Mapping nucleosome location using FS-Seq

Barry Milavetz, Brenna Hanson, Kincaid Rowbotham, and Jacob Haugen

Department of Biomedical Sciences

School of Medicine

University of North Dakota

Grand Forks, ND 58202

Abstract/Summary

The organization of nucleosomes in eukaryotic chromatin is thought to play a critical in the regulation of the biological function of the chromatin. Because of this potential role in regulation, a number of techniques have been developed which combine chromatin fragmentation around nucleosomes with next generation sequencing to map the location of nucleosomes in chromatin. In this section a procedure using a kit from New England Biolabs (NEB NEXT Ultra II FS DNA library prep Kit) to fragment chromatin in preparation for next generation sequencing is described and compared to other available procedures for mapping nucleosome location.

Key Words

NGS Nucleosomes Chromatin Sequencing Phasing

INTRODUCTION

It has been known for many years that eukaryotic DNA is found within the nucleus of a cell organized with histones to form chromatin. The basic building block of chromatin is the nucleosome which consists of approximately 145 base pairs of DNA wrapped around a histone octomer core containing two copies each of histone H2A, H2B, H3, and H4. As shown in Figure 1 for the eukaryotic virus Simian Virus 40 (SV40), the nucleosomes typically appear as "beads on a string" in chromatin.

Figure 1 also shows a short region of DNA indicated by an arrow which appears to lack at least one nucleosome. This region of "naked" DNA is found in the SV40 regulatory region (1-3) and for obvious reasons has been referred to as a "nucleosome-free region" (NFR). The presence of a specialized chromatin structure, such as the NFR found in SV40 chromatin, characterized by specific nucleosome location and/or histone modifications appears to be a general characteristic of genes poised for transcription or actively transcribing (4).

Initially the presence of modified chromatin structure in regulatory regions of cellular genes was determined by differences in susceptibility to nucleases such as DNAaseI (5), since naked DNA would be expected to digest more quickly than DNA present in nucleosomes. While nuclease sensitivity has been a valuable tool to map the location of putative regulatory regions, it is relatively speaking a low-resolution tool. However, with the development of robust next generation sequencing (NGS) techniques it is now possible to directly map the location of nucleosomes in chromatin to determine whether there is a NFR generated or whether the initiation of transcription results in other changes in nucleosome location.

The NGS workflow Figure 2 consists of four steps; fragmentation of chromatin into nucleosomes, preparation of DNA libraries from the DNA in the nucleosome-sized chromatin fragments, NGS sequencing of the libraries, and bioinformatic analysis of the sequencing data to determine the location of the DNA fragments present in the library.

One of the keys to the NGS workflow is the procedure used to fragment the DNA into nucleosomes. Originally, chromatin was fragmented by micrococcal nuclease and the strategy was referred to as MN-Seq (6). Fragmentation of chromatin by micrococcal nuclease is based upon the fact that the Micrococcal nuclease is a double strand specific endonuclease that would be expected to cleave DNA in the linker region of chromatin without digesting the DNA present in a nucleosome (6). While this is generally true, the procedure has two major disadvantages. First, the specificity of micrococcal nuclease for linker region DNA is relative and because of this, it is necessary to titrate the amount of nuclease, temperature of digestion, and length of digestion to optimize the generation of nucleosome-sized fragments ((6). Second, for reasons that are not completely understood some nucleosomal DNA is much more sensitive to digestion than other nucleosomal DNA (6).

A second strategy for mapping nucleosomes is known as "Assay for Transposase Accessible Chromatin" (ATAC)-Seq. ATAC-Seq uses a transposase (Tagment from Illumina) in vitro to fragment the chromatin by targeting open regions of the chromatin (7). With this procedure the chromatin is fragmented in linker regions and libraries are prepared in one step, since the transposase introduces the linkers needed for library amplification and sequencing. The major disadvantage of the ATAC-Seq procedure is that transposition is very sensitive to higher-order chromatin structure and because of this, the nucleosomes generated tend to be

located in the regulatory region of active genes (7). This result has led to ATAC-Seq being used as an assay for open chromatin (7).

Chromatin Immunoprecipitation sequencing (ChIP-Seq) has also been used to map nucleosomes but primarily for those nucleosomes containing a specific form of histone modification (8). In this procedure antibody is used to target an epitope on a histone, the chromatin is fragmented typically by sonication and the fragments containing the epitope separated from other fragments. Again libraries are prepared from the former and sequenced. The major disadvantage with this procedure is that it is specific to the antibody being used. Frequently that results in only a small number of nucleosomes being mapped in a particular target chromatin.

We have recently described a fourth method for mapping nucleosomes in viral chromatin based upon a kit (New England Biolabs FS) which utilizes a proprietary procedure for fragmenting chromatin (9). We have used this procedure to map the location of nucleosomes in SV40 chromatin and found that the procedure (FS-Seq) yields nucleosome maps similar to but not identical to the maps that we obtain using either MN-Seq, ATAC-Seq, or ChIP-Seq. In this chapter we describe the procedures that we have used for mapping nucleosomes by FS-Seq.

BASIC PROTOCOL FOR PREPARING SEQUENCING LIBRARIES FROM CHROMATIN FRAGMENTED USING THE FS KIT

This protocol describes procedures for fragmenting SV40 chromatin using the proprietary reagents in the New England Biolabs NEXT Ultra II FS DNA library prep Kit and preparing DNA sequencing libraries from the fragmented chromatin using the New England BioLabs NEXT Ultra II DNA library prep Kit and E7335S Mulitplex oligos for Illumina. The protocol includes a procedure using submerged agarose gel electrophoresis to select and purify the subset of the library members which contain insert fragments of SV40 DNA sized from approximately 60 base pairs to 200 base pairs.

MATERIALS

NEB library kit NEXT Ultra II FS DNA library prep Kit

NEB library kit NEXT Ultra II DNA library prep Kit and Mulitplex oligos for Illumina (for example, E7335S)

AMPure XP (Beckman Coulter, #A63880)

Ethanol (Sigma-Aldrich, #459844)

Nuclease-free water (Ambion, #fAM9937)

Certified low-Melt Agarose (Bio-Rad, #1613101)

Certified low-Melt Agarose (Bio-Rad, #1613112)

Agarose (Sigma-Aldrich A6877)

DNA Size Markers

SsoAdvanced Universal SYBR Green Supermix (Biorad, # 172-5274)

Primers to target genomic DNA

Agarose gel sample buffer (see Reagents and Solutions)

GelGreen Nucleic Acid Stain (EmbTec, EC-1995)

Zymoclean Gel DNA Recovery Kit (Zymo Research, # D400)

Monarch PCR & DNA Cleanup Kit (New England BioLabs #T1030S)

MiSeq Reagent Kit v3 (150 cycle)

10 µl Graduated Filter Tips

2 µl Pipetman

10 µl Pipetman

200 µl Graduated Filter Tips

200 µl Pipetman

1000ul Filter Tips

1000 µl Pipetman

200 µl thin-wall PCR Tubes (VWR, # 20170-012)

Eppendorf MasterCycler Personal PCR

BioRad CFX Connect real-time PCR system

Eppendorf snap-cap microcentrifuge flex tubes (Fisher Scientific, #022264111)

Minicentrifuge LabDoctor 12 (MidSci)

Savant DNA 120 SpeedVac concentrator

Power supply Bio-Rad Power Pac 3000 at 125 constant volts

Submerged Agarose Gel Apparatus

EmbiTec PrepOne Sapphire

METHODS

Fragmentation of SV40 chromatin using the New England Biolabs NEXT Ultra II FS DNA library prep Kit

1. We have previously described in detail the procedures that we use to prepare various forms of SV40 chromatin including functionally active minichromosomes and the chromatin from virus particles (10). We have used similar procedures to successfully prepare chromatin for analysis from other small DNA viruses including bovine papillomavirus (BPV) and human papillomavirus (HPV).
2. In preparation for the fragmentation procedure, we first set the block on the Eppindorf Master Cycler Personal PCR to 4^0 C and cooled the appropriate number of the thin-walled PCR tubes for 10 minutes in the block. Typically three or four samples including

controls are fragmented at the same time. When working with a new source of chromatin we would do three samples at the same time; the first would be the input chromatin, the second input chromatin with the supplied reaction buffer, and the third input chromatin, reaction buffer and enzyme.

3. We also set up a Monarch column for each sample and labeled the columns and flex tubes to receive the purified DNA following column purification of the fragmented chromatin.
4. Each PCR tube in a set then received 13 ul of SV40 chromatin using the 200 ul pipetman. The buffer control and enzyme tubes then received 3.5 ul of the reaction buffer using the 10 ul pipetman and the enzyme tube received an additional 1 ul of the enzyme mix from the FS kit using the 2ul pipetman. Finally, using the 10 ul pipetman set to 10 ul, the liquid in each tube was mixed by drawing the liquid into the 10 ul tip followed by forcing the liquid back into the PCR tube.
5. Fragmentation was carried out for varying times and temperatures to optimize the generation of fragments. Typically fragmentation was done for 5 minutes to 15 minutes and at either4^0 C or 37^0 C. The lower temperature was tested first since at 4^0 C nucleosomes would not be expected to slide appreciably and the results would be expected to most closely resemble the original organization of nucleosomes in the chromatin. However, if only limited fragmentation occurred we then tested the higher temperature.
6. Following fragmentation the reaction in each PCR tube was stopped by the addition of 100 ul of binding buffer (Monarch Kit) using the 200 ul pipetman. The binding buffer and sample in each tube was mixed as above using the 200 ul pipetman, and then added to the Monarch column. The DNA was purified according to the protocol and reagents supplied in the Monarch kit. The purified bound DNA was eluted with 13 ul of nuclease-free water (Ambion) and stored at -20^0 C until used in subsequent steps.
7. The extent of fragmentation in each of the samples was determined by qPCR. The assay is based upon the idea that PCR will only amplify a particular region of DNA if the DNA is intact. By comparing the amount of DNA amplification product from the untreated sample to the amount of product following addition of reaction buffer and a mixture of reaction buffer and enzyme mix it is possible to determine the extent of fragmentation which occurred in the buffer alone or with the mixture of buffer and enzymes at each of the fragmentation conditions. In order to analyze SV40 chromatin we have a number of sets of primers all of which yield amplifications products between 200 to 400 base pairs in size. The primer sets recognize various regions of the SV40 genome including the regulatory region, coding region and between the coding regions. We will sometimes compare the extent of fragmentation in the regulatory region to other genomic sites to determine if the regulatory region is hypersensitive to fragmentation in certain forms of SV40 chromatin. Routinely we use the set of primers that are found between the two coding regions, 5'AAAATGAAGATGGTGGGGAGAAGAA3' and 5'GACTCGAGGTGAAATTTGTGATGCT3', which recognize a fragment approximately 250 base pairs in size. We prepare a master PCR mix containing 10 µl of 2X BioRad amplification buffer (SsoAdvanced Universal SYBR Green Supermix), 0.2 µl of each primer (100 µM) and nuclease-free water to 20 µl total volume per sample to be analyzed. 1 ul of the Monarch purified sample DNA is added to a 20 ul volume of the PCR mix and the tubes placed in the BioRad CFX Connect real-time PCR system. The mixture in each PCR tube is preheated in the PCR machine for 10 minutes to activate

the DNA polymerase and then the DNA is amplified for 35 cycles at 95°C for 1 minute to denature the DNA, 54°C for 1 minute to anneal the DNA, and 72°C for 1 minute DNA extension.

8. Following the completion of the PCR amplification the cycle threshold for the untreated sample is compared to the corresponding cycle threshold for the sample containing only the reaction buffer and the sample containing the reaction buffer and enzyme mix. If typical fragmentation is occurring with the FS kit we would expect to observe from no change to a one cycle increase in threshold for the sample containing only buffer and a two to four cycle increase in threshold for the sample containing buffer and enzyme mix. The changes in cycle threshold for the sample containing the buffer and enzyme in the example above would indicate that approximately 75% to 90% of the input chromatin had been fragmented.

Preparation of sequencing libraries from FS fragmented DNA using the New England Biolabs NEXT Ultra II DNA library prep Kit

1. The fragmented DNA obtained using the FS kit was then used for the preparation of sequencing libraries using an NEB NEXT Ultra II DNA library prep Kit designed for sequencing on an Illumina sequencing platform. All biochemical manipulations associated with the preparation of libraries with this kit were performed in a BSL-II hood. An Eppindorf Master Cycler Personal PCR located in the hood was set to a block temperature of 4^0 C and a lid temperature of 65^0C. With the heated lid up, sterile thin-walled PCR tubes being used for library preparation were placed in the 4^0 C block of the cycler and cooled for at least 10 minutes. For most library preparations we generated 8 libraries at one time. At the same time that the cycler was being set up we added 11 ul using the 10 ul pipetman of adapter dilution buffer (NEB) to another PCR tube and placed this tube in a -20^0 C freezer for later use.
2. Libraries were prepared according to the protocol supplied with the NEXT Ultra II library prep kit with minor modifications. In the first step of the protocol the fragmented DNA obtained from chromatin was end-repaired using an enzyme mix supplied with the kit. The purified DNA from the fragmentation step was diluted to 25 ul total volume using the 200 ul pipetman with nuclease-free water and added to a thin wall PCR precooled in the cycler block. To each PCR tube we then added 3.5 ul of repair buffer with the 10 ul pipetman and 1.5 ul of repair enzymes using the 2 ul pipetman both of which were supplied in the kit. The temperature of the cycler block was raised to 20^0 C and the DNA and reagents incubated for 30 minutes. The temperature was then raised to 65^0 C for 30 minutes to inactivate enzymes with the lid closed to keep the temperature even throughout the tubes followed by opening of the lid and cooling the block to 4^0 C.
3. The cooled PCR tubes were centrifuged in the LabDoctor 12 minicentrifuge in the hood to ensure that all the liquid in each of the tubes was located at the bottom of the tubes to maintain the proper concentration of reactants. The tubes were returned to the block in the cycler at 4^0 C. The next step in the protocol from the kit was the ligation of adapters on to the ends of the end-repaired DNA present in each of the tubes.
4. Using the 2 ul pipetman, 0.5 ul of adapter was added to the thawed 11 ul of adapter dilution buffer placed into the cycler block and thoroughly mixed. Next using the 10 ul pipetman 2.5 ul of adapter was added to each sample tube, followed by 1 ul of enhacer again using the 2 ul pipetman. Finally, using the 200 ul pipetman, 15 ul of ligation mix

from the kit was added with thorough mixing to each tube and tubes were incubated in the block for 15 minutes at 20^0 C.

5. The tubes in the block were cooled to 4^0 C for the addition of the USER enzyme. Using the 2 ul pipetman, 1.5 ul of USER (from the NEB primer kit) was added to each tube, the tubes mixed using the 10 ul pipetman set to 10 ul, and incubated at 37^0 C for 15 minutes. At the end of this incubation the tubes were cooled to 4^0 C, centrifuged in the LabDoctor 12 minicentrifuge, and stored in a -20^0 C freezer before purification.
6. The libraries were then purified through a series of steps including column purification, submerged agarose gel size selection, and AMPure purification prior to sequencing. The libraries were column purified using the Monarch kits in part to concentrate the libraries. The frozen libraries were thawed to room temperature in a BSL II hood, 250 ul of binding buffer (Monarch kit) added using a 1 ml pipetman with mixing, and the library in binding buffer added to the column. The sample was centrifuged for 1 minute in LabDoctor 12 mincentrifuge and then the bound DNA washed with 194 ul of wash buffer (from the kit). The library DNA was eluted with 25 ul of nuclease free water (200 ul pipetman) and collected in a sterile flex tube. The eluted libraries were then dried using a Savant DNA 120 SpeedVac concentrator set to heat off and slow dry. This typically took approximately 40 minutes.
7. Each library was size-selected on a mixture of low-melting temperature agarose and standard agarose in order to maximize the number of library members that had the correct size inserts. For a 50 ml total gel volume which was used in our gel apparatus, 0.9 gm of low melt agarose (Certified low-Melt Agarose, Bio-Rad, #1613112) and 0.1 gm of regular agarose (Certified low-Melt Agarose (Bio-Rad, #1613101) added to a 100 ml bottle. For each analysis, 350 ml of buffer was prepared in a 500 ml bottle by diluting a 50X stock buffer (see reagents and solutions) kept at 4^0 C. 50 ml of the running buffer was added to the agarose in the bottle and the mixture was heated in a microwave until all of the agarose dissolved. When completely dissolved 2 ul of GelGreen Nucleic Acid Stain was added to the agarose and the gel was then poured into the gel apparatus.
8. When the agarose was solidified (typically approximately 1 hour at room temperature) the gel was covered in running buffer and the sample wells were loaded. Typically a gel contained 10 sample well and was loaded with size markers in lane 2, lane 6, and lane 10. For our work we use PCR amplification products approximately 120 and 300 base pairs in size. The libraries to be size selected were loaded into lanes 4 and 8 by adding 10 ul of sample buffer to the dried libraries to resuspend the DNA and then the liquid placed in the sample well. The samples and size markers were electrophoresed for approximately 1 hour and 15 minutes with the voltage set to 125 volts or until the blue dye in the sample buffer was at the end of the gel.
9. The gel was removed from the electrophoresis apparatus and placed onto the EmbiTec PrepOne Sapphire. The Sapphire was turned on to illuminate the gel and the center of the bands in the marker lanes were sliced with a blade. The Sapphire was turned off and the gel transferred to paper towels for cutting out of the libraries. Using the slices in the DNA marker lanes as guides the portion of the gel from the library lanes that contained the same sized DNA were cut out. The portion of the gel containing the correct sized DNA was sliced into small pieces and added to 800 ul of binding buffer from the Zymoclean Gel DNA Recovery Kit in a flex tube. The tube was shaken repeatedly until all of the agarose gel had dissolved.

10. The dissolved library was added to the Gel DNA recovery column and purified according to the protocol supplied by the kit. Following the required washes the library DNA was eluted in 25 ul of nuclease-free water. The eluted library was dried as above in the Savant DNA 120 SpeedVac concentrator.
11. The dried library was resuspended in 5 ul of nuclease-free water in preparation for PCR amplification with appropriate primers. Libraries were amplified in a total volume of 160 ul of amplification buffer. The buffer was prepared by adding 80 ul using a 200 ul pipetman of 2X SsoAdvanced Universal SYBR Green Supermix, 1.6 ul of universal primer (NEB Mulitplex oligos for Illumina) using a 2 ul pipetman, 1.6 ul of an indexed primer using a 2 ul pipetman (NEB Mulitplex oligos for Illumina), and water 80 ul nuclease-free water using a 200 ul pipetman.
12. Following thorough mixing of the amplification buffer, a 10 ul aliquot was transferred to a PCR tube with the200 ul pipetman to be used as a non-DNA control. 2.5 ul of the library DNA was added to the remaining 150 ul of amplification buffer using the 10 ul pipetman and the DNA was thoroughly mixed in the buffer. A 10 ul aliquot was removed with the 200 ul pipetman and placed into a PCR tube. The 140 ul of amplification buffer remaining was stored in a freezer at -20^0 C until needed. The non-DNA control and library DNA PCR tubes were then placed in a BioRad CFX Connect real-time PCR system and amplified using 1 minute cycles of 60^0, 72^0, and 95^0. Following amplification the peak for the amplification containing the library DNA was determined from the cycle threshold data generated and the remaining 140 ul was divided into four aliquots of approximately 35 ul each using a200 ul pipetman and then amplified to the cycle threshold empirically determined.
13. Following amplification the amplified library DNA was purified by AMPure. All manipulation of the amplified libraries was performed in a BSL II hood. The amplification buffer in the four tubes were combined and 95 ul of AMPure was added to the tube using a 200 ul pipetman, the contents mixed thoroughly and then transferred to flex tube. The combined contents were incubated at room temperature for 10 minutes to allow the library DNA to bind to the AMPure beads.
14. Following the incubation the tube was centrifuged to ensure that the contents was all at the bottom of the tube, and the tube was placed into a magnetic stand to separate the magnetic beads with bound DNA from the DNA-depleted liquid. The stand was placed on its side while the magnetic beads were bound so that the bound beads would be located approximately half way up the tube and not at the bottom. This was done so that when the liquid was removed there was less chance that the beads would be dislodged. After a ten minute incubation to allow the beads to be separated from the liquid, the stand was placed upright in order to allow the liquid to collect at the bottom of the tube. The DNA-depleted liquid was removed with a 200 ul pipetman set at 200 ul taking care not to dislodge any of the magnetic beads.
15. The beads on the side of the tube were washed twice with 400 ul of a wash solution consisting of 80% ethanol and 20% nuclease-free water which was prepared right before use using a 1 ml pipeteman. Following removal of the second wash the beads were air dried for 10 minutes in the hood.
16. The tube containing the magnetic beads was removed from the magnetic rack and 16 ul of nuclease-free water was added using a 10 ul pipetman set to 8 ul. The beads and water were thoroughly mixed by vortexing and incubated for ten minutes at room temperature. Following the incubation, the mixture was centrifuged in a Minicentrifuge

LabDoctor 12 for 10 seconds to force the beads and liquid to the bottom of the flex tube. The tube was then placed back into the magnetic stand and incubated for an additional 5 minutes to allow the beads to bind to the side of the bottom of the tube and separate from the nuclease-free water which contains the eluted library DNA.

17. A 12 ul aliquot of the nuclease-free water containing the library DNA was very carefully removed from the tube using a 10 ul pipetman set to 6 ul and placed in a new sterile flex tube. This aliquot is stored in the freezer at -20^0 and would be submitted for DNA sequencing if it meets our quality control. The remaining aliquot of the library (4 ul) is also stored in the freezer and eventually analyzed by submerged agarose gel electrophoresis to determine the size and amount of nucleosome-sized DNA in the library.
18. The quality of the amplified libraries was determined using submerged agarose gel electrophoresis. In preparation for analysis of the library DNA we prepared running buffer and an agarose gel. The running buffer was prepared by adding 7 ml of a 50X stock TAE buffer to a 500 ml bottle and adding 350 ml of distilled purified water. To identify the location of DNA in the gel 17.5 ul of ethidium bromide was added to the running buffer using a 200 ul pipetman. In a 100 ml bottle we added 1.4 g agarose (Sigma-Aldrich) and 50 ml of the running buffer and heated the mixture in the microwave to dissolve the agarose. When the agarose was completely dissolved we add 2.5 ul of the ethidium bromide solution, swirl the agarose, and pour it into the gel apparatus.
19. When the gel has cooled and hardened, it is covered with running buffer. The library sample is suspended in 10 ul of sample buffer using a 10 ul pipetman and added to a sample well. In a well adjacent to the library sample we add a DNA marker and subject the samples to submerged electrophoresis for approximately 1 hour and 15 minutes with the voltage set to 125 volts. The gel is removed and the DNA present in the gel visualized on a LiCor Oddyssey FC. A high quality library would be expected to show only a fairly tight band around the size of a nucleosome with added adapters at approximately 250 base pairs in size.
20. Libraries that are of sufficient quality are then used for DNA sequencing.
21. Libraries are sequenced on an Illumina MiSeq using a MiSeq Reagent Kit v3 (150 cycle) in the sequencing core at the University of North Dakota. Typically 20 to 25 individual libraries are sequenced at the same time. Because of the small size of the SV40 genome this usually results in enough reads per library to adequately cover the genome.

Bioinformatic analyses

1. Following sequencing of libraries the data files generated are analyzed using standard bioinformatics software. First, the FASTQ files generated by sequencing are subjected to an initial quality control analysis using FASTQC v.0.11.2 (11). Second, the adapters attached to the ends of the insert DNA during the preparation of the libraries were removed using Scythe v0.98 (12). Third, quality trimming was performed using sickle v1.33 (13), and reads with a phred score less than 30 and reads smaller than 45 base pairs were discarded. Fourth, the reads corresponding to African green monkey (*Chlorocebus sabaeus* 1.1) and human (hg19) sequences were removed following alignment to their respective genomes. While we continue to do this we have found that it has little effect on the actual virus reads. Fifth, the reads present in the FASTQ files remaining after these treatments were then aligned

to the SV40 genome (RefSeq ACC: NC_001669.1) cut between nucleotides 2666 and 2667 using Bowtie2 v2.2.4 (14). Cutting the genome was necessary to display the data as a linear map because the SV40 genome is normally found as a circle. Sixth, duplicate reads were removed using the Picard Tools (Broad) Mark Duplicates function. Seventh, bam files were generated using an awk script from each biological library replicate with filtering for specific size ranges of the DNA. Nucleosome-sized DNA was identified using filtered reads from 100-150 base pairs in size, while potential transcription factor binding sites were identified using reads filtered between 60-99 base pairs in size.

2. The bam files generated are displayed for comparison purposes as merged heatmaps. Typically, a minimum of four libraries generated from different biological replicates are sequenced to generate each heatmap. The individual bam files are normalized first so that all are weighted equally using samtools v1.3.1 (15) and then merged using the R programming language . Bedgraphs are normalized to 1X coverage from filtered deduplicated reads using DeepTools v2.5.4 (16). Finally, heatmaps were generated from the Z-scores of the normalized coverage and displayed using IGV v2.3.52 (17).

*REA*GENTS AND SOLUTIONS

Ethidium Bromide Stain (0.5 µg/µl) 50 mg ethidium bromide, add water to 100 ml

TAE stock (50X) 242g of tris base dissolved in 750ml water. Add 57.1ml glacial acetic acid and 100ml EDTA. Adjust final volume to 1L. Bring the pH to 8.5.

TAE Electrophoresis Running Buffer To 20 ml TAE (50X) stock buffer, add 980ml water

Agarose gel sample buffer 6.0ml 10% SDS, 2.0 ml of 0.1 M EDTA, 50 ml glycerol 1% Coomassie blue, and water to 100 ml

Notes

Comparison of FS-Seq to other methods for mapping nucleosomes

A comparison between the mapping results obtained from FS-Seq and other mapping techniques for the SV40 chromatin found in virus particles is shown in Figure 3.

The figure compares the location of nucleosomes using FS-Seq to ATAC-Seq, MN-Seq, and ChIP-Seq using antibodies which recognize nucleosomes containing H3K9me3 and H4K20me1. It is apparent that many of the brighter yellow bands which represent favored nucleosome locations in the chromatin are present in a number of mapping techniques. For example a very bright band is located in the enhancer using the FS, ATAC, and ChIP-Seq with antibody to H3K9me3 procedures. Similarly, many of the less bright bands also appear to be present in the maps obtained by more than one technique. We include the results with ChIP-

Seq using antibody to H3K9me3 and H4K20me1 to demonstrate that the pattern of nucleosomes obtained with FS-Seq most likely represents the nucleosome locations in the major form of SV40 chromatin that was present in the virus particles since the FS pattern closely resemble the ChIP-Seq result with H3K9me3 antibody which is present in a significant fraction of viral chromatin (18) and does not resemble as much the result with H4K20me1 which only appears to be present in a small amount of the chromatin.

The figure also shows that each mapping technique has preferred sites of action. As is well known, ATAC-Seq targets open chromatin which typically is found in the regulatory region of active genes (7). In marked contrast MN-Seq can under-represent nucleosomes located in regulatory regions if digestion occurs at higher temperatures or for longer periods of time (6). This has led to the suggestion that nucleosomes in certain regulatory regions are "fragile" due to the histone modifications are associated factors that allow them to be targeted by MN-Seq more efficiently than the rest of the nucleosomes in a gene (6). Of course, ChIP-Seq results will vary with the antibody used, since the nucleosomes containing the antibody target is likely to be present at preferred sites in the genome. Based upon our experience with viral chromatin, FS-Seq appears to generate maps that are most similar to ATAC-Seq with more nucleosomes present or ChIP-Seq for target histone modifications that are present more or less throughout the genome.

The FS-Seq procedure is relatively rapid and simple. There are a minimal number of manipulations and the overall procedure can be completed over a 30 minute to 1 hour timeframe depending upon the fragmentation time. The FS kit is designed as a one-step procedure for fragmenting DNA and preparing libraries. The kit accomplishes this by inactivating the fragmenting enzymes during an incubation at 65^0C for 30 minutes. Following inactivation, adapters are ligated to the fragmented DNA as in a usual NEXT Ultra procedure. We chose not to do this because we were concerned that heating the chromatin-enzyme mixture to 65^0C for 30 minutes would be likely to result in over fragmentation of the viral chromatin.

We have not used FS-Seq to map the location of nucleosomes in cellular chromatin. However, we believe that a workflow similar to that which is used for ATAC-Seq would likely work with FS-Seq for this purpose as well. This workflow would consist of preparing nuclei from cells followed by resuspension of the nuclei in reaction buffer and addition of fragmentation enzymes to allow for fragmentation. Fragmentation would be assayed by qPCR measurement of the amount of a target gene or genes that is found in the buffer following incubation at different temperatures or times. Since FS-Seq is enzyme-based like ATAC-Seq and the two procedures yield similar results with viral chromatin, it seems likely that it would also tend to target open chromatin and might be an alternative strategy for analyzing open chromatin.

Specific considerations when preparing fragmented chromatin for FS-Seq

In working with viral chromatin we always determine the relative amount of DNA present in a sample using qPCR. In order to ultimately obtain useful sequencing data following FS-Seq, we have experimentally determined that for a genome that is SV40 in size (5243 base pairs) we need the input chromatin to have a cycle threshold of less than 20 cycles. This is due to the fact that the fragmentation by FS typically results in a shift of the cycle threshold to around 25 cycles. As noted below as long as the amount of DNA in the fragmented samples are in this range useful libraries can be prepared. With larger genomes it is likely that in order to

obtain sufficient coverage of the genome more input chromatin will probably be needed. With SV40 used this way we can obtain anywhere from around 500 reads per library sample to 5,000 reads. Since there are only about 24 nucleosomes in SV40 this is sufficient coverage.

Based upon the relatively large numbers of samples that we have analyzed we have noted that occasionally we will have a sample of disrupted virus which does not appear to fragment very well. At this time we do not know why this is the case, because we have successfully fragmented a number of other samples of chromatin from disrupted virus. We believe that this may be due to the presence of inhibitors remaining with the chromatin. For example, we use a high concentration of dithiothrieitol to disrupt the virus particles and this may be the reason for the problem. We have not noted this issue with other samples that were not prepared in the presence of dithiothreitol. We are presently investigating whether the dithiothreitol is responsible for this inhibition and if so whether there are alternative ways to purify the chromatin from disrupted virus to prevent this inhibition. This observation shows the importance of at least initially using the two controls listed, chromatin alone and chromatin with buffer, since with qPCR of the samples it is possible to quickly determine the extent of chromatin fragmentation. We have also observed that with some samples but not all we observed approximately a one cycle reduction in the sample that contains only added buffer. We believe that this is most likely due to the activation of endogenous nucleases that are present in the biological preparations. This is somewhat variable and appears to depend upon a number of parameters including the time point in SV40 infection and the number of passages that the cells used for SV40 infections have undergone. Typically, in these situations we would still observe a 2 to 3 cycle increase in the cycle threshold number in the sample containing buffer and enzyme compared to buffer alone.

For our fragmentation studies we have usually tried to use the lowest incubation temperature possible to try to limit any natural movement of nucleosomes since nucleosome movement might occur at higher temperatures. However, we have not studied this closely and do not know if it is an important consideration.

Specific considerations when preparing libraries following FS fragmented chromatin

We have found that as long as the cycle threshold of the fragmented chromatin is less than or equal to approximately 25 cycles we can obtain high quality libraries. We judge this by analyzing the quality of the library as described in the methods section by agarose gel electrophoresis and looking for a single relatively sharp band in the region of the gel where we would expect to find nucleosome-sized DNA fragments with attached adapters. When fragmenting a new form of chromatin (such as a new virus) we generally will amplify the prepared library after column purification and analyze it on a gel similarly. In this case we are looking for a smear of amplification products from the correct size to a size significantly larger but with a clear increase in the number of products at the correct size. If the products are all larger than the correct size, it indicates that fragmentation was not done for long enough or at a high enough temperature. We would redo the fragmentation taking this into account and adjust the fragmentation conditions accordingly. When the cycle threshold is greater than 25 cycles we find that the libraries sometimes meet our quality requirements but in some cases do not. We have found from experience that if a library does not show a single broad band at the correct size but instead shows multiple bands or a smear of library elements it is unlikely to yield good sequencing data. For those libraries we have found it better to redo the fragmentation and library preparation with more or a different sample instead of sequencing the available library.

We have found that the purification of the proper sized fragments by submerged gel electrophoresis is an important step. Without this step the fraction of DNA fragments of the proper size is very small and most of the DNA sequenced is discarded during the bioinformatics analysis because it is too large. It is important that the slice of gel used for purification is sufficiently large to include the library elements containing inserts from approximately 60 base pairs to 200 base pairs and not larger elements.

Specific considerations when analyzing sequencing data generated by FS-Seq

In our bioinformatic analysis we include only inserts that are from 100 base pairs to 150 base pairs for nucleosomes. It would be possible to use reads that are somewhat larger but it makes it more difficult to determine exactly where the center of the associated nucleosome is supposed to be. This is because for the larger DNA inserts it is not possible to know whether the nucleosome is found in the center of the DNA or at one of the ends.

We chose to display our data as normalized merged heatmaps. We did this because in our studies on SV40 we noticed that one characteristic of viral regulation in our system appeared to be relatively small changes in nucleosome position (19) and thought that the heatmaps showed this best. However, the data could also be displayed as bedgraphs following normalizing and merging of data. We have chosen to use normalizing and merging as a way to remove some of the biological variability which appears in our system. As indicated above, we also use a minimum of four biological replicates in our studies so that the merged data is from at least four samples. Frequently we have used as many as 10 biological replicates if there appears to be variability. One place in the heatmaps that this can be seen is in the width of the bands in the heatmaps. When bands appear to be relatively broad it probably occurs at a site in the genome where there are more than one preferred nucleosome location.

In our recent publication (9) describing the mapping of nucleosomes on the SV40 genome we also describe how we used the sequencing data from shorter reads to look for the location of transcription factor binding. With the shorter read analysis we found a number of reads that corresponded to the position of SP1 binding at its cognate sequence. Of course, since we are only looking at reads it is not possible to exclude the possibility that there are other factors also bound at this site. Never-the-less, an analysis of shorter reads using the FS-Seq sequencing data may help to identify sites of interest for binding by factors in chromatin.

This work was funded by a grant from the National Institutes of Health, AI142011 (to B.M.) The authors thank Ms. Corina Murphy and New England Biolabs for the generous gift of reagents.

References

1. Scott WA, Wigmore DJ. Sites in simian virus 40 chromatin which are preferentially cleaved by endonucleases. Cell. 1978;15(4):1511-8. PubMed PMID: 215331.
2. Varshavsky AJ, Sundin OH, Bohn MJ. SV40 viral minichromosome: preferential exposure of the origin of replication as probed by restriction endonucleases. Nucleic Acids Res. 1978;5(10):3469-77. PubMed PMID: 214758; PMCID: 342688.
3. Waldeck W, Fohring B, Chowdhury K, Gruss P, Sauer G. Origin of DNA replication in papovavirus chromatin is recognized by endogenous endonuclease. Proc Natl Acad Sci U S A. 1978;75(12):5964-8. PubMed PMID: 216004; PMCID: 393097.
4. Parmar JJ, Padinhateeri R. Nucleosome positioning and chromatin organization. Curr Opin Struct Biol. 2020;64:111-8. doi: 10.1016/j.sbi.2020.06.021. PubMed PMID: 32731156.
5. Weintraub H, Groudine M. Chromosomal subunits in active genes have an altered conformation. Science. 1976;193(4256):848-56. doi: 10.1126/science.948749. PubMed PMID: 948749.
6. Voong LN, Xi L, Wang JP, Wang X. Genome-wide Mapping of the Nucleosome Landscape by Micrococcal Nuclease and Chemical Mapping. Trends Genet. 2017;33(8):495-507. doi: 10.1016/j.tig.2017.05.007. PubMed PMID: 28693826; PMCID: PMC5536840.
7. Sun Y, Miao N, Sun T. Detect accessible chromatin using ATAC-sequencing, from principle to applications. Hereditas. 2019;156:29. doi: 10.1186/s41065-019-0105-9. PubMed PMID: 31427911; PMCID: PMC6696680.
8. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009;10(10):669-80. doi: 10.1038/nrg2641. PubMed PMID: 19736561; PMCID: PMC3191340.
9. Milavetz B, Haugen J, Rowbotham K. Comparing a new method for mapping nucleosomes in simian virus 40 chromatin to standard procedures. Epigenetics. 2020:1-10. doi: 10.1080/15592294.2020.1814487. PubMed PMID: 32838633.
10. Balakrishnan L, Milavetz B. Epigenetic Analysis of SV40 Minichromosomes. Curr Protoc Microbiol. 2017;46:14F 3 1-F 3 26. doi: 10.1002/cpmc.35. PubMed PMID: 28800155.
11. Andrews S. FastQC: a quality control for high throughput sequence data2010.
12. Buffalo V. Scythe: A bayesian adapter trimmer2011.
13. Joshi NA, Fass JN. sickle - A windowed adaptive trimming tool for FASTQ files using quality2011.
14. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25. doi: 10.1186/gb-2009-10-3-r25. PubMed PMID: 19261174; PMCID: 2690996.
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PMCID: PMC2723002.
16. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5. doi: 10.1093/nar/gkw257. PubMed PMID: 27079975; PMCID: PMC4987876.
17. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6. doi: 10.1038/nbt.1754. PubMed PMID: 21221095; PMCID: PMC3346182.
18. Milavetz B, Kallestad L, Gefroh A, Adams N, Woods E, Balakrishnan L. Virion-mediated transfer of SV40 epigenetic information. Epigenetics. 2012;7(6):528-34. doi: 10.4161/epi.20057. PubMed PMID: 22507897; PMCID: 3398982.

19. Kumar MA, Kasti K, Balakrishnan L, Milavetz B. Directed Nucleosome Sliding in SV40 Minichromosomes During the Formation of the Virus Particle Exposes DNA Sequences Required for Early Transcription. J Virol. 2018. doi: 10.1128/JVI.01678-18. PubMed PMID: 30518654.
20. Kube D, Milavetz B. Differential regulation by SV40 T-antigen binding at site I defines two distinct classes of nucleosome-free promoter. Anat Rec. 1996;244(1):28-32. doi: 10.1002/(SICI)1097-0185(199601)244:1<28::AID-AR3>3.0.CO;2-B. PubMed PMID: 8838421.

Figures

Figure 1

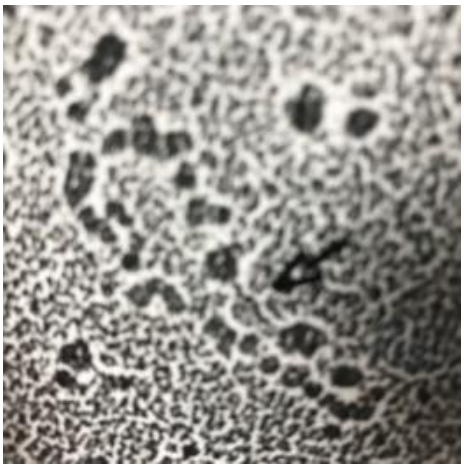

Figure 2

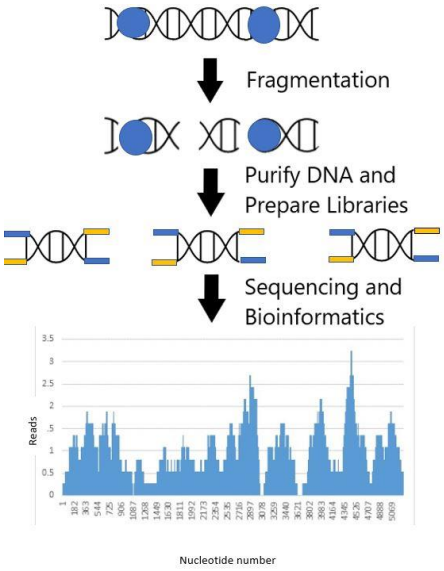

Figure 3

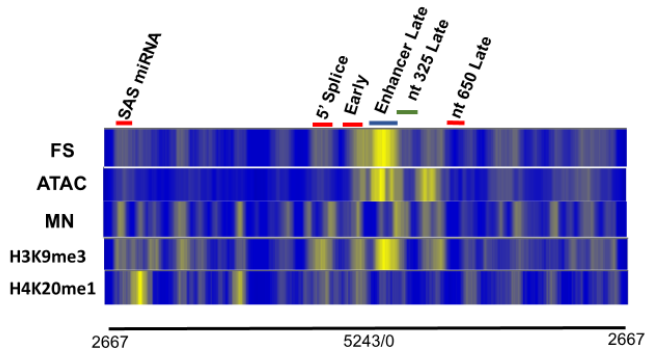

Figure Legends

Figure 1, SV40 minichromosome showing the “beads on a string” structure of chromatin. The arrow indicates the location of a nucleosome-free region. SV40 chromosomes were prepared and analyzed by electron microscopy (20).

Figure 2. Workflow for mapping nucleosomes using chromatin fragmentation and next-generation sequencing. Blue circle = nucleosome. Blue rectangle = adapter 1. Gold rectangle = adapter 2.

Figure 3. Mapping nucleosomes using FS-Seq, ATAC-Seq, MN-Seq, and ChIP-Seq. The chromatin was obtained from SV40 virus particles and analyzed by each of the different procedures as previously described (9). ChIP-Seq is shown using antibody targeting H3K9me3 and targeting H4K20me1 in nucleosomes.