies. The membranes were washed in PBS containing 0.1% (w/v) Tween 20. Antibodies directed against histone H3 (1:1000) or trimethylated lysine 4 of histone H3 (1:1000) were used. The secondary antibody, used at 1:4000, was goat antirabbit conjugated to alkaline phosphatase and processed according to the manufacturer's instructions.

2.3 | Establishment of DNA shearing

A culture volume corresponding to 2 × 10⁸ was collected by centrifugation at 4°C, 1650 g for 2 min. Cell pellet was resuspended in 1-ml ChIP lysis buffer (1% SDS, 10-mM EDTA, 50-mM Tris-Cl, pH 8.0, and 0.25× protease inhibitor cocktail [Roche]) and transferred to Beckman Coulter 4-ml Polycarbonate Thick Wall Centrifuge Tubes (13 × 64 mm). For shearing of genomic DNA, we used a Sonic Dismembrator System with a 1/2 in. probe, Model 505; 117 V, 50/60 Hz (Fisher Scientific Company, No. FB505110), settings: 1 s ON/1 s OFF, amplitude 50%). To test conditions needed for efficient fragmentation of *Chromochloris* genomic DNA, we used 2, 6, and 10 s of sonication. After sonication, 100 μl of each cell lysate (corresponding to 2 × 10⁷ cells) was transferred to a new 1.5-ml Eppendorf tube. Three hundred microliters of ChIP lysis buffer was added. DNA was extracted using phenol/chloroform extraction as follows: 500 μl of phenol/chloroform/isoamyl alcohol (25:24:1) was added to the cell lysates, and samples were mixed and centrifuged for 10 min at 4°C and 16,200 g. The aqueous phase was transferred to 500-μl chloroform/isoamyl alcohol (24:1). Samples were mixed and centrifuged for 10 min at 4°C and 16,200 g, and this step was repeated once. After centrifugation, the aqueous phase containing the DNA was added to a new Eppendorf tube containing 50 μl of 3-M Na acetate (pH 5.3), and 1 ml of 100%



ethanol was added. Samples were mixed and kept at -20°C for at least 2 h. DNA was pelleted by a 15-min centrifugation at 4°C at 16,200 g. DNA pellets were washed once with 70% ethanol, and the DNA pellet was air dried for at least 15 min at room temperature. DNA was resuspended in 50- μl TE supplemented with 10- $\mu\text{g}/\mu\text{l}$ DNase-free RNase A. Ten microliters of each sample was supplemented with $6\times$ DNA loading dye and loaded on a 1.2% agarose gel supplemented with SYBR gold.

2.4 | Optimizing formaldehyde cross-linking

A culture volume corresponding to 2×10^8 was collected by centrifugation at 4°C , 1650 g for 2 min. Cell pellets were washed with 50-ml KH buffer (20-mM K-HEPES, pH 7.4, 40-mM KCl). This is necessary because residual Tris from the growth medium would quench formaldehyde and lead to irreproducible results. After washing, cells were collected by centrifugation for 2 min at 4°C and 1650 g. Cell pellets were resuspended in 10-ml cross-linking buffer (20-mM K-HEPES, 40-mM KCl supplemented with no or 0.1%, 0.3%, 0.7%, and 1% formaldehyde). This solution was always made fresh because formaldehyde is unstable. Optimal formaldehyde concentration for cross-linking DNA to chromatin in *Chromochloris* was determined based on efficient ChIP using an antibody cross-reactive against an unmodified N-terminus of histone H3. We used the promoter region of *RBCS* as target genomic region during optimization. qPCR was performed on a Bio-Rad CFX96 Touch Real-Time PCR Detection System using iTAQ Mastermix according to the manufacturer's instructions. Primers used were as follows: *RBCS*promfor: CAATGCAAGCAGTTCGCATG and *RBCS*promrev: ACGGAGGACTTGGAATGAC.

2.5 | Optimized ChIP protocol

A culture volume corresponding to 2×10^8 cells was collected by centrifugation at 4°C , 1650 g for 2 min. Cell pellets were washed with 50-ml KH buffer. After washing, cells were collected by centrifugation for 2 min at 4°C and 1650 g. Cell pellets were resuspended in 10-ml cross-linking buffer (20-mM HEPES-KOH, 40-mM KCl supplemented with 0.7% formaldehyde). The cell pellet was resuspended in 1-ml ChIP lysis buffer (1% SDS, 10-mM EDTA, 50-mM Tris-Cl, pH 8.0, and $0.25\times$ protease inhibitor cocktail [Roche]) and transferred to Beckman Coulter 4-ml Polycarbonate Thick Wall Centrifuge Tubes (13 \times 64 mm). For shearing of genomic DNA, we used a Sonic Dismembrator System with a 1/2 in. probe, Model 505; 117 V, 50/60 Hz (Fisher Scientific Company, No. FB505110) at settings: 1 s ON/1 s OFF, amplitude 50%. After sonication for 10 s, cell lysate was centrifuged for 10 min at 4°C and 16,200 g (this is to clear the lysate from starch and cell debris). We generated aliquots of 100- μl input chromatin in 1.5-ml tubes and flash froze them in liquid nitrogen (each aliquot corresponds to chromatin from $\sim 2 \times 10^7$ cells). The input chromatin can be stored for several months at -80°C , but avoid multiple freeze/thaw cycles. We used 50 mg of Protein-A-Sepharose beads (Sigma,

P3391-1G), resuspended beads in 1-ml ChIP buffer (1.1% Triton X-100, 1.2-mM EDTA, 167-mM NaCl, and 16.7-mM Tris-HCl, pH 8) and incubated them for 30 min at 4°C . Swollen beads were washed two times with 500- μl ChIP buffer. After the last wash, supernatant was discarded, and we added 500- μl ChIP buffer (gives ~ 750 - μl suspension with swollen beads). We did not add sonicated lambda DNA, because this would interfere with DNA library generation. For ChIP, 5 μl of anti-H3 or 10 μl of anti-H3K4me3 antibodies was premixed with 10- μl BSA solution (10 mg/ml) and incubated for at least 30 min on ice. For each antibody employed, two 100- μl aliquots of chromatin solution were thawed on ice (one for a control without antibody and one for the antibody to be tested). Nine hundred microliters of ChIP buffer was added to each aliquot (from this step on, we consequently used filter tips to avoid DNA contaminations). Samples were centrifuged for 20 s at 16,100 g and 4°C , and supernatant was transferred to microcentrifuge tubes containing the prepared antibody solutions. Samples were mixed on a rotation wheel for 1 h at 4°C . After 1 h, samples were centrifuged for 20 s at 16,200 g and 4°C , and supernatants were transferred to new microcentrifuge tubes containing 60- μl Sepharose beads. Samples were mixed on a rotation wheel for 2 h at 4°C . After incubation, samples were centrifuged for 20 s at 16,200 g and 4°C and supernatants discarded. Sepharose beads were washed once with washing buffer 1 (0.1% SDS, 1% Triton X-100, and 2-mM EDTA, pH 8) containing 150-mM NaCl, once with washing buffer 1 containing 500-mM NaCl, once with washing buffer 2 (250-mM LiCl, 1% Nonidet P-40, 1% Na deoxycholate, 1-mM EDTA, and 10-mM Tris-HCl, pH 8), and twice with TE (1-mM EDTA and 10-mM Tris-HCl, pH 8). Cross-links were reverted by an overnight incubation at 65°C after addition of NaCl to a final concentration of 0.5 M, 1% SDS, and 0.1-M NaHCO_3 (made fresh) for 15 min at 65°C . This step was repeated once, and eluates were pooled. At this point, another aliquot of input chromatin was also thawed, and 400 μl of ChIP lysis buffer was added and incubated overnight at 65°C (input DNA control). In order to remove proteins in the precipitates, we did add 10 μl of 0.5-M EDTA (pH 8.0), 20 μl of 1-M Tris-HCl (pH 8.0), and 2.1 μl of proteinase K (10 mg/ml) and incubated the samples for 1 h at 55°C . DNA extraction was performed by one extraction with 500- μl phenol/chloroform/isoamyl alcohol and one extraction using 500- μl chloroform/isoamyl alcohol. DNA was precipitated by adding 50 μl of 3-M Na acetate (pH 5.2), 2.5 μl of glycogen (2.5 $\mu\text{g}/\mu\text{l}$) (this will give an otherwise invisible pellet), and 1 ml of 100% EtOH and incubating for at least 2 h at -20°C . Samples were centrifuged for 20 min at 16,200 g and 4°C . DNA pellet was dried and resuspended in 50 μl of 10-mM Tris-HCl buffer (pH 8.0).

2.6 | ChIP library preparation and sequencing

Five nanogram of ChIP'ed DNA was treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT) using the KAPA-HyperPrep kit (KAPA Biosystems). The ligated products were enriched with 8 cycles of PCR (KAPA Biosystems) and size selected to 200–500 bp with Total Pure NGS beads (Omega Bio-Tek). The prepared

libraries were quantified using KAPA Biosystems' next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Sequencing of the flow cell was performed on the Illumina NextSeq500 sequencer using NextSeq500 NextSeq HO kits, v2, following a 2×151 indexed run recipe.

2.7 | ChIP read preprocessing and filtering

2×151 sequence data were generated at the DOE Joint Genome Institute (JGI) using the Illumina NextSeq platform. BBDuk version 38.87 (<https://bbtools.jgi.doe.gov>) was used to remove contaminants, trim reads that contained adapter sequence and homopolymers of G's of size 5 or more at the ends of the reads and right quality trim reads where quality drops below 6. BBDuk was used to remove reads that contained 1 or more "N" bases and had an average quality score across the read less than 10 or had minimum length ≤ 49 bp. Reads mapped with BBDuk (v. 38.87) to masked human, cat, dog, mouse, and common microbial contaminant references at 93% identity were removed.

2.8 | ChIP read alignment and peak calling

Filtered reads were aligned to the *Chromochloris zofingiensis* reference genome (Roth et al., 2017) using BWA mem (version 0.7.17). Only uniquely mapped reads were retained. Duplicated reads were removed by Picard (v. 2.22.9) (<http://broadinstitute.github.io/picard/>; Broad Institute) MarkDuplicates tool. Finally, the remaining reads were used for peak calling by MACS2 (v. 2.1.1) (Zhang et al., 2008) with parameters "--call-summits --nomodel --extsize 147 -c." Input control libraries were generated and used for peak calling and downstream analysis. To visualize and plot data, bigwig files were created using bedGraphToBigWig (v. 4) (Kent et al., 2010) and Deeptools (v. 3.1.3) (Ramirez et al., 2016) was used to generate summary signal plots and heatmaps.

2.9 | RNA extraction, sequencing, and transcriptome analyses

A culture volume corresponding to 5×10^7 cells was collected by centrifugation at 4°C , 3500 rpm for 2 min. Cell pellet was resuspended in 0.2-ml RLC buffer (Qiagen), flash frozen in liquid nitrogen, and ground to a fine powder using mortar and pestle. Sample was added to 700- μl TRIzol and mixed overhead before the addition of 200- μl chloroform/isoamyl alcohol. Samples were vigorously shaken, then centrifuged for 10 min at 4°C and 13,200 rpm. Supernatant was added to 700- μl isopropanol. RNA was precipitated at -20°C overnight and washed with 70% ethanol, air dried, and resuspended in 40- μl RNase-free water. DNase I digest and cleanup was performed using Zymo RNA Clean & Concentrator Kit (RCC) according to the manufacturer's instructions.

RNA was converted into cDNA and made into sequence ready libraries with the KAPA RNA-Seq Kit (KAPA Biosystems). RNA-Seq libraries were sequenced with 150-bp single-end reads on a HiSeq 2500. Data were aligned to the ChrZofV5 release of the *C. zofingiensis* genome with RNA STAR. Determination of counts per gene and transcript abundance in transcripts per million (TPMs) were made with DESeq2.

2.10 | Data availability

Data are available from the US NCBI Short Read Archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra>) using the following accession numbers: SRP354587, SRP354588, SRP354586 (ChIP-Seq data) and SRP355099, SRP355100 (RNA-Seq data).

3 | RESULTS

3.1 | An optimized ChIP-Seq framework for Chromochloris—An overview of the most critical parameters

ChIP involves cross-linking of the chromatin-bound proteins by formaldehyde, followed by sonication to obtain small DNA fragments. Immunoprecipitation of cross-linked, fragmented material is then carried out using specific antibodies against the DNA-binding protein of interest. As pointed out by us and others, some parameters are crucial for a successful ChIP experiment and need to be individually adjusted for each respective organism (Das et al., 2004). Accordingly, the optimization of the most critical steps for a ChIP protocol in *Chromochloris* is outlined in more detail below.

In order to be able to identify the location of the histone-bound DNA sequence of interest with proper resolution, it is critical to break the DNA to an average fragment size of around 250 bp (approximate average nucleosome spacing). The hydrodynamic shearing method using sonication is commonly used for this purpose (Orlando et al., 1997; Spencer et al., 2003). A total of three sonication times were tested on *Chromochloris* samples: 2, 6, and 10×10 s. While intact genomic DNA was still visible after a sonication time of 2×10 s, sonication for 10×10 s yielded the desired average DNA fragment size of ~ 250 (Figure 1a).

A second important factor for a successful ChIP experiment is the choice of the antibody against the protein of interest. Commercially available antibodies are usually raised against human (modified) histones and are most likely cross-reactive with other species, as histones are highly conserved proteins. Unsurprisingly, *Chromochloris* histone H3 shares high sequence similarity with human histone H3, making it likely that commercial antibodies will be functional in our applications (Figure S1). We aimed to use two different antibodies in this study, one cross-reactive against unmodified histone H3 (for optimization purposes) and the other one against trimethylated H3K4. Cross-reactivity and specificity of both antibodies was tested by

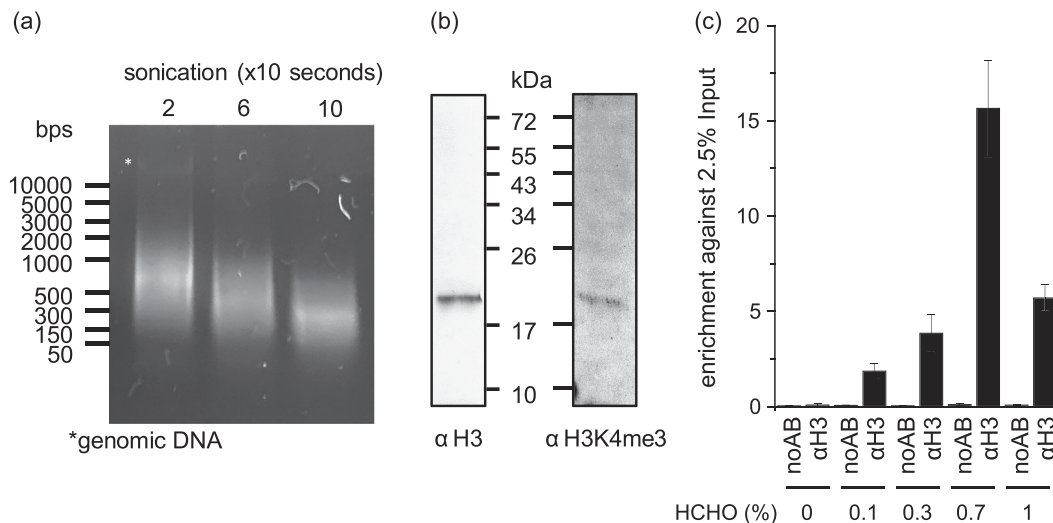


FIGURE 1 Antibody test, optimization of sonication conditions to yield ~250-bp chromatin fragments and concentrations for efficient cross-linking. (a) 2×10^7 cells of *Chromochloris* (*Chr.*) *zofingiensis* were sonicated 2, 6, and 20 times for 10 s. DNA was extracted with phenol/chloroform/isoamyl alcohol (25/24/1 v/v), separated on a 1.5% agarose gel and stained with SYBR gold. (b) Whole-cell proteins corresponding to 2×10^5 cells from *Chr. zofingiensis* were separated on an SDS-polyacrylamide (15% monomer) gel and analyzed by immunodetection using antibodies against histone H3 or histone H3K4me3 as indicated. (c) 2×10^7 cells were incubated for 10 min with formaldehyde (HCHO) at the concentrations indicated. The reaction was quenched by the addition of 125-mM glycine. Antibodies against histone H3 were used for ChIP (with no antibody as a background control). Chromatin-immunoprecipitated DNA was extracted and amplified using primers targeting the *RBCS* promoter region. Shown are averages and StDEV of three amplification reactions

immunoblotting using *Chromochloris* cell lysates for both of which we saw a single, clear band at the expected molecular weight (Figure 1b).

Cross-linking of the starting material prior to ChIP is used to ensure that the chromatin structure is preserved during sample collection and during the remainder of the ChIP procedure (Solomon et al., 1988; Solomon & Varshavsky, 1985). Formaldehyde is the most commonly used cross-linking agent in ChIP experiments as it penetrates most cell walls and because cross-links are reversible by incubation at higher temperatures; commonly 65°C is used for the purpose of decross-linking (Orlando et al., 1997). In addition, formaldehyde forms bonds that span a distance of approximately 2 Å (Dedon et al., 1991), resulting in DNA-protein, RNA-protein, and protein-protein cross-linking. While one needs sufficient amounts of formaldehyde for efficient DNA-protein cross-linking, it is also crucial not to overcross-link the sample to avoid the loss of antibody recognition sites and thus loss of ChIP'd DNA. Many published ChIP protocols use a standard amount of 1% formaldehyde, but we have seen in the past that different concentrations can give better signal to noise ratios in ChIP experiments of algae (Strenkert et al., 2011). For the determination of the optimal cross-linker concentrations, ChIP was performed using an antibody targeting unmodified histone H3 (as opposed to a certain histone mark) as we expected reasonable nucleosome occupancy at the whole-genome level, including at a promoter region of the gene encoding the small subunit of Rubisco, *RBCS*. We tested formaldehyde at concentrations of 0%, 0.1%, 0.3%, 0.7%, and 1 percent. Samples without an antibody (noAB) were used as negative controls, which allowed us to estimate DNA enrichment over background contamination if histone-DNA cross-linking was successful. While all

formaldehyde concentrations tested led to significant enrichment of ChIP'd DNA over background (Figure 2c), the use of 0.7% formaldehyde yielded the largest amounts of ChIP'd *RBCS* promoter DNA and the highest signal to noise (Figure 2c). Lower formaldehyde concentrations gave insufficient cross-linking, and higher concentrations resulted in overcross-linking, both of which reduced ChIP efficiency. Therefore, we recommend the use of 0.7% formaldehyde.

3.2 | A survey of the genome-wide H3K4me3 distribution pattern in *Chromochloris*

To further test suitability of our optimized ChIP framework for deep sequencing applications, we determined H3K4me3 enrichment at a genome-wide level. H3K4me3 is a so-called narrow-peak histone mark that seems to be specifically enriched at TSSs of protein coding genes in a variety of different organisms including yeast, mammals, and land plants (Barski et al., 2007; Bernstein et al., 2005; Pokholok et al., 2005; Santos-Rosa et al., 2002; Zhang et al., 2009). Accordingly, assessment of data quality will be straightforward. In our ChIP experiments, one possible negative control, ChIPs performed without an antibody (noAB controls) did not yield sufficient amounts of DNA for library preparation and subsequent deep sequencing. It is certainly possible to identify DNA-binding sites based on relative enrichment of different chromosomal regions even without a negative control; however, some chromosomal regions can be significantly enriched even when sequencing genomic DNA alone (Rozowsky et al., 2009). Accordingly, non-ChIP'd, genomic DNA is often analyzed in parallel

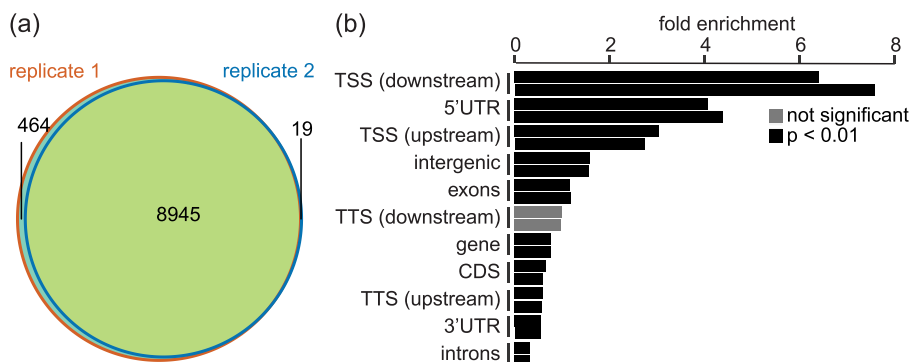


FIGURE 2 H3K4me3 peak calling and genomic distribution in *Chromochloris*. (a) Total number of narrow H3K4me3 peaks in each replicate and the overlap between both replicates. (b) Relative fold enrichment of H3K4me3 at different genomic regions as indicated. Each individual sample represents one independent replicate. The fold enrichment represents the number of peaks that overlap with a region divided by the number of peaks that would overlap with the respective region if they were randomly distributed across the genome

TABLE 1 H3K4me3 ChIP-Seq mapping statistics

	Total reads	Mapped reads	Mapped reads (%)	Unique deduplicate reads
H3K4me3 ChIP Rep1	137,855,434	136,896,574	99.3	76,945,677
H3K4me3 ChIP Rep2	125,105,126	123,333,772	98.6	21,906,729
Input DNA	188,843,106	186,436,803	98.8	163,415,936

as negative control for ChIP-Seq experiments to account for this issue. To this end, we included non-ChIP'ed genomic DNA as our control. We interrogated H3K4 trimethylation from duplicate samples in *Chromochloris* cultures. mRNA from the same cultures was obtained in parallel and used for RNA-Seq analysis, with the intention to later analyze it alongside the ChIP-Seq data. By aligning H3K4me3 ChIP-Seq reads to the reference genome, we noted that the majority of reads (>98%) could be mapped to the nuclear genome, which suggested that the sequencing data were of a high quality (Table 1). Model-based analysis using ChIP-Seq (MACS) software (Zhang et al., 2008) was used for peak calling. We identified a total of 8945 H3K4me3 consensus peaks in both replicates, with the duplicate samples showing excellent overlap between each other (Figure 2a).

The distribution of H3K4me3 peaks along the *Chromochloris* genome was further characterized using 11 groups that included distinct genic regions and observed predominant H3K4me3 enrichment at TSSs of genes (Figure 2b). Genome-wide profiling of H3K4me3 revealed an average H3K4me3 peak width of ± 500 bp around the TSSs in the majority of genes (Figure 3).

3.3 | H3K4me3 and gene expression

To get an integrated view on H3K4me3 enrichment and its relationship to gene expression, we compared transcript abundances of individual genes with H3K4me3 enrichment at their respective TSSs. An increase in transcript abundance seems to correlate with the chances of a gene harboring a H3K4me3 at its TSS (Figure 4). When we

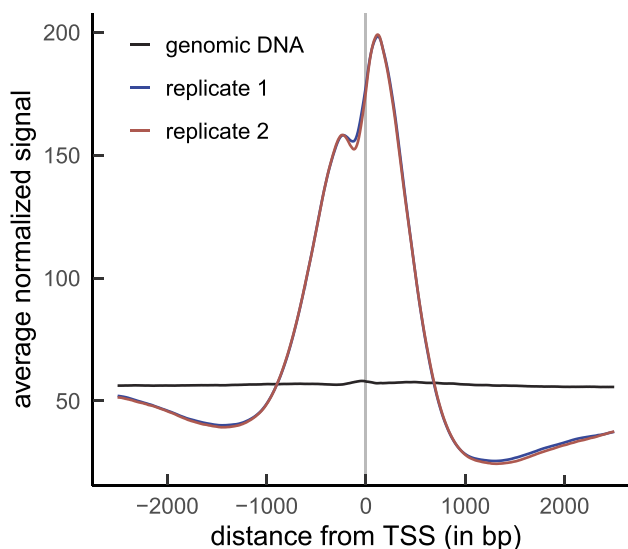


FIGURE 3 Genome-wide distribution pattern of H3K4me3 in *Chromochloris*. Shown is the average normalized H3K4me3 enrichment for both replicates (blue and red) and input genomic DNA (black) at the TSSs (gray line) and ± 2500 bps from TSSs. Values were normalized by scaling based on the number of uniquely mapped, nonduplicate reads

further interrogated transcript abundances of genes that harbor a H3K4me3 peak at their TSS and compared it with those that do not, again, we saw a clear, positive correlation between active transcription and the likelihood of a gene being marked with H3K4me3 (Figure 5).

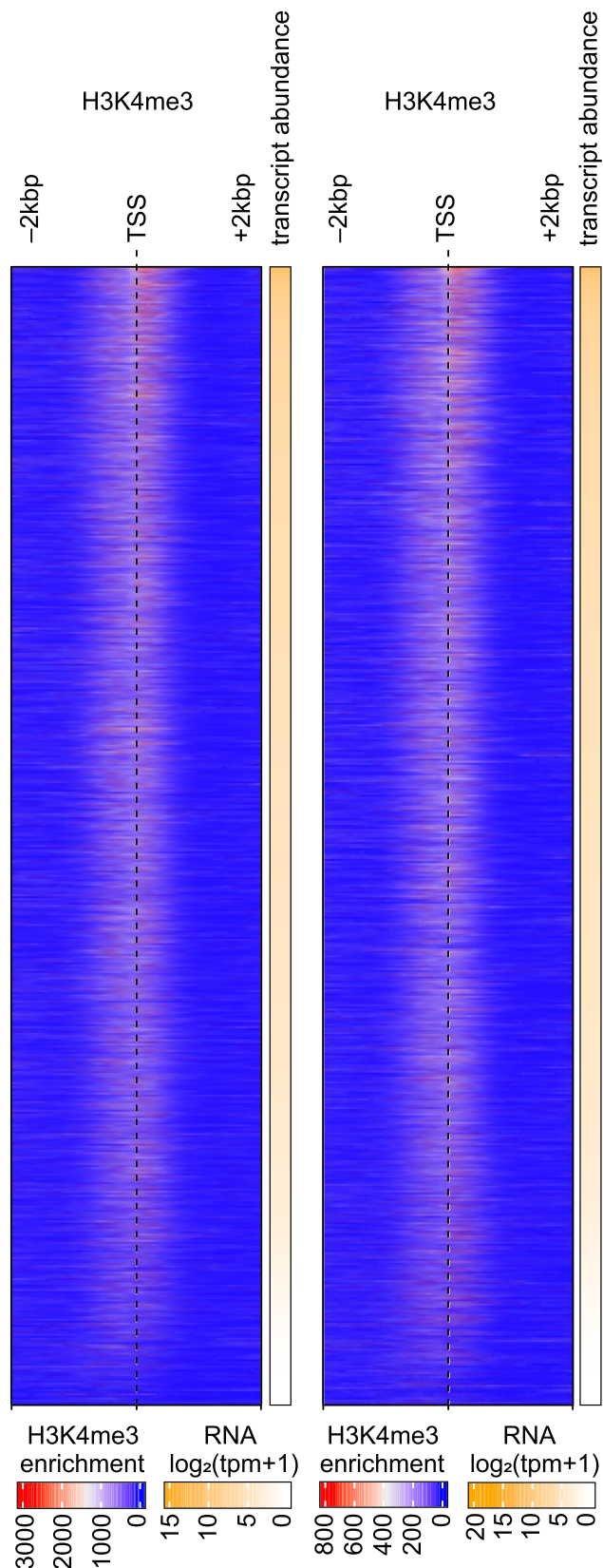


FIGURE 4 H3K4me3 enrichment ranked by gene expression. Shown are H3K4me3 values visualized as heat map. H3K4me3 values are sorted based on the respective expression values (as transcripts per million [TPMs]) for both replicates at the TSSs (dashed line) and ± 2000 bp from TSSs

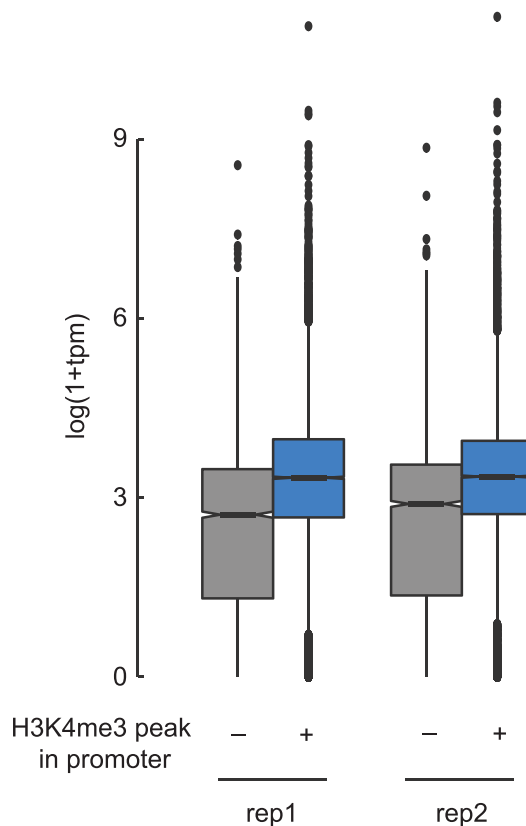


FIGURE 5 Integrated analyses of H3K4me3 enrichment and gene expression. Transcript abundance values (in transcripts per million [TPMs]) from genes that harbor a H3K4me3 peak at their TSSs (blue) versus all others (gray) are visualized using boxplots

4 | DISCUSSION

Epigenetic regulation is a prerequisite for mediating transcriptional responses in general and for the fine tuning of genetically programmed changes in particular. ChIP-Seq is a powerful tool to assess changes in histone mark patterns and to identify transcription factor-binding sites. Most importantly, some critical parameters in the procedure need to be optimized in each organism individually to ensure proper data quality and reproducibility between samples.

Given the conserved, genome-wide distribution pattern of H3K4me3 in many diverse taxa, ChIP-Seq using H3K4me3 was facilitated as a proof of principle application to illustrate desired data quality of our optimized ChIP-Seq framework. H3K4me3 was shown to be a narrow-peak histone mark that is enriched at TSSs of genes in yeast, fly, human, algae, and plants (Barski et al., 2007; Bernstein et al., 2005; Pokholok et al., 2005; Santos-Rosa et al., 2002; Zhang et al., 2009). Unsurprisingly, and in line with a conserved role for H3K4me3 in gene activation, *Chromochloris* genes are also marked with H3K4me3 at their TSSs. Notably, the likelihood of a gene being marked by H3K4me3 was higher, if the underlying gene was transcribed.

The optimized ChIP-Seq framework established herein will pave the way for studying epigenetic changes and transcription factor-

binding sites in the emerging reference alga *Chromochloris*. Moreover, the coverage achieved by our ChIP-Seq frame work will be sufficient for quantitative recovery of narrow- and broad-peak histone marks alike. This work represents a critical first step in establishing chromatin state dynamics in *Chromochloris*.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under Award Number DE-SC0018301. We also thank Kris K. Niyogi for critical reading of the manuscript and kindly providing the histone H3 antibody. This research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative and used resources at the DOE Joint Genome Institute which is a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and operated under Contract No. DE-AC02-05CH11231.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

ORCID

Daniela Strenkert  <https://orcid.org/0000-0002-3420-1332>
 Matthew Mingay  <https://orcid.org/0000-0001-7917-1432>
 Stefan Schmollinger  <https://orcid.org/0000-0002-7487-8014>
 Cindy Chen  <https://orcid.org/0000-0002-7539-103X>
 Ronan C. O'Malley  <https://orcid.org/0000-0002-4907-0339>
 Sabeeha S. Merchant  <https://orcid.org/0000-0002-2594-509X>

REFERENCES

- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–837. <https://doi.org/10.1016/j.cell.2007.05.009>
- Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., & Schreiber, S. L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences*, *99*(13), 8695–8700. <https://doi.org/10.1073/pnas.082249499>
- Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J. III, Gingeras, T. R., Schreiber, S. L., & Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, *120*, 169–181. <https://doi.org/10.1016/j.cell.2005.01.001>
- Bowler, C., Benvenuto, G., Laflamme, P., Molino, D., Probst, A. V., Tariq, M., & Paszkowski, J. (2004). Chromatin techniques for plant cells. *The Plant Journal*, *39*(5), 776–789. <https://doi.org/10.1111/j.1365-313X.2004.02169.x>
- Das, P. M., Ramachandran, K., VanWert, J., & Singal, R. (2004). Chromatin immunoprecipitation assay. *BioTechniques*, *37*(6), 961–969. <https://doi.org/10.2144/04376RV01>
- Dedon, P. C., Soultis, J. A., Allis, C. D., & Gorovsky, M. A. (1991). A simplified formaldehyde fixation and immunoprecipitation technique for studying protein-DNA interactions. *Analytical Biochemistry*, *197*, 83–90. [https://doi.org/10.1016/0003-2697\(91\)90359-2](https://doi.org/10.1016/0003-2697(91)90359-2)
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, *130*, 77–88. <https://doi.org/10.1016/j.cell.2007.05.042>
- Haring, M., Offermann, S., Danker, T., Horst, I., Peterhänsel, C., & Stam, M. (2007). Chromatin immunoprecipitation: Optimization, quantitative analysis and data normalization. *Plant Methods*, *3*, 11. <https://doi.org/10.1186/1746-4811-3-11>
- He, G. M., Zhu, X. P., Elling, A. A., Chen, L. B., Wang, X. F., Guo, L., Liang, M. Z., He, H., Zhang, H. Y., Chen, F. F., Qi, Y. J., Chen, R. S., & Deng, X. W. (2010). Global epigenetic and transcriptional trends among two Rice subspecies and their reciprocal hybrids. *The Plant Cell*, *22*, 17–33. <https://doi.org/10.1105/tpc.109.072041>
- Hecht, A., & Grunstein, M. (1999). Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction. *Methods in Enzymology*, *304*, 399–414. [https://doi.org/10.1016/S0076-6879\(99\)04024-0](https://doi.org/10.1016/S0076-6879(99)04024-0)
- Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., & Karolchik, D. (2010). BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics*, *26*(17), 2204–2207. <https://doi.org/10.1093/bioinformatics/btq351>
- Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S. S., & Malasarn, D. (2011). A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *The Plant Journal*, *66*(5), 770–780. <https://doi.org/10.1111/j.1365-313X.2011.04537.x>
- Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, *128*(4), 707–719. <https://doi.org/10.1016/j.cell.2007.01.015>
- Luger, K. (2003). Structure and dynamic behavior of nucleosomes. *Current Opinion in Genetics & Development*, *13*, 127–135. [https://doi.org/10.1016/S0959-437X\(03\)00026-1](https://doi.org/10.1016/S0959-437X(03)00026-1)
- Martin, C., & Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nature Reviews. Molecular Cell Biology*, *6*(11), 838–849. <https://doi.org/10.1038/nrm1761>
- Ng, H. H., Robert, F., Young, R. A., & Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating pol II provides a localized mark and memory of recent transcriptional activity. *Molecular Cell*, *11*(3), 709–719. [https://doi.org/10.1016/S1097-2765\(03\)00092-3](https://doi.org/10.1016/S1097-2765(03)00092-3)
- Oh, S., Park, S., & Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genetics*, *4*(8), e1000077. <https://doi.org/10.1371/journal.pgen.1000077>
- Orlando, V., Strutt, H., & Paro, R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods*, *11*, 205–214. <https://doi.org/10.1006/meth.1996.0407>
- Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., & Young, R. A. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, *122*(4), 517–527. <https://doi.org/10.1016/j.cell.2005.06.026>
- Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dundar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*, W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Ricardi, M. M., González, R. M., & Iusem, N. D. (2010). Protocol: Fine-tuning of a chromatin immunoprecipitation (ChIP) protocol in tomato. *Plant Methods*, *6*, 1746–4811. <https://doi.org/10.1186/1746-4811-6-11>
- Roth, M. S., Cokus, S. J., Gallaher, S. D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C. A., Merchant, S. S., Pellegrini, M., & Niyogi, K. K. (2017). Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *P Natl Acad Sci USA*, *114*(21), E4296–E4305. <https://doi.org/10.1073/pnas.1619928114>



- Rozowsky, J., Euskirchen, G., Auerbach, R. K., Zhang, Z. D. D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., & Gerstein, M. B. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotechnology*, 27(1), 66–75. <https://doi.org/10.1038/nbt.1518>
- Ruthenburg, A. J., Allis, C. D., & Wysocka, J. (2007). Methylation of lysine 4 on histone H3: Intricacy of writing and reading a single epigenetic mark. *Molecular Cell*, 25, 15–30. <https://doi.org/10.1016/j.molcel.2006.12.014>
- Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J., & Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*, 419(6905), 407–411. <https://doi.org/10.1038/nature01080>
- Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature Cell Biology*, 6(1), 73–77. <https://doi.org/10.1038/ncb1076>
- Schübeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., Bell, S. P., & Groudine, M. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes & Development*, 18(11), 1263–1271. <https://doi.org/10.1101/gad.1198204>
- Sequeira-Mendes, J., Araguez, I., Peiro, R., Mendez-Giraldez, R., Zhang, X. Y., Jacobsen, S. E., Bastolla, U., & Gutierrez, C. (2014). The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *The Plant Cell*, 26(6), 2351–2366. <https://doi.org/10.1105/tpc.114.124578>
- Solomon, M. J., Larsen, P. L., & Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell*, 53(6), 937–947. [https://doi.org/10.1016/S0092-8674\(88\)90469-2](https://doi.org/10.1016/S0092-8674(88)90469-2)
- Solomon, M. J., & Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures. *Proceedings of the National Academy of Sciences*, 82, 6470–6474. <https://doi.org/10.1073/pnas.82.19.6470>
- Spencer, V. A., Sun, J. M., Li, L., & Davie, J. R. (2003). Chromatin immunoprecipitation: A tool for studying histone acetylation and transcription factor binding. *Methods*, 31(1), 67–75. [https://doi.org/10.1016/S1046-2023\(03\)00089-6](https://doi.org/10.1016/S1046-2023(03)00089-6)
- Strenkert, D., Schmollinger, S., & Schroda, M. (2011). Protocol: Methodology for chromatin immunoprecipitation (ChIP) in *Chlamydomonas reinhardtii*. *Plant Methods*, 7(1), 35. <https://doi.org/10.1186/1746-4811-7-35>
- Wang, X. F., Elling, A. A., Li, X. Y., Li, N., Peng, Z. Y., He, G. M., Sun, H., Qi, Y. J., Liu, X. S., & Deng, X. W. (2009). Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *The Plant Cell*, 21(4), 1053–1069. <https://doi.org/10.1105/tpc.109.065714>
- Workman, J. L., & Kingston, R. E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annual Review of Biochemistry*, 67, 545–579. <https://doi.org/10.1146/annurev.biochem.67.1.545>
- Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., & Jacobsen, S. E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biology*, 10(6), R62. <https://doi.org/10.1186/gb-2009-10-6-r62>
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9, R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
- Zong, W., Zhong, X. C., You, J., & Xiong, L. Z. (2013). Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. *Plant Molecular Biology*, 81(1–2), 175–188. <https://doi.org/10.1007/s11103-012-9990-2>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Strenkert, D., Mingay, M., Schmollinger, S., Chen, C., O'Malley, R. C., & Merchant, S. S. (2022). An optimized ChIP-Seq framework for profiling histone modifications in *Chromochloris zofingiensis*. *Plant Direct*, 6(3), e392. <https://doi.org/10.1002/pld3.392>