HHS Public Access

Author manuscript

Curr Protoc Mol Biol. Author manuscript; available in PMC 2015 April 21.

Published in final edited form as:

Curr Protoc Mol Biol. 2013 July ; 0 27: Unit–21.27. doi:10.1002/0471142727.mb2127s103.

Genome-scale Mapping of DNaseI Hypersensitivity

Sam John[1,*], **Peter J. Sabo**[1,*], **Theresa K. Canfield**[1], **Kristen Lee**[1], **Shinny Vong**[1], **Molly Weaver**[1], **Hao Wang**[1], **Jeff Vierstra**[1], **Alex P. Reynolds**[1], **Robert E. Thurman**[1], and **John A. Stamatoyannopoulos**[1,2,#]

[1]Department of Genome Sciences, University of Washington, 1705 NE Pacific St., Box 357730, Seattle, Washington 98195, USA

[2]Department of Medicine, University of Washington, Box 357720, Seattle, Washington 98195, USA

Abstract

DNaseI-seq is a global and high-resolution method that uses the non-specific endonuclease DNaseI to map chromatin accessibility. These accessible regions, designated as DNaseI hypersensitive sites (DHSs), define the regulatory features, (eg. promoters, enhancers, insulators, locus control regions) of complex genomes. In this unit, we will describe systematic methods for nuclei isolation, digestion of nuclei with limiting concentrations of DNaseI and the biochemical fractionation of DNaseI hypersensitive sites in preparation for high-throughput sequencing. DNaseI-seq is an unbiased and robust method that is not predicated on an a priori understanding of regulatory patterns or chromatin features.

Keywords

Chromatin; Nucleosome; DNaseI-seq; Transcription; Regulatory DNA

INTRODUCTION

DNaseI hypersensitive site (DHS) mapping has emerged as a powerful tool for genome-wide mapping and profiling of regulatory DNA. The central underpinning in understanding fundamental biological processes such as gene expression, replication, genome organization and their effects on human disease necessitates a methodical documentation of the regulatory compartment of the genome. Prior applications that identified DHSs (Wu et. al., 1979a, b; Wu, 1980; Stalder, 1980; Emerson, et. al., 1985; Zimarino and Wu, 1987; Gross and Garrard 1988; Elgin, 1990), such as Southern hybridization, involved the slow capillary transfer of nuclease digested DNAs to membranes followed by the use of end-labeled radioactive probes. While there have been methodological advances in recent years that have provided incremental improvements (Dorschner et. al., 2004; Sabo et. al., 2004a, b; Sabo et. al., 2006; Boyle et. al., 2008), the coupling of biochemical fractionation and enrichment of DNaseI hypersensitive sites with next-generation sequencing has increased throughput,

[#]Correspondence: jstam@u.washington.edu.
[*]These authors contributed equally

specificity and sensitivity in an unparalleled manner (Thurman et al., 2012a; Neph et. al., 2012a; John et. al., 2011).

This unit contains four basic protocols. BASIC PROTOCOL 1 depicts methods that describe the preparation of nuclei from cells in culture and digestion of chromatin nuclei with limiting quantities of DNaseI. BASIC PROTOCOL 2 describes the biochemical fractionation and purification of nuclease digested DNAs. BASIC PROTOCOL 3 illustrates DHS library construction methods. This includes sample processing specifics, such as adaptor ligation and amplification, in preparation for sequencing on the Illumina Genome Analyzer. Quality control steps that allow for the accurate identification of samples that provide the best likelihood of generating high quality sequencing datasets are outlined in BASIC PROTOCOL 4.

Data from the Illumina Genome Analyzer can be aligned and analyzed using manufacturer provided software. While data analysis is an integral and important aspect in extracting the most useful information from any genome-wide data set, it remains beyond the scope of this protocol to delve into the fundamentals of such analysis. While a number of peak-calling algorithms that aid in the downstream processing and analysis of epigenomics data have been described in the literature (Zhang et. al., 2008; Boyle et. al., 2008; Ji et. al., 2011; Rozowsky et. al., 2009; Valouev et. al., 2008; Jothi et. al., 2008; Heinz et. al., 2010) our experience suggests that the best tools for the determination of nuclease sensitive regions are based on the scan statistic algorithm (Thurman et. al., 2012b, c;). The generation of high quality data is paramount for analyses and comparisons across genome-wide DNaseI-seq datasets. Quality control metrics are provided that can accurately identify DNaseI-digested samples with a high probability of producing quality data.

DNaseI-seq provides an effective platform that is both rapid and reproducible and can provide genome-wide level readouts of regulatory elements across any cell or tissue type in any sequenced species.

Parts of this unit have been published in Nature Methods (P. Sabo, M.S. Kuehn et. al., Nature Methods 3:511–8 (2006)) and Nature Protocols (S. John, P. Sabo, et. al., in press).

BASIC PROTOCOL 1: PREPARATION OF NUCLEI AND DIGESTION WITH DNaseI

This protocol describes a method for purifying nuclei from mammalian cells in culture. Isolated nuclei are treated with limiting quantities of DNaseI to liberate the most accessible regions in chromatin. These short fragments are subsequently purified by size selection over sucrose gradients in preparation for sequencing on high-throughput platforms (Fig. 1).

Materials

- 25 to 50 million mammalian cells
- AG501-X8 (D) 20–50 Mesh Resin (Bio-Rad, cat.no. 143-6425)
- Calcium Chloride 1M (Boston Bioproducts, cat.no. MT-140)

- Complete EDTA-free Protease Inhibitor (Roche Applied Science cat.no. 04-693-132-001)
- Deoxyribonuclease I (Sigma-Aldrich, cat.no. D4527)
- EDTA 0.5M (pH 8.0) (Applied Biosystems, cat.no. AM9262)
- EGTA 0.5M (pH 8.0) (Boston Bioproducts, cat.no. BM-151)
- Glycerol Molecular Biology Grade (Roche Applied Science, cat.no. 03-117-502-001)
- IGEPAL®CA-630 (Sigma-Aldrich, cat.no. I8896)
- NaCl 5M (Mediatech Inc., cat.no. 46-032-CV)
- $MgCl_2$ 1M (Applied Biosystems, cat.no. AM9530G)
- Pefabloc SC Plus (Roche Applied Science, cat.no. 11-873-601-001)
- Potassium Chloride (Sigma-Aldrich, cat.no. P4504)
- Proteinase K (Sigma-Aldrich, cat.no. P4850)
- Ribonuclease A (Sigma-Aldrich, cat.no. R4642)
- SDS 10% Solution (Applied Biosystems, cat.no. AM9823)
- Spermidine Free Base (MP Biomedicals Inc., cat.no. 0215206801)
- Spermine Free Base (MP Biomedicals Inc., cat.no 0215207001)
- Tris-HCl 1M (pH 8.0) (Mediatech Inc., cat.no. 46-031-CM)
- Tris-HCl 1M (pH 7.5) (Mediatech Inc., cat.no. 46-030-CM)
- 1, 4-Dithiothreitol (Sigma-Aldrich, cat. no. D9680)
- Sterile Milli-Q or Molecular Biology Grade Water
- Biosafety Cabinet (Nuaire, model no. NU-425-600)
- Benchtop Centrifuge (Eppendorf, model no. 5810R)
- Swinging Bucket Rotor (Eppendorf, cat.no. A-4-62)
- 2, Digital 10 liter Water Bath (Fisher Scientific, cat. no. 15-462-10Q)
- Inverted Phase Contrast Microscope (Nikon, model no. TS-100)
- Phase Hemocytometer (VWR, cat. no. 15170-263)
- 10ml syringe (Becton Dickinson cat.no 309646)
- 18 gauge 1 inch PrecisionGlide needle (Becton Dickinson cat.no. 305195)
- 15ml Polypropylene Conical Centrifuge Tubes (Corning, cat.no 430766)
- 50ml Polypropylene Conical Centrifuge Tubes (Corning, cat.no 430828) Pipetman p20 (Rainin, cat. no. L-20)
- Pipetman p200 (Rainin, cat. no. L-200)

- Pipetman p1000 (Rainin, cat. no. L-1000)
- 20 µl pipet tips (Rainin, cat. no. RT-L10F)
- 200 µl pipet tips (Rainin, cat. no. RT-L200F)
- 1000 µl pipet tips (Rainin, cat. no. RT-L1000F)
- Graduated pipets (5, 10, 25, 50ml)
- Microfuge tubes 1.7 ml (Genesee Scientific, cat.no. 24-282)

Prior to Nuclei Isolation

1 Add 1 complete EDTA-free protease inhibitor tablet to Buffer A (1 tablet per 50ml solution, see recipe for Buffer A in REAGENTS AND SOLUTIONS) and dissolve. Keep on ice.

2 Prepare a fresh 2X IGEPAL CA-630 solution in Buffer A. It is recommended that the concentration of detergent that works best on a given cell type be determined empirically on a small sampling of cells. For example, if cells are most effectively lysed at a detergent concentration of 0.1%, a 0.2 % solution (2X IGEPAL CA-630) is prepared. See Step 10 in BASIC PROTOCOL 1 and Critical Parameters and Troubleshooting section on Preparation of nuclei. See REAGENTS AND SOLUTIONS for preparation of 10% deionized IGEPAL.

3 Add spermine free base and spermidine free base to Stop Buffer (see recipe in REAGENTS AND SOLUTIONS). If SDS in the Stop Buffer has precipitated out of solution, warm to 37°C to dissolve SDS **prior** to adding spermine and spermidine.

4 Prepare fresh 1X DNaseI Digestion Buffer (see recipe in REAGENTS AND SOLUTIONS) by diluting 10X DNaseI Digestion Buffer (1:9) with Buffer A.

5 Aliquot 1X DNaseI Digestion Buffer into 15ml conical tubes. The number of tubes is determined by the number of DNaseI treatments being performed. For example, for two individual digestions of 5–10 million nuclei, 1 ml of 1X DNaseI Digestion Buffer is aliquoted into two separate 15 ml conical tubes.

6 Warm Stop Buffer and 1X DNaseI Digestion Buffer (minus DNaseI) in 37°C water bath. Allow solutions to equilibrate for 60 minutes prior to use.

7 Pre-cool centrifuge to 4°C. All centrifugations should be done at 4°C.

Nuclei Isolation from Cultured Cells

8 Cell suspensions are counted using a hemocytometer or an automatic cell counter. The numbers of cells necessary for a given experiment are aliquoted into a conical tube.

For a single DNaseI digestion, 5 million cells are needed (assuming 100% recovery of nuclei). Since multiple DNaseI concentrations are employed in a typical experiment, it is common to start with 25 to 50 million cells.

9 Cells are pelleted for 5 minutes at 500 × g at 4°C and washed with cold PBS.

10 Resuspend the washed cell pellet gently and thoroughly in ice-cold Buffer A at 5 million cells per ml. Add an equal volume of ice-cold 2X IGEPAL CA-630 solution to cells.

From this step onward, work quickly. The optimal concentration of IGEPAL CA-630 is calibrated to maximize nuclear recovery with minimal damage to nuclear integrity and must be determined empirically for each cell type. Also see Critical Parameters and Troubleshooting section on Preparation of nuclei. We have found the optimal ***final*** *IGEPAL CA-630 concentration to range from .01% to .5%.*

11 The suspension is gently mixed and inverted gently 10 times and kept on ice for 8 minutes.

12 Nuclei are pelleted for 5 minutes at 500 × g at 4°C.

Some types of cells/nuclei are exquisitely sensitive to centrifugal forces and may need to be spun at lower speeds. The nuclear pelletforms a loose pellet and can be easily dislodged. The supernatant needs to be carefully aspirated.

13 The nuclear pellet is washed with fresh ice-cold Buffer A (1 ml per 5 million nuclei).

14 The efficiency of nuclear isolation can be gauged by counting nuclei using a hemocytometer under a phase-contrast microscope.

10 μl of the nuclear suspension counted using a hemocytometer. Nuclei counts should be equivalent or slightly lower than the initial cell count. Typical recoveries of nuclei range between 80–90%. Individual nuclei should be observed with no visible clumping. Clumping during the incubation of cells with detergent is an indication of excessive IGEPAL resulting in nuclear lysis.

15 Nuclei are aliquoted into individual 15-ml conical tubes for DNaseI treatment.

It is recommended that 5 million nuclei be used per DNaseI digestion (at a concentration of 10 million nuclei per ml).

16 Centrifuge nuclei for 5 minutes at 500 × g at 4°C.

17 Aspirate supernatant from nuclei pellets and place on ice in preparation for DNaseI digestion.

DNaseI digestion of nuclei

18 The Stop Buffer and aliquots of 1X DNaseI Digestion Buffer (number of aliquots depend on the number of DNaseI concentrations being used) should be equilibrated at 37°C in a water bath. (see Prior to Nuclei Isolation above, step 6).

Typically 4 to 5 DNaseI concentrations are used in the optimization of DNaseI digestion. Also see Critical Parameters and Troubleshooting section on DNaseI digestion of nuclei.

19 Add 5 μl proteinase K per ml of Stop Buffer. The Stop Buffer with proteinase K can be kept at 37°C during the course of the experiment.

With the exception of the DNaseI digestion step, which is performed at 37°C, nuclei should be kept on ice for the duration of experiment.

20 The appropriate amount of DNaseI enzyme is added to individual 1X DNaseI Digestion Buffer aliquots (For example: For an 80 unit/ml digestion, add 8 μl of 10 units/μl stock DNaseI enzyme to 1ml of 1X DNaseI Digestion Buffer). Mix completely by gentle pipetting.

Do not vortex as the enzyme denatures easily with aeration. We have found the optimal DNaseI concentration to be between 40 and 160 units/ml for cells studied. The optimal concentration will need to be determined empirically for each cell type being studied. Using a range of DNaseI concentrations, we identify concentrations where high levels of cutting occur in DHSs with little cutting in non-DHS regions. Such differences can be readily determined using a combination of QC gels and q-PCR analysis (see BASIC PROTOCOL 3*). DNaseI lots should always be calibrated prior to use.*

21 Place tubes with nuclei in 37°C water bath and allow the nuclear pellet to equilibrate for 1 minute.

This step and the following steps should be timed carefully.

22 Gently re-suspend nuclei (at a concentration of 10 million nuclei per ml) with pre-warmed 1X DNaseI Digestion Buffer, containing the desired concentration of enzyme.

Pipet gently using wide-bore tips to ensure even suspension.

23 Digest nuclei for 3 minutes at 37°C in a water bath.

24 Reactions are terminated by adding an equal volume of Stop Buffer to the DNaseI reaction tube. Mix by inverting. Transfer tubes to a 55°C water bath.

25 DNaseI digested nuclei are deproteinized for 1 hr at 55°C.

26 Samples can be stored at 4°C. Samples have been found to be stable for more than 2 years at 4°C.

Prior to QC steps (gel electrophoresis (QC gel, Fig. 2) and q-PCR, Fig. 3), incubate the samples at 37°C for 30 minutes with 1.5 μl of 30 mg/ml RNaseA per ml of DNaseI-digested sample.

BASIC PROTOCOL 2: BIOCHEMICAL FRACTIONATION OF DNaseI HYPERSENSITIVE SITES

DNaseI digestion of chromatin in intact nuclei releases small DNA products that arise from multiple cutting events over short distances within accessible chromatin (Sabo et. al., 2006; John et. al., 2011; Thurman et. al., 2012a; Neph et. al., 2012a). These short fragments can then be purified and enriched by size selection over sucrose cushions.

Materials

- Phase Lock Heavy Tubes (5 Prime, cat.no. 2302850)
- Centricon 10 kDa MWCO (EMD Millipore, cat.no. UFC 701008)
- Sucrose BioUltra Molecular Biology Grade (Sigma-Aldrich, cat.no. 84097)
- Open-top thick walled polyallomer tubes (Seton Scientific, cat.no. 5032)
- Benchtop Centrifuge (Eppendorf, model no. 5810R)
- Swinging Bucket Rotor (Eppendorf, cat.no. A-4-62)
- Standard laboratory ultracentrifuge (eg. Beckman LE-80)
- Phenol-Chloroform-Isoamyl Alcohol (Invitrogen, cat.no. 15593-031)
- SyBr Green I (Invitrogen, cat.no. S7563)
- Qiagen MinElute PCR purification kit (Qiagen, cat.no. 28004)
- Qiagen MinElute Gel extraction kit (Qiagen, cat.no. 28604)
- Microfuge tubes 1.7 ml (Genesee Scientific, cat.no. 24-282)
- QIAvac 24 Plus Vacuum manifold (Qiagen, cat.no. 19413)

DNA clean-up and concentration

1 DNaseI, Proteinase K and RNase A digested samples (from Step 10 above) are incubated at 55°C for 10 minutes.

2 Transfer samples to 15 ml PhaseLock Heavy tubes by gently pouring.

To avoid shearing of fragile, high-molecular weight DNAs, samples are poured into PhaseLock tubes. Avoid pipetting.

3 Add 2 ml Phenol-Chloroform-Isoamyl alcohol and mix by gentle rocking for 10 minutes at room temperature.

4 Samples in PhaseLock tubes are spun at 4,000 × g at room temperature for 10 minutes.

5 Add 2 ml chloroform and mix by gentle rocking for 10 minutes at room temperature.

6 Spin at 4,000 × g at room temperature for 10 minutes.

7 Decant aqueous (top) phase directly into a 15 ml Centricon 10kDa MWCO column.

8 Spin at 4,000 × g at room temperature for 25 minutes to concentrate samples to ~250 µl.

Phenol-chloroform extracted DNAs (2 ml total for a 10 million nuclei digestion) are concentrated to an approximate volume of 250 µl in a Centricon (10 kDa MWCO) by spinning in a benchtop centrifuge. The time for concentration is entirely dependent on the volumes being centrifuged. We monitor the extent of

concentration by periodically stopping the centrifugation and estimating the volume of the retentate. Typically, 2 ml of sample (10 million nuclei equivalents) are concentrated in less than 30 minutes.

Size selection of DNaseI hypersensitive sites

9 During sample concentration, a 9% sucrose cushion is prepared (see recipe in REAGENTS AND SOLUTIONS). 7.5 ml of a 9% sucrose solution is layered in an open-top, thick-walled polyallomer ultracentrifuge tube (Seton Scientific). The total volume accommodated by the tube is 8 ml.

10 The concentrated DNA sample (~250 µl) is gently layered on top of the sucrose.

An interface should be visible between the sample and the cushion. Tubes need to be weighed prior to spinning in an ultracentrifuge. Weight imbalances can be corrected with sucrose buffer.

11 Samples are centrifuged in a SW41 swinging bucket rotor using a Beckman LE-80 Ultracentrifuge at 25,000 × g for 24 hours at 20°C. Use slow acceleration and no brake during the deceleration.

12 Pipette 600 µl fractions from the top of each of the gradients.

Each fraction is collected with minimal perturbation to the gradient as possible. An alternative to manual collection is the use of an automated liquid handling system (eg. BioMek), if available. This is a more reproducible and less labor intensive method for collecting fractions.

13 Load 10 µl of each fraction on a 2% agarose gel (1X TAE), 100V for 1 hour.

14 Post-stain with 1X SYBR Green I for 1 hour.

15 DNA fragments are detected by imaging on a scanner (eg. Typhoon 9200 scanner).

*To account for the thickness of the gel, use a +3 mm focal plane setting when scanning on the Typhoon 9200 (*Fig. 4).

16 Fractions containing fragments smaller than 500 bp are pooled in a 15 ml conical tube and 3X sample volume of Qiagen QG buffer (found in Qiagen MinElute Gel extraction kit) is added and mixed.

17 Add 1X (of starting sample volume) of 100% isopropanol. Mix and let sit for 15 min at room temperature.

18 Samples are bound and washed on Qiagen MinElute columns. [*Author; please expand on what you mean by ‘using a vacuum manifold’, e.g., where is it attached, and also please clarify how the syringe is used in applying and efficiently binding the samples to the Qiagen column. E.g., Vacuum attached on the bottom and sample loaded into syringe?] Because of the large volumes involved, we highly recommend performing this step using a vacuum manifold, that facilitates the processing of large volume samples, by connecting a 10 ml syringe to the top of a minelute column which in turn is directly connected to the

vacuum manifold. Samples from Step 16 above can then be applied to the syringe.

19 After samples have been loaded onto Minelute columns, they are allowed to incubate with Qiagen wash buffer (PE) for 5 minutes before being filtered through the column.

20 DNA is eluted with two consecutive 12 µl elutions of 1mM Tris-HCl (pH 8). After adding 12 µl of elution buffer to the MinElute membrane, the column is allowed to incubate for 5 min before elution.

21 DNA concentrations are determined using a fluorometer (eg. Qubit from Invitrogen). Starting with 5–10 million cells, we typically obtain 10–50 ng of DNA fragments (<500 bp).

Samples can be stored at −20°C until further use.

BASIC PROTOCOL 3: PREPARATION OF DHS LIBRARIES FOR SEQUENCING ON THE ILLUMINA GENOME ANALYZER

This unit describes the preparation of size-selected, DNaseI hypersensitive DNA for high-throughput sequencing. These DNAs are ligated with adaptors and lightly amplified to generate a DHS library that can be subsequently sequenced on the Illumina sequencing platform.

Materials

- Sterile Milli-Q or Molecular Biology Grade Water
- SyBr Green I
- Qiagen MinElute Gel extraction kit (Qiagen, cat.no. 28604)
- Klenow DNA polymerase (New England Biolabs, cat. No M0210L)
- Klenow (3′ to 5′ exo−) (New England Biolabs, cat. No M0212L)
- T4 DNA Ligase (New England Biolabs, cat. No M0202L)
- NEB Quick Ligase (New England Biolabs, cat. no M2200L)
- Kapa HiFi Library Amplification Kit (KAPA Biosystems, KK2611)
- Kapa Hifi Library Quantification Kit (KAPA Biosystems, KK4824)
- Standard laboratory PCR machine (eg. BioRad DNA Engine)
- Microfuge tubes 1.7 ml (Genesee Scientific, cat.no. 24-282)

Library Preparation of DHS samples

End-repair

1 Prepare the following reaction on ice. Total volume is 30 µl.

18.75 µl	Purified DNA (5–50 ng; all DNA from Step 14 in BASIC PROTOCOL 2)
2.5 µl	T4 DNA Ligase Buffer + 10 mM ATP
1.0 µl	dNTP mix (10 mM each)
1.25 µl	T4 DNA Polymerase
0.25 µl	Klenow DNA Polymerase
1.25 µl	T4 PNK

2 Incubate for 30 minutes at 20°C.

3 Add 125 µl of Buffer PB (from Qiagen MinElute kit and purify using MinElute columns. Elute in 16 µl of kit provided elution buffer (EB).

A-tailing

4 Prepare the following reaction on ice. Total volume is 25 µl.

16 µl	Blunted DNA (from Step 3 above)
2.5 µl	10X NEB Buffer 2
5.0 µl	dATP (1mM)
1.5 µl	Klenow (3′ to 5′ exo−)

5 Incubate for 30 minutes at 37°C

6 Add 125 µl of Buffer PB and purify using MinElute as in Step 3 in End-repair section. Elute in 10 µl of Buffer EB.

Adapter ligation

7 Prepare the following reaction on ice. Total volume is 25 µl.

10 µl	A-tailed DNA (10 µl from Step 3 above)
12.5 µl	2X Quick Ligase Buffer
1.25 µl	Illumina oligo adapter mix (110 dilution)
1.25 µl	DNA Quick Ligase

8 Ligate adapters for 15 min at 20°C

9 Add 125 µl of Buffer PB and purify using MinElute columns. Elute in 20 µl of Buffer EB.

Store excess post-ligation products at −20°C

PCR amplification of adapter ligated fragments

10 Prepare the following reaction on ice. Total volume is 50 µl.

5 µl Adapter ligated DNA (5 µl from above)

1 μl PCR Primer 1.0 (1:2 dilution)

1 μl PCR Primer 2.0 (1:2 dilution)

25 μl KAPA Hifi HotStart Taq polymerase

18 μl Water

11 Reactions are amplified with the following parameters:

30″ 98°C, [10″ 98°C, 30″ 65°C, 30″ 72°C] × 12 cycles, 5″ 72°C.

It is important to minimize the number of PCR cycles as excessive amplification can lead to template selection, library bias and partial accessibility profiles.

12 Add loading dye and 5 μl (1:500) SyBr Green I to PCR reaction. Load entire PCR reaction on a preparative 2% agarose gel (1X TAE).

13 Run gel for 40 min at 75V.

14 Image on a Typhoon 9200 scanner (using +3 mm focal plane setting).

15 Cut the amplified library band and add 6X (gel mass) volume of QG buffer to gel slice and incubate 30 minutes at room temperature. Rotate by inversion.

16 Add 2X (gel mass) volume of isopropanol and vortex.

17 Load sample onto a MinElute column and wash and elute (in 15 μl of EB) as described in Steps 11–13 in BASIC PROTOCOL 2.

18 The 15 μl of eluate is used to re-elute residual material on the column. The membrane is allowed to soak in elution buffer for 5 min before eluting.

19 Libraries can be quantified using the KAPA Hifi Library Quantification Kit (according to manufacturer's recommendations). Optional step.

PAUSE POINT Samples can be stored at −20°C prior to sequencing.

20 The high-throughput sequencing of DNaseI libraries on the Illumina platform is performed as per manufacturer's recommendations.

The band of interest is typically the most prominent band on the gel. DNaseI libraries are usually in the size range of 200–500 bp. Libraries can be further quantitated using KAPA Hifi Library quantification kits (follow recommendations according to manufacturer). PCR Primers 1.0 and 2.0 (used in Step 1 of 'PCR amplification of adapter ligated fragments' above) are standard Illumina primers. These steps for library preparation have been optimized for the Illumina platform.

BASIC PROTOCOL 4: QUALITY CONTROL FOR DNaseI DIGESTION

The construction and sequencing of DHS libraries on next-generation sequencing platforms is expensive. It is, therefore, important to introduce cost-effective checkpoints to assess sample quality prior to proceeding to this step. Two easy checks, a simple agarose gel that provides a rough but informative visual readout of the extent of digestion or a quantitative

PCR based method using hallmark accessible chromatin and region-specific primers to gauge the amount of digestion provide safeguards that will ensure the generation of high-quality genome-wide accessibility data.

Materials

- Sterile Milli-Q or Molecular Biology Grade Water
- SyBr Green I
- Qiagen MinElute PCR Purification kit (Qiagen, cat.no. 28004)
- Klenow DNA polymerase (New England Biolabs, cat. No M0210L)
- q-PCR reagents and standard laboratory q-PCR machine (eg. BioRad iCycler)

QC Gel

1 From Step 10 of 'DNaseI digestion of nuclei', load 10,000 nuclei equivalents on a 1% agarose gel.

2 Run at 100V for 1 hr.

3 Post-stain in 1X (1:10,000) SYBR Green I in 1X TAE gel buffer

4 Image on a Typhoon 9200 scanner (using +3 mm focal plane setting)

A light smear as seen in Fig. 2 (40U and 60U lanes, top panel) is a qualitative indicator of a good digestion. Extensive digestion (as observed with 100 units of digestion) or laddering (lower panel of Fig. 2) is suggestive of over-digestion with DNaseI or cells undergoing apoptosis.

Quantitative PCR for measuring DNaseI digestion specificity

5 100 μl of DNaseI-digested DNA (from Step 10 of 'DNaseI digestion of nuclei') is purified over Qiagen PCR purification columns.

The reference sample for q-PCR reactions is the undigested control sample. Before purification over Qiagen MinElute columns, the viscous control sample is sheared by pipetting vigorously until the sample is no longer viscous. Primer pairs and sequences for DHSs that are accessible in many cell types can be obtained from Supplementary Table 3 in Sabo et. al., 2006. Also see section in Critical Parameters and Troubleshooting on Assessment of digestion specificity using quantitative PCR.

6 DNA is quantitated using a nanodrop.

7 10 ng of DNA from each DNaseI condition is used in a real-time q-PCR reaction. The exact methodology of amplicon detection is dependent on the real-time PCR machine being used (eg. SyBR Green in a Roche LightCycler or BioRad iCycler). In principle, regions that are hypersensitive amplify less efficiently than identical regions in the control (undigested) sample. A general rule of thumb is that a strong DHS is reflected by an approximate 70% reduction in amplification efficiency (Fig. 3). It is important that DNaseI-insensitive

regions show no reduction in amplification compared with control (undigested) DNA samples.

REAGENTS AND SOLUTIONS

10X DNaseI buffer

To prepare 10X DNaseI buffer, add 3 ml of 1M CaCl2 and 7.5 ml of 5M NaCl to 39.5 ml sterile Milli-Q water. This solution can be stored at room temp for 1 year.

60 mM $CaCl_2$

750 mM NaCl

3M KCl solution

Dissolve 7.46 grams KCl in 100 ml Milli-Q H_2O. Autoclave 15 min at 121°C. This solution can be stored at room temp for 1 year.

Spermine (0.5M)

Dissolve 1 gram spermine free base in 9.89 ml Milli-Q water. Aliquot 0.5 ml aliquots into 1.5 ml microfuge tubes. Store for 12 months at −20°C ! **CAUTION** Wear gloves during preparation because spermine causes burns.

Spermidine (0.5M)

Dissolve 1 gram spermidine free base in 13.77ml final volume sterile Milli-Q water. Aliquot 1 ml in 1.5 ml microfuge tubes. Store for 12 months at −20°C. ! **CAUTION** Wear gloves during preparation because spermidine can cause severe burns and is irritating to skin, eyes and respiratory system.

10% Deionized IGEPAL CA-630 vol/vol

Warm stock bottle of IGEPAL CA-630 to 37°C as it is quite viscous at room temperature. Make a 10% solution of IGEPAL CA-630 by adding 4ml of warmed IGEPAL (dispensed via a 5ml syringe attached to an 18 gauge needle) to 36ml of Milli-Q sterile dH_20. Vortex until solubilized. Add 2 grams AG501-X8 resin to the 40ml 10% IGEPAL solution. This resin will ‘deionize’ the solution. ‘Spent’ resin will be golden in color; the solution is still deionized if the resin is a combination of blue and golden beads. Store solution in an aluminum-wrapped conical tube for 6 months at 4°C.

100mM Dithiothreitol

Dissolve 100 mg dithiothreitol in 6.48 ml sterile Milli-Q H_20. Aliquot 1 ml into screwcap tubes. Store for 12 months at −20. ! **CAUTION** Wear gloves and work in fume hood during preparation irritating to skin, eyes and respiratory system.

DNaseI Storage Buffer

For 10 ml of Storage Buffer, combine 4.53 ml sterile Milli-Q H_20, 200 μl of 1 M Tris-HCl (pH7.5), 100 μl of 5M NaCl, 20 μl of 1M $CaCl_2$, 100 μl of 100 mM dithiothreitol, 50 μl of

20 mg/ml Pefabloc SC Plus and 5 ml molecular biology grade glycerol. Mix by inverting and place on ice. DNaseI Storage Buffer is made fresh as needed.

20mM Tris-HCl (pH 7.6)

50 mM NaCl

2 mM $MgCl_2$

2 mM $CaCl_2$

1 mM Dithiothreitol

0.1 mg/ml Pefabloc SC Plus

50% Glycerol

DNaseI 10u/μl

Gently solubilize on ice an entire bottle of DNaseI Type II from Bovine Pancreas in DNaseI Storage Buffer to a final concentration of 10u/μl. Dispense 50 to 100 μl aliquots into 1.5 ml microfuge tubes and store for 12 months at −20°C. ▲ **Critical** DNaseI is easily denatured. Do not vortex or pipet vigorously as this will decrease DNaseI enzymatic activity.

Buffer A

For 1 liter Buffer A add 15 ml of 1M Tris-HCl (pH 8.0), 3 ml of 5M NaCl, 20 ml of 3M KCl, 2 ml of 0.5M EDTA (pH 8.0), 1 ml of 0.5M EGTA (pH 8.0), and 1 ml of 0.5M spermidine to 918 ml sterile Milli-Q water. Mix thoroughly by inverting and store at 4°C. Discard after 1 week.

15mM Tris-HCl (pH 8.0)

15mM NaCl, 60 mM KCl

1mM EDTA (pH 8.0)

0.5mM EGTA (pH 8.0)

0.5mM spermidine

Stop Buffer

For 100 ml of buffer add 5 ml of 1M Tris-HCl (pH 8.0), 2 ml of 5M NaCl, 1 ml of 10% SDS and 20 ml of 0.5M EDTA (pH8.0) to 72 ml sterile Milli-Q water. Aliquot 10 ml into 15 ml tubes and stored at 4°C for 1 month. Note: SDS will precipitate out of solution. On day of use warm solution to 37°C for 5 minutes and mix until SDS goes back into solution. Add 20 μl of 0.5 M Spermidine and 6 μl of 0.5 M Spermine and hold at 37°C.

50 mM Tris-HCl (pH 8.0),

100 mM NaCl,

0.1% SDS,

100 mM EDTA (pH 8.0)

1mM spermidine

0.3mM spermine

1X DNaseI Digestion Buffer

In a 15 ml conical tube add 1ml 10X DNaseI Digestion Buffer to 9ml Buffer A. Place in 37°C water bath and allow to equilibrate to 37°C for 60 minutes prior to use. To be made fresh on the day of use.

9% sucrose solution

For 1 liter 9% sucrose solution, add 90g sucrose, 200 ml 5M NaCl, 20 ml 1M Tris-HCl (pH 8.0), 10 ml 0.5 M EDTA and sterile Milli-Q water to 1 liter.

COMMENTARY

Background Information

Over the last decade, the assembly of multiple complete genomes has led to an immense compilation of sequence data. This has also led to the eventual predicament--that of assigning functionality to elements within these genomes. Numerous diverse epigenomic profiling approaches have been developed to address this limitation. The most commonly used methods include ChIP-seq, which, given an antibody, can provide genomic coordinates for any transcription factor, co-activator or histone modification and RNA-seq which can reveal the complete transcriptional status within a cell line or tissue. These techniques, while useful and instructive, have their inherent biases. ChIP-seq is dependent on having a good antibody and knowing what transcription factor to query while RNA-seq identifies expressed genes but provides little insight on what controls their expression.

DNaseI has been used as an unbiased probe of accessible compartments within the higher-order, three-dimensional architecture of the genome. These ‘open’ regions, referred to as DNaseI hypersensitive sites (DHSs), invariably portend the regulatory apparatus of the genome. While the DNaseI assay and its variants have been described in the literature for over 30 years (Wu et. al., 1979a,b; Wu, 1980; Stalder, 1980; Emerson, et. al., 1985; Zimarino and Wu, 1987; Gross and Garrard 1988; Elgin, 1990; Dorschner et. al., 2004; Sabo et. al., 2006), they have been limited to analyzing specific genes or genomic regions using a time-consuming ‘region-by-region’ strategy.

We have developed DNaseI-seq as a robust assay, broadly applicable to any cell type or tissue to identify the totality of regulatory elements in a given genome. Its application has now yielded hundreds of DNaseI-seq datasets across a variety of species and scores of cell types and tissues (www.uwencode.org, www.roadmapepigenomics.org, and www.epigenomebrowser.org). Taken together, these data have identified millions of regulatory elements including promoters, enhancers and insulators. These data also highlight the remarkable cell-specific organization of accessibility landscapes which in turn help inform the connections between open chromatin and the cell-restricted expression of genes (Thurman et. al., 2012a; Neph et. al., 2012a; John et. al., 2011). This assay has been extended to achieve genomic DNaseI footprinting (Neph et. al., 2012a) at base-par

resolution which has helped uncover a number of novel cell-type exclusive DNA binding motifs for transcription factors that will likely play a role in the maintenance of cellular lineages, genome organization and gene expression.

The importance of generating high-quality DNaseI-digested samples is essential for the extraction of useful information from DNaseI-seq datasets. We have, therefore, provided both experimental and computation-based quality control (QC) steps that allow for the identification of high quality and high signal-to-noise DNaseI-seq samples and datasets (see below).

Critical Parameters and Troubleshooting

A successful DNaseI-seq experiment is predicated on a few key steps in the protocol. In this section, we highlight critical steps that need attention and optimization.

Preparation of nuclei—The preparation of high quality nuclei is a necessary pre-requisite for the generation of first-rate sequencing data from a DNaseI-seq experiment. The amount of detergent needed to lyse cells to make nuclei varies significantly across cell lines and can range from 0.01% to 0.5% IGEPAL. It is, therefore, imperative that test studies be performed to calibrate the optimum amount of detergent that maximizes nuclear recovery with minimal damage to structure and integrity. The efficiency of detergent lysis can be examined under a phase-contrast light microscope. Small numbers of cells can be used for standardization. While 5–10 million cells are used in a typical DNaseI experiment, only 1,000–10,000 cells need be used in optimization studies.

DNaseI digestion of nuclei—A key step in the DNaseI-seq assay is finding the concentration of enzyme that results in the release of the most accessible compartments of the genome (DNaseI hypersensitive sites) while minimizing digestion of bulk chromatin. DNaseI concentrations that are too low or too high can lead to sporadic digestion or overdigestion of chromatin. Cleavage by the enzyme, therefore, needs to be precisely optimized using a range of DNaseI concentrations (Fig. 2). Ideal DNaseI concentrations can vary significantly between different cell and tissue types. Although both times of digestion and amounts of enzyme can be adjusted in this assay, we have found that varying enzyme concentrations with a fixed digestion time (3 min at 37°C) provides the best approach for standardization. The steps detailed in BASIC PROTOCOL 4 (Quality Control) and below list the detection methodologies used to evaluate the extent of DNaseI digestion.

Quality control

QC gels: We suggest two simple quality control steps be implemented to serve as important checkpoints prior to committing to sequencing. In the first QC step, an agarose gel is used to electrophorese undigested DNA and DNAs digested with several concentrations of DNaseI. The gel profiles (Fig. 2 upper and lower panels) provide a ‘reality check’ in helping determine levels of endogenous nuclease activity and the extent of digestion with DNaseI. Extensive smearing below the parental genomic DNA band is indicative of significant endogenous nuclease activity, overdigested chromatin or sick cells.

We also recommend digesting naked, genomic DNA with DNaseI to generate similarly sized fragments as DNaseI digested nuclear chromatin. Sequencing these fragments reveal the intrinsic sequence preferences of the enzyme. Our genome-wide experiments (Neph et. al., 2012a) indicate that the accessibility landscapes observed in a given cell type are not significantly influenced by enzymatic sequence preferences.

Assessment of digestion specificity using quantitative PCR—In the second QC step, a quantitative determination of DNaseI digestion can be made using digested DNAs and primers to hypersensitive sites commonly observed in many cell types (genomic co-ordinates and primer sequences available in Supplementary Table 3 in Sabo et. al., 2006). Measurements of amplicon efficiencies at DNaseI hypersensitive sites, DNaseI-refractive sites and undigested genomic DNAs allows for the identification of quality samples with high signal to noise ratios (Fig. 3) (Dorschner et. al., 2004). These samples can then be queued for size fractionation and processed into sequencing libraries.

Size-fractionation of DHSs—An incorrectly prepared sucrose cushion (see REAGENTS AND SOLUTIONS) can result in poor size separation of DNaseI digested fragments (~1–2% of the genome) from bulk undigested DNA. Make certain that NaCl is present at a final concentration of 1M. The manual collection of samples requires gentle pipetting from the top of the collection tube with minimal disturbance of the gradient. Automated liquid collection systems can greatly facilitate the reproducibility and ease of collecting fractions.

DNaseI—There can be significant lot to lot variability of DNaseI stocks. Using identical amounts of nuclease is not a reliable method in estimating cutting efficiencies of nuclear chromatin. The QC steps (described in BASIC PROTOCOL 4) provide useful checks in evaluating enzymatic activity before proceeding with library construction and sequencing.

Numbers of cells—Only a few thousand cells are needed for optimization experiments, however, 5–10 million cells are required to generate the 10–50 ng of DHS DNA needed to capture the complex accessibility landscape in a mammalian genome. These numbers of cells are easy to obtain for cell lines or abundant tissues but can become an issue for limiting biological samples.

Reproducibility—While the DNaseI-seq assay is a robust assay, we would recommend generating biological replicates using independently-derived batches of cells. In general, we have found negligible variation across nuclear preparations in experimental repeats (Sabo et. al., 2006).

Additional validation—Further corroboration of DHSs identified by DNaseI-seq approaches can be substantiated by ‘gold-standard’ methodologies such as the Southern hybridization assay. The indirect end-labels needed to map hypersensitive sites depend on the exact genomic region being probed. An analysis of both methods, in a study that examined megabases of DNA across 12 cell lines and 7500 independent Southern blots, validated the strong agreement between both approaches and confirmed the DNaseI-seq assay as highly reliable, sensitive and specific in identifying DHSs (Sabo et. al., 2006).

Data analysis—One of the most exigent aspects of any genome-wide experiment is the computational processing and sophisticated analysis needed to generate interpretable data. It is beyond the purview of this unit to provide algorithmic details for the identification of nuclease sensitive domains. We guide the investigator to tools that we and others have generated that will aid in the discovery of regions of enrichment obtained in a DNaseI-seq assay or any other epigenomic profiling method (http://www.uwencode.org/proj/hotspot-ptih) (Thurman et. al., 2012b, c; Zhang et. al., 2008; Boyle et. al., 2008; Ji et. al., 2011; Rozowsky et. al., 2009; Valouev et. al., 2008; Jothi et. al., 2008; Heinz et. al, 2010). Additional computational resources for processing, analysis and comparisons of DNaseI-seq with other genome-wide epigenomic datasets can be found at http://code.google.com/p/bedops (Neph et. al., 2012b).

Final quality assessment—The generation of high quality and high signal-to-noise datasets is critical for meaningful comparisons across DNaseI-seq experiments. We have, therefore, introduced an important quality measure called SPOT (Signal Portion Of Tags) that uses tag-enriched regions identified by Hotspot (or any other peak caller) to quantify a 'SPOT' score, which is computed as the percentage of all mapped reads that fall within hotspots called on a fixed number of sub-sampled tags (5 million) from the original dataset (Thurman et. al., 2012b,c). We use a SPOT score cutoff of .4 or greater for DNaseI-seq data as this has been determined to mark high-quality samples (Fig. 5). Given the ability to barcode samples and sequence as many as a dozen reactions per lane on the Illumina platform, we customarily perform pilot scale sequencing of multiple DNaseI libraries (~ 5 million unique reads per library) and then compute a SPOT score for each library. A DNaseI library with a good SPOT score is then promoted into the pipeline for expansion sequencing. SPOT is a routinely used parameter for quality evaluation of data generated by sequencing consortia such as ENCODE and Roadmap.

Anticipated Results

Starting with 5–10 million nuclei of DNaseI digested chromatin, one can anticipate purifying 10–50 ng of DHS DNA after size-fractionation. Once the DHS libraries have been sequenced and processed, they can be best visualized using a genome browser such as the UCSC Genome Browser (University of California Santa Cruz). DNaseI hypersensitive sites can be viewed as peaks of DNaseI cleavage densities of varying accessibilities across the entire genome (an example for a specific region across a small subset of cell types is shown in (Fig. 6).

Our laboratory and sequencing consortia have generated 100s of DNaseI profiles in primary and established cell lines as well as in mouse and human tissues (links to data are listed above). An analysis of these data have established numerous interesting connections between DHSs and cell-specific gene expression, DNA methylation, chromatin remodeling, transcription factor occupancy, long range interactions and human disease states (Thurman et. al., 2012a, Neph et. al., 2012a, Humbert et. al., 2012, John et. al., 2011).

Time Consideration

This detailed protocol provides the investigator with a comprehensive set of instructions to proceed from cells to sequencing libraries. Starting with BASIC PROTOCOL 1, the entire procedure up to (but not including) high-throughput sequencing takes approximately 4 days. Specifically, the preparation and digestion of nuclei will take 3–4 h on day 1. On day 2–3, DNA clean-up and running and fractionating sucrose gradients will consume ~ 27 h (BASIC PROTOCOL 2). This includes 24 hrs of centrifugation. Preparation of DHS libraries for sequencing on the Illumina Genome Analyzer takes ~ 8 h on days 3–4 (BASIC PROTOCOL 3). Quality control checks (QC gel and/or q-PCR) can be performed on day 2 after Basic Protocol 1 and can be performed concurrently with sample centrifugation.

BASIC PROTOCOL 1: Day 1

Buffer preparation	1 h
Nuclei isolation	1–2 h
DNaseI digestion of nuclei	1–2 h

BASIC PROTOCOL 2: Day 2–3

DNA clean-up	2 h
Running sucrose gradients	24 h
Fractionation of gradients	2 h

BASIC PROTOCOL 3: Day 3–4

Library preparation	8 h

BASIC PROTOCOL 4: Day 2

QC gel	1–2 h
Quantitative profiling	4–5 h (optional)

Acknowledgments

We thank Rachel Ludwig for help with the preparation of figures and members of the Stamatoyannopoulos laboratory for their valuable comments and insight. Parts of this unit have been published in Nature Methods (P. Sabo et. al., Nature Methods 3:511-8 (2006)) and Nature Protocols (S. John, P. Sabo, et. al., in press).

References

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008; 132:311–22. [PubMed: 18243105]

Dorschner MO, Hawrylycz M, Humbert R, Wallace JC, Shafer A, Kawamoto J, Mack J, Hall R, Goldy J, Sabo PJ, Kohli A, Li Q, McArthur M, Stamatoyannopoulos JA. High-throughput localization of

functional elements by quantitative chromatin profiling. Nat Methods. 2004; 1:219–25. [PubMed: 15782197]

Elgin SC. Chromatin structure and gene activity. Curr Opin Cell Biol. 1990; 2:437–45. [PubMed: 2198896]

Emerson BM, Lewis CD, Felsenfeld G. Interaction of specific nuclear factors with the nuclease-hypersensitive region of the chicken adult beta-globin gene: nature of the binding domain. Cell. 1985; 41:21–30. [PubMed: 2986843]

Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. Annu Rev Biochem. 1988; 57:159–97. [PubMed: 3052270]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–89. [PubMed: 20513432]

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Widespread variation in regulatory DNA associated with common human disease. Science. 2012; 337:1190–95. [PubMed: 22955828]

Ji H, Jiang H, Ma W, Wong WH. Using CisGenome to analyze ChIP-chip and ChIP-seq data. Curr Protoc Bioinformatics. 2011; Chapter 2(Unit2.13)

John S, Sabo PJ, Thurman RE, Sung MH, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat Genet. 2011; 43:264–8. [PubMed: 21258342]

Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. 2008; 36:5221–31. [PubMed: 18684996]

Neph S, Vierstra J, Stergachis AB, Reynolds AP, et al. An expansive human regulatory lexicon encoded in transcription factor footprints. Nature. 2012a; 489:83–90. [PubMed: 22955618]

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, Sandstrom R, Humbert R, Stamatoyannopoulos JA. BEDOPS: high-performance genomic feature operations. Bioinformatics. 2012b; 28:1919–20. [PubMed: 22576172]

Sabo PJ, Humbert R, Hawrylycz M, Wallace JC, Dorschner MO, McArthur M, Stamatoyannopoulos JA. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. Proc Natl Acad Sci U S A. 2004a; 101:4537–42. [PubMed: 15070753]

Sabo PJ, Hawrylycz M, Wallace JC, Humbert R, Yu M, Shafer A, Kawamoto J, Hall R, Mack J, Dorschner MO, McArthur M, Stamatoyannopoulos JA. Discovery of functional noncoding elements by digital analysis of chromatin structure. Proc Natl Acad Sci U S A. 2004b; 101:16837–42. [PubMed: 15550541]

Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A, Weaver M, Shafer A, Lee K, Neri F, Humbert R, Singer MA, Richmond TA, Dorschner MO, McArthur M, Hawrylycz M, Green RD, Navas PA, Noble WS, Stamatoyannopoulos JA. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat Methods. 2006; 3:511–8. [PubMed: 16791208]

Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol. 2009; 27:66–75. [PubMed: 19122651]

Stalder J, Larsen A, Engel JD, Dolan M, Groudine M, Weintraub H. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. Cell. 1980; 20:451–60. [PubMed: 7388947]

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]

Thurman RE, Shoresh N, Rynes E, John S, Sabo PJ, Navas P, Stamatoyannopoulos JA. An automated quality measurement for epigenomics assays. in prep.

Thurman RE, Hawrylycz M, Kuehn S, Haugen E, Stamatoyannopoulos JA. Hotspot: a scan-statistic algorithm for identifying enriched regions of short-read sequence tags. in prep.

Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods. 2008; 5:829–34. [PubMed: 19160518]

Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SC. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell. 1979a; 16:797–806. [PubMed: 455449]

Wu C, Wong YC, Elgin SC. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell. 1979b; 16:807–14. [PubMed: 455450]

Wu C. The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature. 1980; 286:854–60. [PubMed: 6774262]

Zimarino V, Wu C. Induction of sequence-specific binding of Drosophila heat shock activator protein without protein synthesis. Nature. 1987; 327:727–30. [PubMed: 3600771]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]

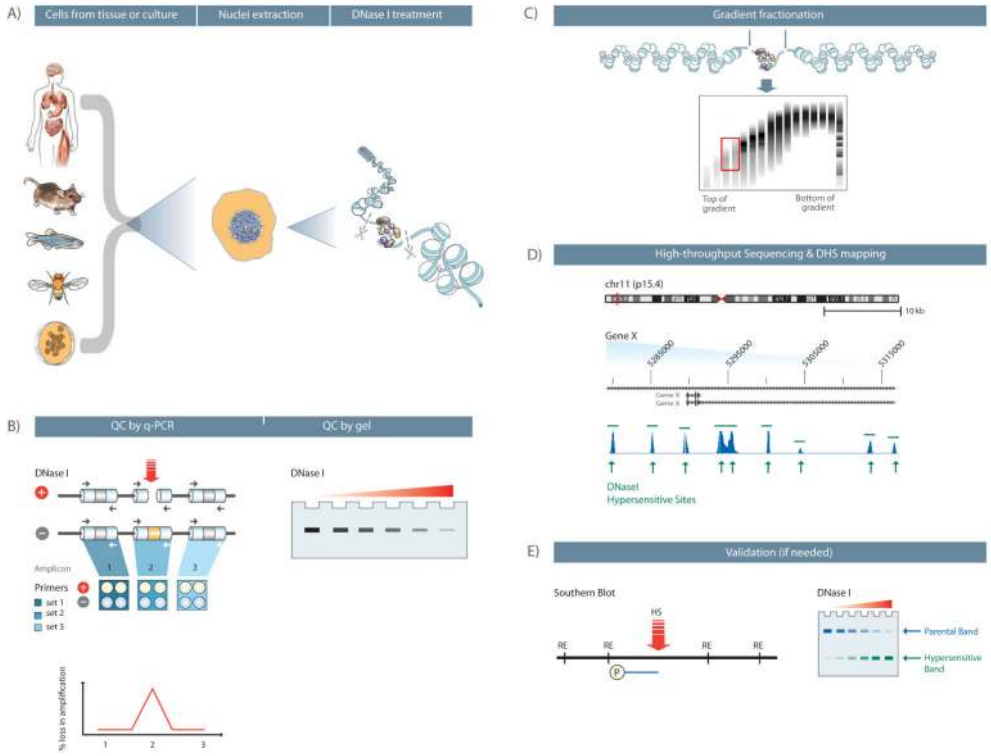

Figure 1. A schematic of the DNaseI-seq assay: from cells to hypersensitive sites

An overall outline of the DNaseI assay starting with (A) making nuclei from cells in culture and lightly digesting chromatin with DNaseI (B) QC checkpoints to assess digestion (C) size fractionation of DNaseI hypersensitive sites (D) high-throughput sequencing of DHS libraries and (E) validation of DNaseI-seq obtained data with traditional blotting approaches.

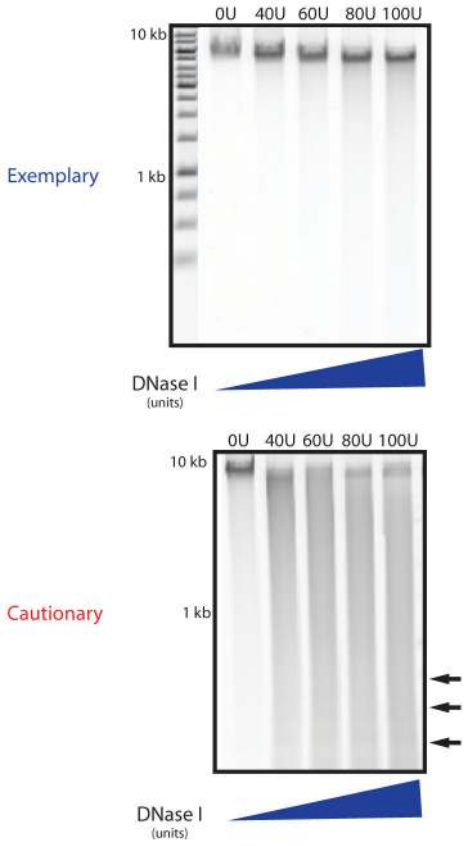

Figure 2. Quality control: checkpoint prior to high-throughput sequencing

QC gels. DNaseI-digested, deproteinized samples are warmed to 55°C and 10 μl from each of the digested samples are run out on 1% agarose gels. The light smear observed with 40U and 60U of DNaseI (top panel) is indicative of an appropriately digested sample. The extensive digestion, smearing and laddering observed with the 40U and 60U digests (lower panel) is symptomatic of a poor sample.

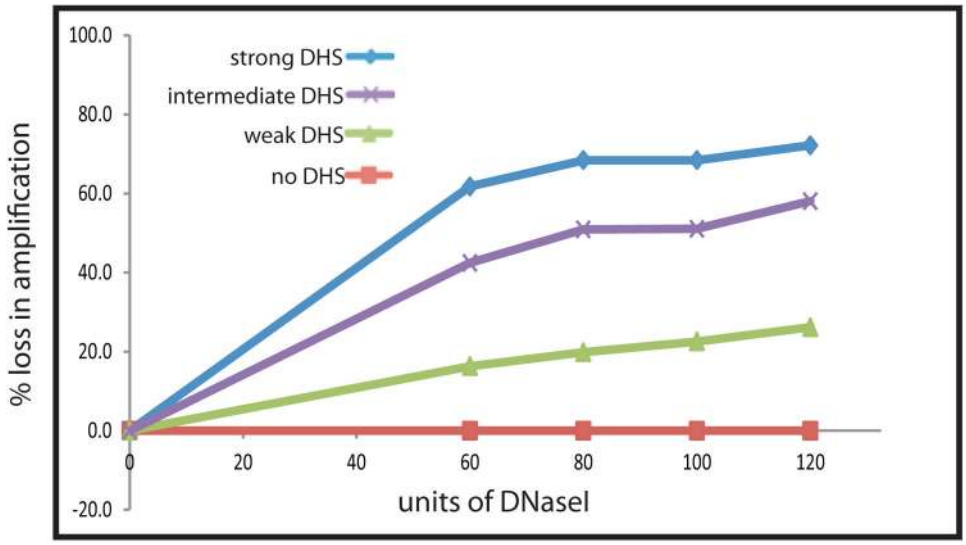

Figure 3. Quality control: checkpoint prior to high-throughput sequencing

Q-PCR validation of DHSs. Primers are designed to strong, intermediate or weak DNaseI sites as well as to control DNaseI-insensitive regions (see Sabo 2006 for list of regions and primers). A DNaseI sensitivity index is obtained by measuring the loss in amplification at the various regions analyzed across a range of DNaseI concentrations.

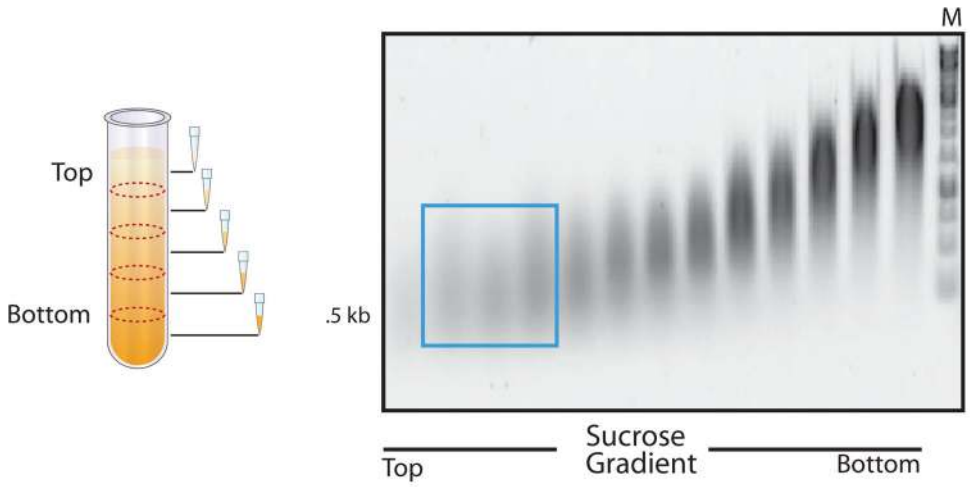

Figure 4. Size-fractionation of DNaseI hypersensitive sites

Sucrose gradient profiles. Small DNaseI fragments generated from the 'two-hit' digestion kinetics of DNaseI are isolated by size fractionation on sucrose gradients.

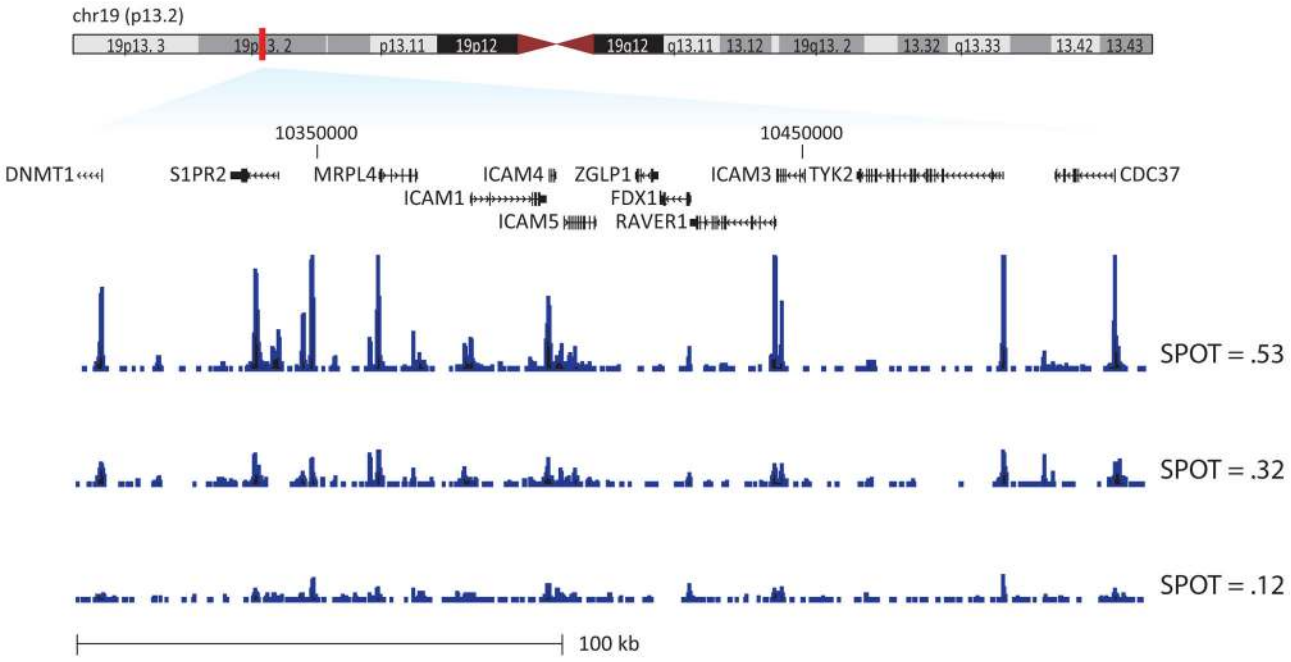

Figure 5. SPOT scores accurately predict high quality DNaseI-seq data sets

Three independent DNaseI-seq experiments, performed in the same cell line (lung), are ranked from high (0.53) to low SPOT scores (0.12). SPOT scores are computed as the fraction of all tags that fall within hotspots and are validated by reproducibility and biological reference standards [27]. Empirical observations indicate that DNaseI-seq SPOT scores greater than .4 represent high-quality datasets.

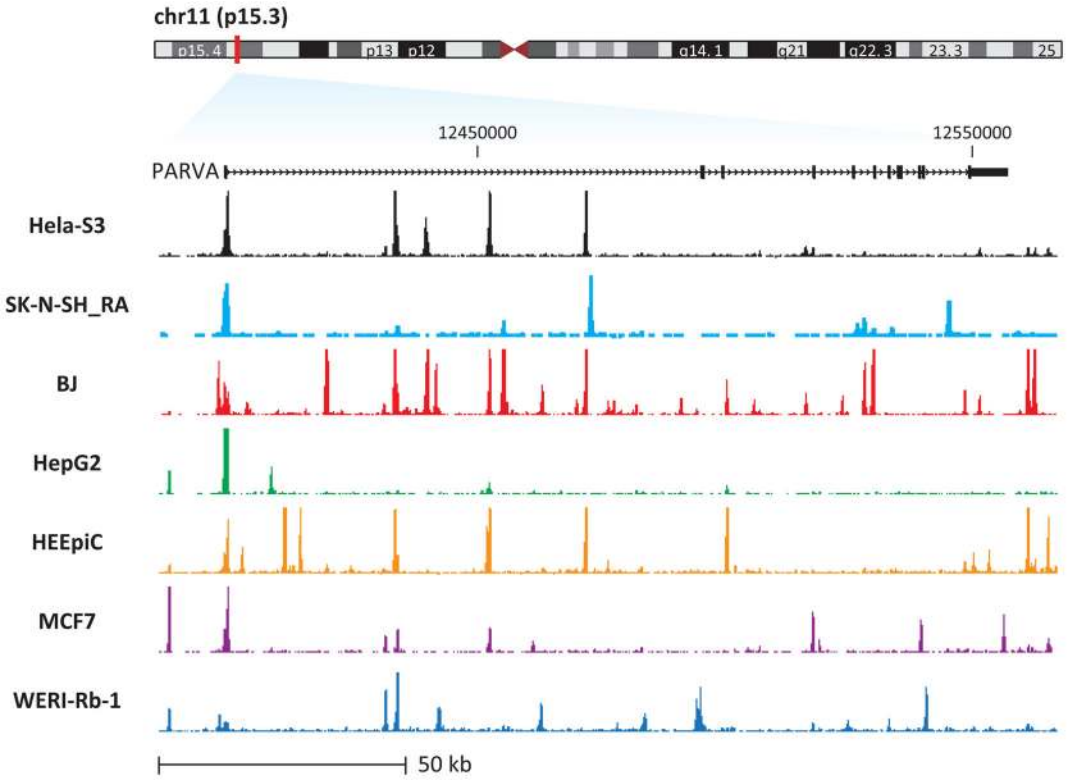

Figure 6. Anticipated results

Cell-specific chromatin accessibility. DNaseI cleavage sites are shown for a 100 kb region across a few primary and transformed cell types. While some DNaseI sites are shared across many cell types, the vast majority of DHSs exhibit cell-type specific distribution patterns (see Thurman et. al., 2012a for details).