Video Article

Chromatin Immunoprecipitation of Murine Brown Adipose Tissue

Maria Dafne Cardamone[1], Joseph Orofino[1], Adam Labadorf[2], Valentina Perissi[1]

[1]Biochemistry Department, Boston University School of Medicine

[2]Bioinformatic Hub, Boston University

Correspondence to: Valentina Perissi at vperissi@bu.edu

URL: https://www.jove.com/video/58682
DOI: doi:10.3791/58682

Keywords: Genetics, Issue 141, Chromatin immunoprecipitation, brown adipose tissue, next generation sequencing, mouse, fat

Date Published: 11/21/2018

Citation: Cardamone, M.D., Orofino, J., Labadorf, A., Perissi, V. Chromatin Immunoprecipitation of Murine Brown Adipose Tissue. *J. Vis. Exp.* (141), e58682, doi:10.3791/58682 (2018).

Abstract

Most cellular processes are regulated by transcriptional modulation of specific gene programs. Such modulation is achieved through the combined actions of a wide range of transcription factors (TFs) and cofactors mediating transcriptional activation or repression *via* changes in chromatin structure. Chromatin immunoprecipitation (ChIP) is a useful molecular biology approach for mapping histone modifications and profiling transcription factors/cofactors binding to DNA, thus providing a snapshot of the dynamic nuclear changes occurring during different biological processes.

To study transcriptional regulation in adipose tissue, samples derived from *in vitro* cell cultures of immortalized or primary cell lines are often favored in ChIP assays because of the abundance of starting material and reduced biological variability. However, these models represent a limited snapshot of the actual chromatin state in living organisms. Thus, there is a critical need for optimized protocols to perform ChIP on adipose tissue samples derived from animal models.

Here we describe a protocol for efficient ChIP-seq of both histone modifications and non-histone proteins in brown adipose tissue (BAT) isolated from a mouse. The protocol is optimized for investigating genome-wide localization of proteins of interest and epigenetic markers in the BAT, which is a morphologically and physiologically distinct tissue amongst fat depots.

Video Link

The video component of this article can be found at https://www.jove.com/video/58682/

Introduction

While white adipose tissue (WAT) is specialized for energy storage, brown adipose tissue (BAT) dissipates energy in the form of heat due to its ability to convert carbohydrates and lipids into thermal energy *via* mitochondrial uncoupling[1]. Because of this specialized function, the BAT depot is required for maintenance of body temperature in physiological conditions and in response to cold exposure. While gene expression changes during BAT differentiation and upon thermogenic stress have been extensively studied in vivo and in vitro, the molecular mechanisms underlying these changes have been mostly dissected in immortalized cell lines and primary pre-adipocytes, with the exception of several in vivo studies[2,3,4,5].

Regulation of specific gene expression programs through transcriptional regulation is achieved by coordinated changes in chromatin structure *via* various transcription factors and co-factors actions. Chromatin immunoprecipitation (ChIP) is a valuable molecular biology approach for investigating the recruitment of these factors to DNA and for profiling the associated changes in the chromatin landscape. Key factors for the success of ChIP experiments include optimizations of crosslinking conditions and chromatin shearing consistency throughout different samples, availability of adequate starting material, and, most notably, quality of the antibodies. When performing ChIP from whole tissues, it is also important to consider heterogeneity of the samples and optimize the protocol to improve efficiency of nuclei isolation, with the latter being a particularly sensitive step when working with adipose tissue due to the elevated lipid content. In fact, molecular isolation techniques from whole adipose depots are complicated by the presence of high levels of triglycerides, and protocols must be optimized to increase the amount of chromatin isolation. Finally, when high-throughput sequencing is performed after ChIP-DNA isolation, the sequencing depth is critical for determining the number of peaks that are confidently detected.

Here, we refer to the working standards and general guidelines for ChIP-seq experiments recommended by the ENCODE and modENCODE consortia[6] for best practices, and we focus on a step-by-step description of a protocol optimized for ChIP-seq from BAT. The described protocol allows for efficient isolation of chromatin from adipose tissue to perform genome-wide sequencing for DNA-binding factors with well-defined peaks as well as histone marks with more diffuse signals.

Protocol

The animal handling steps of the protocol have been approved by Boston University's Institutional Animal Care and Use Committee (IACUC).

1. Day 1: Dissection and Preparation of BAT for Chromatin Immunoprecipitation (ChIP)

1. Euthanize mice using a carbon dioxide (CO_2) chamber, and perform the dissection immediately afterwards. Spray mouse fur with 70% ethanol before incision. Place the mouse with its back facing up, then cut open the skin along the neck. Locate the BAT directly under the skin between the shoulders (interscapular; it appears as two lobes, butterfly-shaped, with a thin layer of white fat that needs to be carefully removed[7]).
   NOTE: For a 3 month-old C57BL/6J mouse that is fed a regular chow diet (mouse weight at this age ranges between 27 and 30 g), each BAT lobe is approximately 1 cm long. The total weight of the BAT depot is approximately 0.15 g in both male and female mice.
2. Use 1 BAT pad for 3 ChIPs. Multiple BAT pads can be simultaneously processed until sonication.
3. Crosslink each BAT pad in 10 mL of phosphate-buffered saline (PBS) containing 1% formaldehyde for 15 min rotating at 150 rpm at room temperature (RT).
4. Quench crosslinking by adding 2.5 M glycine to a final concentration of 125 mM for 5 min, rotating at 150 rpm at RT.
5. Transfer the BAT pad to 500 µL of hypotonic lysis buffer (10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSF) and add 2 stainless steel beads (size: 3.2 mm diameter, see **Table of Materials**) for each BAT pad. Use a bead-based tissue homogenizer (see **Table of Materials**) to shred the tissue.
6. Start with 5 min at max power. If needed, extend the time until the tissue is homogenized and single cells are separated. Transfer the samples into a new tube and centrifuge at 4 °C at 16,000 x *g* for 10 min.
7. After centrifugation, transfer the supernatant into a new tube and keep the cell pellet on ice. Centrifuge the supernatant at 4 °C at 16,000 x g for 10 min to prevent loss of floating cells. Discard the final supernatant and resuspend both cell pellets together in 300 µL of SDS lysis buffer (SDS 1%; EDTA 10 mM; Tris-HCl pH 8, 50 mM) with 1x protease inhibitor (see **Table of Materials**). Incubate on ice for 10 min.
8. Sonicate the lysates to generate DNA fragments 200–500 base pairs long. After sonication, remove debris by centrifugation for 10 min at 16,000 x *g* at 4 °C and check the success of sonication *via* DNA electrophoresis on a 1% agarose gel. For a medium-size gel, run at 120 V; a good separation is achieved after 45 min.
   NOTE: Sonication optimization may be required to achieve proper size of fragments. With the chosen sonicator (see **Table of Materials**), this requires 2 times at 320 W output power and 20 KHz frequency for 10 min, using cycles of 30 s on/30 s off.
9. Dilute the supernatant fraction 10-fold (final volume is 3 mL for each BAT pad) in ChIP dilution buffer (SDS 0.01%; Triton 1.1%; EDTA 1.2 mM; Tris-HCl pH 8, 16.7 mM; NaCl 167 mM) with 1x protease inhibitor. Keep aside 1% of this chromatin solution (equal to 30 µL for 3 mL) as the ChIP input. Keep the inputs overnight at 4 °C (until step 2.4).
   Note: The above fraction will provide the input reference for relative quantification of immunoprecipitated material compared to the amount of DNA present in each sample prior to immunoprecipitation.
10. Add the ChIP antibody (see **Table of Materials** for the antibodies used here as an example) to 1 mL of chromatin solution and incubate overnight at 4 °C with rotation. Perform parallel immunoprecipitation with corresponding pre-immune IgG, if available, or normal IgG of the same species (*e.g.*, regular rabbit IgG if the antibody is produced in a rabbit). Alternatively, save 1 mL of chromatin solution for a no-antibody control.
    NOTE: The optimal amount of antibody should be determined experimentally for each new antibody, if the antibody concentration is not known. Otherwise, 2–5 µg of antibody is a reasonable starting amount to use. Use ChIP-grade antibodies whenever possible, or antibodies that have been validated for immunoprecipitation.

2. Day 2: Collection of the Immune Complexes

1. Add 50 µL of Protein A Agarose 50% slurry to the immune complexes and incubate for 1 h at 4 °C with rotation at 150 rpm.
2. **Pellet the beads by centrifugation for 1 min at 100 x *g* at 4 °C. Perform sequential washes of the beads with buffers of increasing stringency.**
   1. First, wash with low-salt wash buffer (150 mM NaCl; Tris-HCl pH 8, 20 mM; EDTA 2 mM; Triton-X 1%; SDS 0.1%), then with high-salt wash buffer (500 mM NaCl; Tris-HCl pH 8, 20 mM; EDTA 2 mM, Triton-X1%; SDS 0.1%), then with LiCl wash (0.25 M LiCl; Tris-HCl pH 8, 10 mM; EDTA 1 mM; Igepal 1%; deoxicholate 1%), and finally with 1x TE (Tris-HCl pH 8, 10 mM; EDTA).
   2. Perform all washes on 0.45 µm PVDF centrifugal filter columns (**Table of Materials**) by first transferring the beads in the columns using 500 µL of low-salt wash buffer. Pellet the beads in the columns for 3 min at 2,500 x *g*. Discard the flow-through and repeat 1 time for all the wash buffers listed in step 2.2.1.
3. After the washes are completed, elute the immune complexes by adding 300 µL of elution buffer 1% SDS, 0.1 M $NaHCO_3$) to the pelleted beads in the columns. Incubate at RT for 30 min on a shaker at 400 rpm. Collect the eluate by spinning down columns for 3 min at 2,500 x *g* in a fresh tube.
4. Reverse the crosslinks by incubating the eluate and inputs collected in step 1.9 at 65 °C overnight.

3. Day 3: Recovery of DNA with Phenol/Chloroform Extraction

1. Add 300 µL of phenol:chloroform:IAA, 1:1 to the eluate and inputs. Vortex for 10 s.
2. Spin at 4 °C for 15 min at 21,000 x *g*. Transfer the supernatant to a fresh tube. Add 3 µL of glycogen, 30 µL of 3 M sodium acetate, and 750 µL of cold 100% ethanol. Vortex for 10 s. Store at -80 °C for 2 h.
3. Spin down at 21,000 x g at 4 °C for 20 min, keep the pellet, and discard the supernatant. Wash pellet with 1 mL of cold 70% EtOH, then spin down at 21,000 x *g* at 4 °C for 5 min. Discard the supernatant and air-dry the pellet.

4. Resuspend the pellet in 70 µL of DNAse-free water.
5. Proceed to step 4 for testing chromatin occupancy on specific genomic regions by quantitative polymerase chain reaction (qPCR) or to step 5 to prepare samples for sequencing.

4. Analysis of ChIP using qPCR (Single/Multiple Genes Readout)

1. Dilute the input DNA from step 3.4 to generate a standard reference curve. Serial dilutions of 1/10, 1/100, and 1/1000 are usually sufficient to capture the linear range, but further dilutions may be required for more concentrated samples.
2. For each primer pair (here, we have used primers targeting NDUFV1 and TOMM20 promoters), prepare two sets of qPCR reactions: one using the diluted input samples from step 4.1, and a second with the DNA samples isolated by immunoprecipitation from step 3.4. Perform each reaction in triplicate.
   NOTE: Reactions can be set up in parallel for multiple genes using 96- or 384-well plates.
3. Set up a reaction volume of 10 µL per well with 5 µL of 2x PCR mix (**Table of Materials**), 1 µL of primer mix (1 µM forward primer and 1 µM reverse primer), and 4 µL of DNA from ChIP (or the input control).
4. Briefly spin down the plate.
5. Run the qPCR reactions on a real-time PCR system using the protocol recommended by the manufacturer for reagent detection (**Table of Materials**). Here we have used the following conditions: 1) 1s at 95°C for 1 cycle; 2) 1 s at 95 °C followed by 20 s at 60 °C for 45 cycles; 3) 15 s at 95 °C followed by 60 s at 60 °C followed by 15 s at 95 °C for 1 cycle.
   NOTE: Check the dissociation curve: the presence of one peak in the thermal dissociation plot suggests a single amplicon is generated from the PCR reaction. If more than one peak is visualized, this is potentially indicative of multiple, non-specific amplicons; in which case, we recommend the design of alternative primer pairs.
6. Determine the linear phase of exponential amplification of the PCR reaction for each primer set by calculating the standard curve using the series of diluted genomic DNA in step 4.1. according to the Ct (cycle threshold) value.
7. If the Ct value of DNA from ChIP and input control are within the linear range of Ct value, calculate the enrichment of DNA from ChIP relative to the input, according to the dilution factor of DNA from ChIP and input control in step 4.1.

5. Amplification of DNA from ChIP for High-throughput Sequencing (Genome-wide Readout)

NOTE: This step can be outsourced to an academic core facility or commercial sequencing company when sequencing capabilities are not available in-house.

1. Prepare a DNA library from the extracted DNA with an appropriate ChIP library preparation kit (see **Table of Materials**) following the manufacturer's instructions.
2. Sequence the library on a high-throughput sequencing system, using 50 bp single-end as readout (see **Table of Materials**).
   NOTE: According to ChIP-seq guidelines suggested by ENCODE[6], a minimum of 10 million reads is recommended for mammalian genomes.

6. Raw Data Analysis

1. Perform quality control on raw sequencing reads using FastQC[8].
   NOTE: Common sequencing quality metrics include per base sequence quality, per base sequence content, per sequence GC content, sequence length, and sequence duplication levels. Generally, a read quality score of Q > 30 across each base position is indicative of high-quality sequencing results. The distribution of GC content across all reads should roughly resemble a normal distribution, with a peak centered around the species' actual genomic GC content percentage.
2. Remove any sequencing adapter contamination and low-quality reads through the read trimming utility Trimmomatic[9].
   NOTE: The implemented default parameters specify the removal of platform dependent adapters, removal of leading and trailing low-quality bases, and use of a 4-base wide sliding window to trim reads when the average quality per base drops below a specified threshold.
3. Align processed reads to the mouse MM10 reference genome[10] with the Bowtie2 aligner[11] using default parameters.
4. Filter aligned reads by removing those with a Bowtie2 mapping quality score (MAPQ) of less than 10 using Samtools[12].
5. Determine various post-alignment quality metrics such as percent reads mapped, strand cross-correlation, cumulative read coverage, and sample replicate reproducibility-.
   Note: Various packages including SPP[13], ChIPqc[14], and Deeptools[15] are available to calculate these quality metrics.
6. Perform peak calling analysis with an input control or appropriate KO sample using the MACS algorithm (MACS2)[16] with default parameters.
7. Remove any called peaks that overlap with known blacklisted regions[17] from the mm10 annotation using Bedtools[18].
   NOTE: Blacklisted regions are areas in the genome that have been shown to have high signal or read counts independent of cell line or sequencing technique. These regions display artificially high artifact signals in certain regions of the genome that can be filtered out of functional genomic sequencing experiments. All replicates should be processed independently through the pipeline and may then be used for irreproducibility discovery rate (IDR). IDR is employed to measure the reproducibility of findings between replicates through the method as described by Li, Brown, Huang, and Bickel[19,20]. The IDR method quantitatively assesses consistency between replicates and is recommended by ENCODE and modENCODE as part of their ChIP-seq guidelines.
8. Use statistically significant peaks for various downstream analyses such as gene ontology[21,22], enrichment, or motif analysis[23].

Representative Results

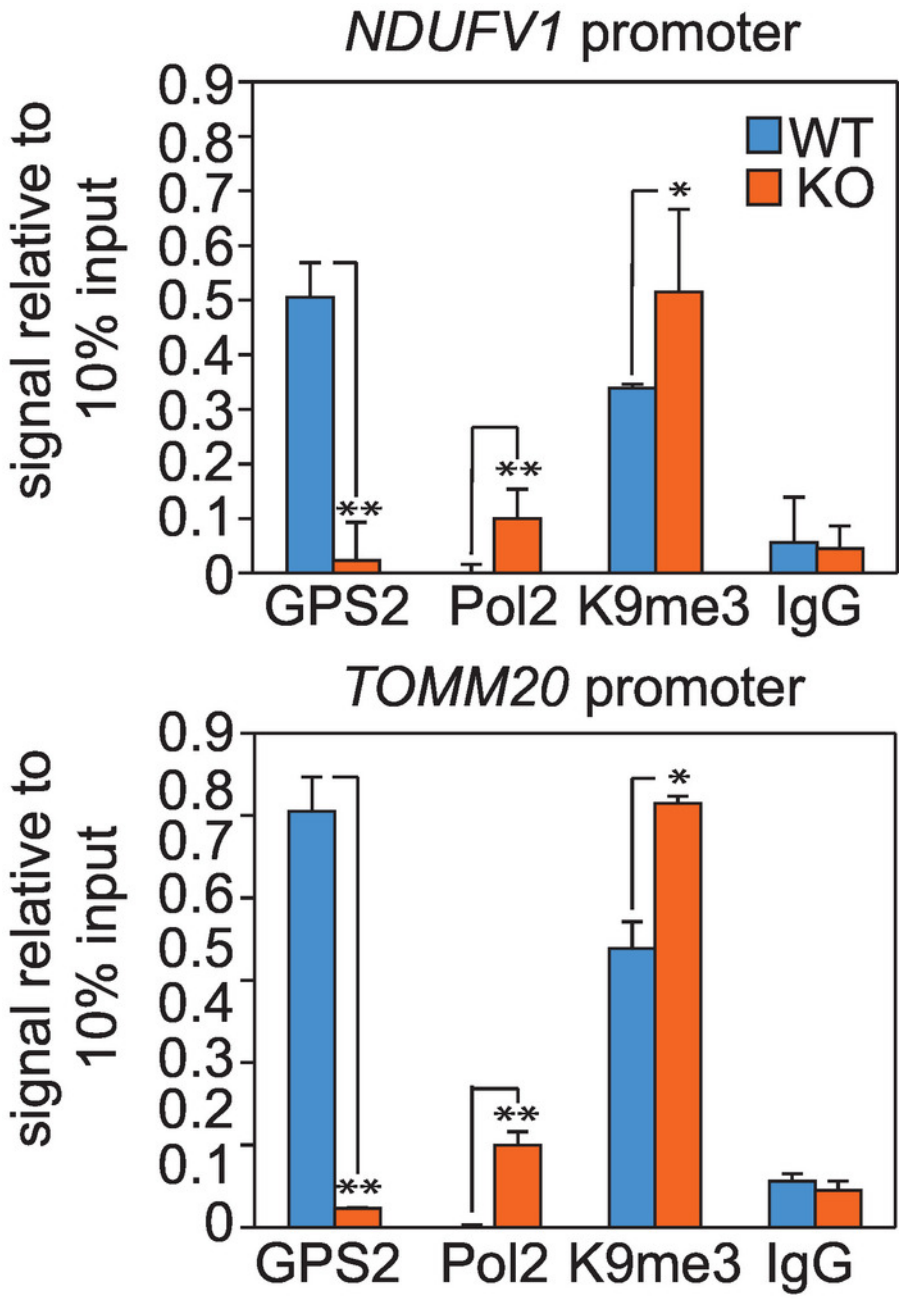

Figure 1: ChIP validation by qPCR. ChIP-qPCR analysis of representative GPS2 target genes *NDUFV1* (left) and *TOMM20* (right) in the BAT of WT and GPS2-AKO mice, showing relative changes in the level of H3K9 methylation and GPS2 and Pol2 binding. The bar graphs represent the sample mean of 3 replicates with *p < 0.05 and **p < 0.01 as calculated with Student's t-test. This figure is modified from Cardamone et al.[2]. Please click here to view a larger version of this figure.

Examples of ChIP-qPCR using BAT isolated from WT or GPS2-AKO mice are shown in **Figure 1**[2]**.** Having found that GPS2 was required for the expression of nuclear-encoded mitochondrial genes in different cell lines[2], we tested the recruitment of GPS2 on two specific nuclear-encoded mitochondrial genes, *NDUFV1* and *TOMM20*, in BAT from mice as an example of a tissue highly enriched in mitochondria. We first compared GPS2 promoter occupancy in GPS2-AKO mice and wild-type littermates, observing an expected decrease in GPS2 binding on selective target genes in the BAT from GPS2-AKO mice . Using the protocol described here, we also recorded increased binding of RNA polymerase 2 (Pol2) and the repressive histone mark H3K9me3 in the BAT from GPS2-AKO mice as compared to wild-type littermates. For all antibodies, the binding was significant as compared to the background signal observed in the IgG control sample. These results confirm the recruitment of GPS2 on selected nuclear-encoded mitochondrial genes and show that GPS2 is required for preventing the accumulation of H3K9me3, and thus required for the stalling of Pol2 transcriptional activity on target promoters. See the **Table of Materials** for more details on the antibodies and the original publication for additional data and more comprehensive discussion of these results[2].

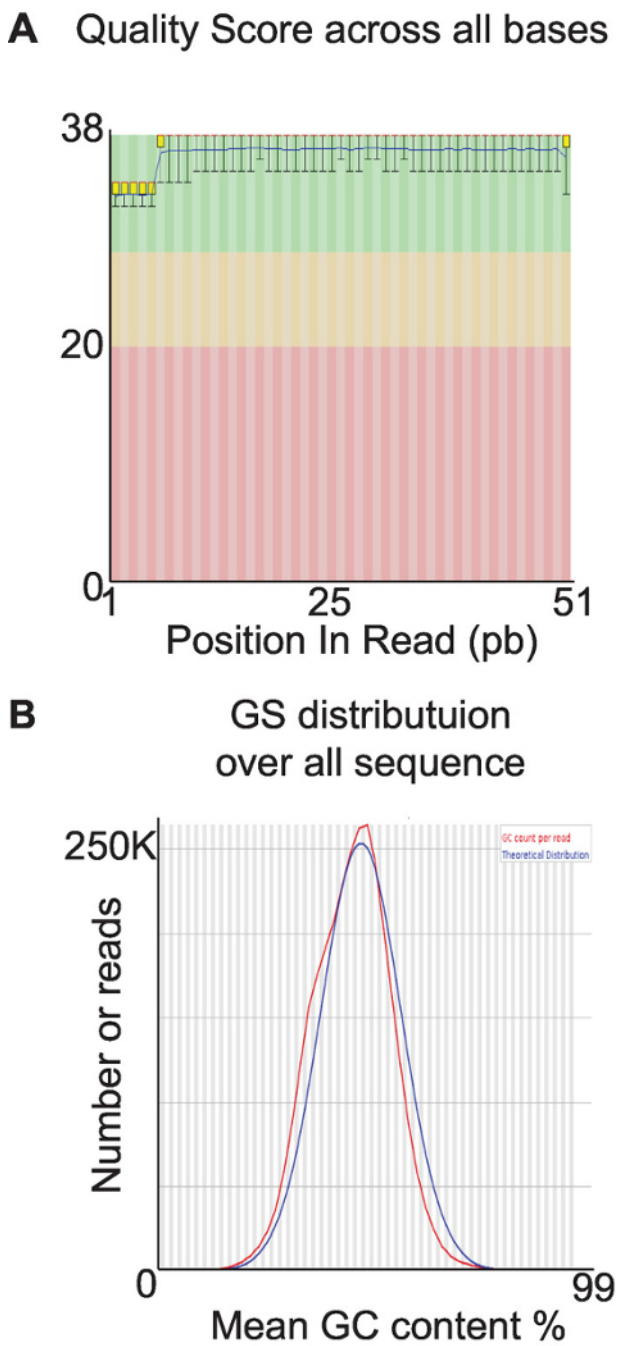

Figure 2: Plot of quality scores from raw sequencing reads isolated from BAT. (**A**) Quality scores reflect a prediction of the probability of an error during base-calling. It is commonly expressed as an integer value, Q = -10log10(P), in which P is the probability of an error in base-calling. Shown above is a per base sequence quality plot generated by FastQC based on raw, unprocessed sequencing reads generated from BAT. (**B**) GC content distribution of reads after adapter removal and read trimming from BAT samples. GC content should roughly resemble a normal distribution with a peak centered around the species' actual fraction of GC genomic content. Please click here to view a larger version of this figure.

In **Figure 2,** sequencing quality scores are shown at a per base level, and reads may be further processed by trimming any residual adapter contamination and removing reads where the average quality score across a sliding window falls below a specified threshold. Also shown is the GC content distribution across reads after adapter removal and read trimming. Sequence quality and GC content distribution are two widely used metrics to evaluate next generation sequencing data before computational analyses are performed. These results demonstrate that the protocol produces a sufficient amount of high-quality chromatin isolated from BAT that is compatible with downstream next-generation sequencing technologies.

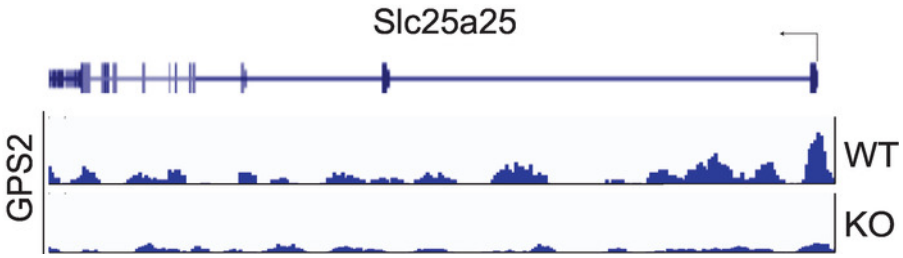

Figure 3: Normalized bigwig files generated from ChIP-seq data from the BAT of WT and GPS2-AKO mice. This figure shows the normalized read coverage along the *Slc25a25* gene locus with a significantly called peak in the promoter region. GPS2 was shown to regulate the expression of nuclear-encoded mitochondrial genes by altering the chromatin state of target gene promoters. These results display the visualization of a significantly called peak on a representative nuclear-encoded mitochondrial gene. Please click here to view a larger version of this figure.

Shown in **Figure 3** are bigWig tracks of the WT and GPS2-AKO aligned BAM files normalized to 1x depth (reads per genome coverage) as implemented in deepTools. Shown is the presence of a statistically significant peak located in the promoter region of the mitochondrial gene *Slc25a25*. The normalized tracks allow visualization of the enrichment of reads corresponding to the called peak and the reduction of reads at this location in the BAT of GPS2-AKO mice relative to wild-type littermates.

Discussion

The protocol described here represents a valuable tool for performing ChIP from murine tissues, specifically optimized for brown adipose tissue. One of the greater challenges in performing ChIP from tissue is recovering a sufficient number of cells during sample preparation. Shearing the BAT using a tissue homogenizer blender coupled with stainless steel beads instead of a canonical glass pestle significantly reduces the number of cells lost due to unbroken tissue. Moreover, homogenizing the tissue directly in a hypotonic buffer helps the release of lipids that can then be easily separated and removed from the nuclei *via* high-speed centrifugation.

Proper sonication is also critical for performing consistent and reproducible ChIP assays. The use of a water bath ultra-sonicator allows for the simultaneous processing of multiple samples, thus improving reproducibility of chromatin shearing and reducing the risk of sample cross-contamination. In addition, use of a temperature-controlled sonication system reduces overheating of the samples, which should be avoided to prevent sample degradation and loss of epitope recognition by the antibody (see **Table of Materials** for details).

Another challenging step is the recovery of immunoprecipitated complexes. Magnetic beads are often preferred for this step because they allow for faster and more effective washes when used together with a magnetic rack. However, they have a lower rate of DNA recovery compared to Protein A Agarose, which can be a serious limitation when working with small amount of starting material. We find that the combination of Protein A Agarose slurry with PVDF 0.45 µm centrifugal filter columns is a great solution to achieve maximum DNA yield with minimal washing time and higher reproducibility.

Regarding the DNA isolation, plenty of columns-based DNA isolation kits can be used. In our experience, one limitation of this approach is the capacity of the columns to have an effect on the yield of DNA recovery. To overcome the problem, we prefer using a more traditional method as phenol/chloroform for DNA extraction.

The listed ChIP-seq pipeline incorporates a number of widely accepted tools and utilities for next generation sequencing experiments. Briefly, the reads are subjected to basic quality control using FastQC and Trimmomatic, then aligned to the mouse MM10 genome using Bowtie2. Surviving reads are filtered by mapping quality before being used for peak calling through the MACS2 algorithm. However, it should be noted that other available and validated tools may also be used depending on the nature of the experiment and data being generated. The use of any customized parameters during pre-processing, alignment, or peak calling may also be appropriate.

References

1. Cannon, B., Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiological Reviews.* (2004).
2. Cardamone, M. D., *et al.* Mitochondrial Retrograde Signaling in Mammals Is Mediated by the Transcriptional Cofactor GPS2 *via* Direct Mitochondria-to-Nucleus Translocation. *Molecular Cell.* **69,** (2018).
3. Emmett, M. J., *et al.* Histone deacetylase 3 prepares brown adipose tissue for acute thermogenic challenge. *Nature.* (2017).
4. Harms, M. J., *et al.* PRDM16 binds MED1 and controls chromatin architecture to determine a brown fat transcriptional program. *Genes & Development.* (2015).
5. Shapira, S. N., *et al.* EBF2 transcriptionally regulates brown adipogenesis *via* the histone reader DPF3 and the BAF chromatin remodeling complex. *Genes & Development.* (2017).
6. Landt, S. G., *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research.* **22**, 1813-1831 (2012).
7. Liisberg Aune, U., Ruiz, L., Kajimura, S. Isolation and Differentiation of Stromal Vascular Cells to Beige/Brite Cells. *Journal of Visualized Experiments.* e50191 (2013).
8. Andrews, S. *FastQC: a quality control tool for high throughput sequence data.* Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (2010).
9. Bolger, A. M., Lohse, M., Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* **30,** 2114-20 (2014).
10. *Genome Browser Gateway.* https://genome.ucsc.edu/cgi-bin/hgGateway?db=mm10 (2018).

11. Langmead, B., Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* **9**, 357-359 (2012).
12. Li, H., *et al.* The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics.* **25** (16), 2078-9 (2009).
13. Kharchenko, P. V, Tolstorukov, M. Y., Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology.* **26**, 1351–1359 (2008).
14. Carroll, T. S., Liang, Z., Salama, R., Stark, R., de Santiago, I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Frontiers in Genetics.* **5**, 75 (2014).
15. Ramírez, F., *et al.* deepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis. *Nucleic Acids Research.* (2016).
16. Zhang, Y., *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology.* **9**, R137 (2008).
17. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489**, (7414), 57-74 (2012).
18. Quinlan, A. R., Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* **26**, 841-842 (2010).
19. Li, Q., Brown, J. B., Huang, H., Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Annals of Applied Statistics.* **5**, (3), 1752-1779 (2011).
20. *Irreproducible Discovery Rate (IDR).* https://github.com/nboley/idr (2018).
21. The Gene Ontology Consortium. Expansion of the Gene Ontology Knowledgebase and Resources. *Nucleic Acids Research.* **45**, D331-D338 (2017).
22. Ashburner, M., *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics.* **25**, 25-29 (2000).
23. Boeva, V. Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. *Frontiers in Genetics.* **7**, 24 (2016).

