Jun 04, 2020   Version 2

🌐 An improved ChEC-seq method for mapping the genome-wide binding of *S. cerevisiae* transcription factors V.2

Rafal Donczew[1], Amélia Lalou[1], Steven Hahn[1]

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109 USA

👤 Rafal Donczew

Protocol status: Working
We use this protocol and it's working

Abstract

ChEC-seq and other nuclease-based methods such as Cut&Run map protein locations on DNA by targeting nuclease activity to specific transcription factors and mapping the resulting DNA cleavages (Schmid et al. 2004; Skene and Henikoff 2017; Zentner et al. 2015). For ChEC-seq, yeast cells expressing a protein-micrococcal nuclease (MNase) fusion are permeabilized, MNase is activated by the addition of calcium, and the resulting DNA fragments are mapped. Potential advantages of this approach include avoiding non-specific protein-DNA crosslinking in highly transcribed regions, efficient mapping of factors that do not directly bind DNA and, more sensitive detection of protein-DNA interactions. We optimized the original ChEC-seq protocol to minimize non-specific DNA cleavage, avoid over digestion at authentic binding sites, and efficiently assay large numbers of factors. We also created a robust data analysis pipeline that incorporates peak calling to map binding sites and quantitative analysis, based on utilization of spike-in DNA, to compare factor-DNA binding under different conditions. We used this modified approach to map genome-wide distributions of the transcription coactivators TFIID and SAGA (Donczew et al. 2020) as well as transcription factors Abf1 and Rap1 (Donczew et al, submitted to Mol Cell).

References

Donczew R, Warfield L, Pacheco D, Erijman A, Hahn S. 2020. Two roles for the yeast transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA.*eLife* 9: e50109.

Schmid M, Durussel T, Laemmli UK. 2004. ChIC and ChEC; genomic mapping of chromatin proteins.*Mol Cell* 16: 147–157.

Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites.*eLife* 6: e21856.

Warfield L, Ramachandran S, Baptista T, Devys D, Tora L, Hahn S. 2017. Transcription of nearly all yeast RNA Polymerase II-transcribed genes is dependent on transcription factor TFIID.*Mol Cell*. 68:118-129

Zentner GE, Kasinathan S, Xin B, Rohs R, Henikoff S. 2015. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo.*Nat Commun* 6: 8733.

Guidelines

General considerations

ChEC-seq and other MNase based approaches have several advantages over ChIP-seq. ChEC-seq is not biased by non-specific crosslinking to highly transcribed regions, by DNA sequences that do not efficiently crosslink with formaldehyde and does not require specific antibodies to the transcription factor (TF) of interest. In practical terms, ChEC-seq experiments are significantly faster and cheaper (no costs associated with antibodies and antibody resins). If cells are ready for harvesting in the morning, the full experiment can be done in one day for multiple samples (alternatively two stopping points are available). In the yeast system, ChEC-seq has also an advantage over the other MNase-based methods developed recently such as Cut&Run and Cut&Tag since ChEC-seq does not require manipulations that may affect gene expression or protein-DNA binding such as digestion of the cell wall and subsequent nuclei isolation. That being said, there are several important considerations associated with ChEC-seq that are discussed below. Since no available genomic technique is free of potential pitfalls, we believe that, for *S. cerevisiae,* ChEC-seq provides a fast, sensitive and robust alternative to ChIP-seq. Both methods complement each other in certain aspects. Depending on the experimental questions being asked, available resources, and other considerations, one or both methods may be used to probe genome-wide binding and function of TFs.

Strain availability

A requirement for ChEC-seq is the need to prepare a yeast strain harboring a fusion between the protein of interest and micrococcal nuclease. In our experience, construction and use of such strains is rapid and straightforward since the tag does not compromise protein function in almost all cases. Even though ChIP-seq experiments can be carried out with a factor-specific antibody (if available), in practice, many researchers still prefer to use one of the epitope tags like Flag, Myc or HA, which also requires strain construction. One-step tagging of yeast strains with a C-terminal MNase fusion is described in (Zentner et al. 2015).

Free MNase control

The DNA cleavage signal generated from a strain expressing MNase with a nuclear import signal not fused to any protein factor serves as an important control in ChEC-seq experiments, which is used in a similar way as the input control in ChIP-seq. The free MNase cleavage pattern provides a basis for peak calling as we expect that the specific interaction of the TF-MNase fusion with chromatin will generate a signal significantly stronger than free MNase. In practice, free MNase generates a cleavage pattern dictated solely by local chromatin accessibility. Thus, free MNase preferentially cleaves exposed DNA including nucleosome depleted regions, but this cleavage shows relatively little variation. When MNase is fused to a TF or coactivator that is localized to specific promoter regions, the cleavage pattern may resemble free MNase when only signal location is considered but the signal intensity at specific and non-specific loci differs significantly.

The free MNase used in ChEC experiments is tagged with a nuclear localization signal for efficient import to the nucleus and should be expressed from a promoter of equal or greater strength compared to the factor of interest (see Donczew et al. 2020 for gene-specific mRNA levels in normal growth conditions). In practice we recommend using a promoter with sufficient activity to ensure that a strain carrying the free MNase will be suitable as a control for mapping a wide range of factors. With the ChEC conditions used here, the free MNase signal is very low compared to that obtained for a TF-MNase fusion with specific DNA localization, even if the promoter of the studied factor is less active than the one driving the expression of free MNase.

Time-course experiments

Earlier applications of ChEC-seq were based on collecting multiple samples for a single experiment, corresponding to different times of MNase treatment. This approach may provide information about the kinetics of the interaction of a TF with DNA, but it is not clear how such data can be used in practice, except to compare the kinetics of factors and free MNase. A drawback of this approach is generation of a large number of samples and increased cost for a limited gain in experimental insights. Based on numerous experiments done in our lab, an efficient alternative is to use a fixed time point after calcium addition (5 minutes) with digestion kinetics limited by a low calcium concentration in the reaction. Such conditions favor the signal from TF-DNA interactions versus random free MNase diffusion. These modified conditions make it feasible to process multiple cultures simultaneously. In practice, we add calcium to consecutive samples every 30 seconds, which allows us to process up to 10 cultures during a single experiment.

Replicate experiments

We recommend collecting at least two biological replicates for ChEC-seq experiments. For mapping of factor binding sites we routinely collect three replicates and use two out of three criterium to identify bound genomic regions. For quantitative analysis, especially when comparing different experimental conditions, it may be advantageous to use an even higher number of replicate experiments. In our experience both ChEC-seq and ChIP-seq sometimes show a significant variation in signal intensity between replicate samples, which makes it hard to identify relatively small changes in binding when using a limited number of replicate samples.

Spike-in DNA

We use MNase digested *D. melanogaster* DNA as a spike-in for quantitative analysis (e.g. treatment/control experiments). For simple mapping of factor binding sites spike-in addition is not necessary because the commonly used peak calling algorithms utilize RPM normalization. Our *D. melanogaster* DNA stock has a concentration of 1 ng/ml. In a typical experiment we supplement Stop Solution with an amount correlated to the OD_{600} measurement of the *S. cerevisiae* culture (volume = OD_{600} x 8 ml).

Limitations related to MNase substrate specificity

When analyzing a ChEC-seq data it is important to realize that MNase cleavage activity can be biased by the local chromatin environment and DNA sequence. Due to this property, data for different factors generated by ChEC-seq carry some qualitative resemblance to free MNase and to each other and it is the signal intensity that primarily discriminates specific versus non-specific interactions. The position of the peak in ChEC-seq data corresponds to the actual binding site as long as it is located clearly in the nucleosome depleted region. For example, we were able to successfully identify known consensus binding motifs in the vicinity of a significant fraction of peaks called for the TFs Abf1 and Rap1 (Donczew et al, submitted to Mol Cell). In other cases, (eg., the TF binding site is adjacent to a nucleosome) the cleavage peak may be shifted some distance from the actual binding site. This can also occur if the binding site is located in a region where nucleosomes are not well positioned. For example, the performance of ChEC-seq within a gene body is significantly decreased compared to ChIP-seq (unpublished results). Consequently, we do not recommend ChEC-seq for mapping transcribing polymerase and elongation factors.

Materials

MATERIALS

⊗ Digitonin **Millipore Sigma Catalog** #300410

⊗ Spermine **Millipore Sigma Catalog** #S3256

⊗ Buffered phenol **Millipore Sigma Catalog** #P4557

⊗ Spermidine **Millipore Sigma Catalog** #85558

⊗ Protease inhibitors **Millipore Sigma Catalog** #04693159001

⊗ Proteinase K **Thermo Fisher Scientific Catalog** #AM2548

⊗ RNase A **Thermo Fisher Scientific Catalog** #EN0531

⊗ Glycogen **Millipore Sigma Catalog** #10901393001

⊗ Mag-Bind TotalPure NGS **Omega Biotek Catalog** #M1378-01

⊗ EDTA **Millipore Sigma Catalog** #E9884

⊗ EGTA **Millipore Sigma Catalog** #E3889

⊗ Tris base **Millipore Sigma Catalog** #TRIS-RO

⊗ Calcium chloride **Millipore Sigma Catalog** #C4901

⊗ Potassium chloride **Millipore Sigma Catalog** #P3911

⊗ SDS **Millipore Sigma Catalog** #436143

Buffers

Buffer A (100 ml)

1.5 ml 1 M Tris-HCl, pH 7.5 (15 mM)

8 ml 1 M KCl (80 mM)

40ml 0.25 M EGTA (0.1 mM)

H_2O to 100 ml

For 10 ml of Buffer A add before use:

-protease inhibitors (to 1x)

-0.9 µl 1.6 M spermine (0.2 mM final)

-3 µl 1 M spermidine (0.3 mM final)

Stop Solution

8 ml 5 M NaCl (400 mM)

8 ml 0.25 M EDTA (20 mM)

1.6 ml 0.25 M EGTA (4 mM)

H_2O to 100 ml

If using spike-in DNA add the appropriate amount to the Stop Solution (see 'Spike-in DNA' in Guidelines).

Comments

2% digitonin - 20 mg/ml in DMSO, store frozen at -80°C, several freezing/thawing cycles have minimal effect on activity, we usually store it in 200 ul aliquots.

We supplement phenol with chloroform and isoamyl alcohol to obtain a 25:24:1 mixture.

Any protease inhibitors preparations are suitable as long as they do not contain EDTA or EGTA.

Before start

Prepare Buffer A and Stop Solution (4 ml and 180 µl required per yeast culture, respectively, if collecting a single time point and undigested control). Buffer A needs to be supplemented with protease inhibitors, spermine and spermidine prior to use (see Materials).

Prepare a single 1.5 ml tube ('stop tube') containing 90 µl Stop Solution, 10 µl 10% SDS (1% final concentration) and spike-in DNA (if used) for each sample to be collected. Routinely we collect two samples for each culture – control (before activation of MNase) and the actual sample after five minutes of digestion.

Cell harvest

1. Grow 50 ml of yeast cells in preferred conditions to $OD_{600} = 0.5 - 0.7$.

2. When the culture approaches the desired OD_{600} set up a heat block or water bath to 30°C and prepare Buffer A.

3. Harvest cells in a 50 ml tube (2000 x g, 3 min).

4. Resuspend cells in 1 ml of Buffer A and transfer suspension to a 1.5 ml tube.

5. Pellet cells (1500 x g, 30 sec) and resuspend in 1 ml of Buffer A.

6. Repeat step 5 once. Pellet cells (1500 x g, 30 sec).

DNA cleavage

7. Resuspend cell pellet in 570 ml of Buffer A. Add 30 ml 2% digitonin (0.1% final), mix and permeabilize cells at 30°C for 5 min in a heat block with shaking (900 rpm).

8. Transfer 100 ml of cell suspension to an appropriate 'stop tube' (control sample) and mix.

   > **Note**
   >
   > We do not routinely sequence the control sample since it does not provide any advantage in data interpretation. In our experience in case of factors which show multiple genomic binding sites a certain enrichment of specific signals can be expected to be seen for the control sample but this minimal MNase activity is not affecting cells well-being in any noticeable way.

9. Add 4 µl of 25 mM $CaCl_2$ (0.2 mM final concentration) to 500 ml of the remaining cell suspension and vortex briefly to mix (low to medium setting). Immediately return the tube to the heat block and start a timer. Keep shaking at 900 rpm.

> **Note**
>
> 0.2 mM final $CaCl_2$ concentration is 10x lower than found in other protocols. We found that this approach limits the background digestion, at the same time allowing for a longer incubation, which in turn makes it feasible to process multiple samples at the same time.

10   After five minutes transfer 100 µl of cell suspension to an appropriate 'stop tube' and mix.

> **Note**
>
> The remaining 400 µl of cell suspension can be used to collect additional time points or technical replicates although we do not do it routinely.

11   Add 4 µl of 20 mg/ml Proteinase K to each collected sample, mix and incubate at 55°C for 30 min.

DNA purification

12   Add 200 µl of phenol/chloroform/isoamyl alcohol, vortex for 10 sec and centrifuge at a maximum speed for 5 min.

13   Remove ~175 µl of an aqueous phase to a new tube. Add 20 mg of glycogen and 500 µl of 100% ethanol. Vortex vigorously to mix and precipitate on dry ice for 10 min.

> **Note**
>
> Protocol can be stopped at this point. Store samples at -80°C.

14   Centrifuge at a maximum speed for 10 min at 4°C to pellet DNA.

15   Discard supernatants and wash pellets with 1 ml 100% ethanol (add ethanol to a tube and remove immediately with a pipette).

16   Spin briefly and remove the residual ethanol with a pipette taking care not to disturb the pellet. Air-dry pellets for 3 min at room temperature.

17   Resuspend DNA in 29 µl TE buffer, add 1 ml of 10 mg/ml RNase A and incubate at 37°C for 15 min.

> **Note**
>
> DNA pellets at this step are sometimes hard to resuspend. We found that spinning the tubes for a few minutes at maximum speed followed by vortexing usually helps. Residual pellet may sometimes be still seen but it is not a reason for concern.

18   Add 60 µl of Mag-Bind beads (2x beads to DNA ratio) and pipet up and down 10x. Incubate the mixture for 10 min at room temperature.

19   Place tubes on a magnetic rack and allow beads to collect for 2 min.

20   Transfer supernatant (~90 µl)) to a new tube containing 106 µl of 10 mM Tris (pH 8.0) and 4 µl of 5 M NaCl (100 mM final concentration). Discard the beads.

> **Note**
>
> Supernatant is enriched in short DNA fragments. Longer DNA fragments arecaptured on the beads and are not useful for further processing. We routinely use Mag-Bind beads, not Ampure XP beads found in other protocols, since they have similar performancefor a significantly lower price.

21   Add 200 µl of phenol/chloroform/isoamyl alcohol, vortex for 10 sec and centrifuge at a maximum speed for 5 min.

22   Remove ~175 µl of an aqueous phase to a new tube. Add 20 mg of glycogen and 500 µl of 100% ethanol. Vortex vigorously to mix and precipitate on dry ice for 10 min.

> **Note**
>
> Protocol can be stopped at this point. Store samples at -80°C.

23   Centrifuge at a maximum speed for 10 min at 4°C to pellet DNA.

24   Discard supernatants and wash pellets with 1 ml 70% ethanol (add ethanol to a tube and remove immediately with a pipette).

25 Spin briefly and remove the residual ethanol with a pipette taking care not to disturb the pellet. Air-dry pellets for 3 min at room temperature.

26 Resuspend DNA in 25 μl 10 mM Tris (pH 8.0).

> **Note**
>
> Samples can be quantified using a high-sensitivity kit but this is not informative because of possible contamination with large DNA fragments, which are not efficiently amplified during library preparation and are not a good substrate for sequencing.

DNA sequencing and data analysis

27 For library preparation we use half of the sample volume (~12 ml), no matter the concentration.

> **Note**
>
> Any DNA library kit is suitable as long as it does not limit the recovery of very short (< 100 bp) DNA fragments since these fragments constitute a significant part of a usable DNA pool. For example, kits which do not require optimizing adapter to sample ratio are not recommended since they are usually based on limiting the recovery of adapter dimers and consequently any short DNA fragments. Our detailed protocol for library preparation is provided in the following articles (Warfield et al. 2017; Donczew et al. 2020).

28 Sequence libraries in a paired-end and 25 bp read length mode.

> **Note**
>
> We routinely use 25 bp read length which is optimal for yeast genome size. 2-3 million reads provide sufficient coverage for a ChEC sample which allows for multiplexing of many samples. In practice, we usually pool together 48 samples for a sequencing on a single lane of a HiSeq 2500 system.

29 We align sequencing reads to a sacCer3 *S. cerevisiae* genome assembly using Bowtie2.

30 Resulting SAM files are converted to tag directories with the HOMER (http://homer.ucsd.edu) 'makeTagDirectory' tool. Peaks are called using HOMER 'findPeaks' tool with optional arguments set to '-o auto -C 0 L 2 F 2', with the free MNase data used as a control. These settings use a default false discovery rate (0.1%) and

require peaks to be enriched 2-fold over the control and 2-fold over the local background. Resulting peak files are converted to BED files using 'pos2bed.pl' program. For each peak, the peak summit is calculated as a mid-range between peak borders Peaks are assigned to promoters if their peak summit location is in the range from −300 to +100 bp relative to the TSS. In rare cases, where more than one peak is assigned to the particular promoter, the one closer to the TSS is used.

31   For quantitative analysis coverage at each base pair of the *S. cerevisiae* genome is calculated as the number of reads that mapped at that position divided by the number of all *D. melanogaster* (spike-in) reads mapped in the sample and multiplied by 10000 (arbitrarily chosen number).