

SUMMARY

Provided herein is a method for analyzing polynucleotides such as genomic DNA. In certain embodiments, the method comprises: (a) treating chromatin isolated from a population of cells with a transposase and molecular tags to produce tagged fragments of polynucleotides; (b) sequencing a portion of the tagged fragments to produce a plurality of sequence reads; and (c) making an epigenetic map of a region of the genome of the cells by mapping information obtained from the sequence reads to the region.

In one aspect of the present disclosure, there is provided a method for analyzing a biological sample, comprising:

(a) contacting chromatin of a genome region of said biological sample with an insertional enzyme complex to produce tagged nucleic acid molecules, wherein said insertional enzyme complex does not comprise an antibody specific to a protein that is part of said chromatin;

(b) performing a nucleic acid assay on said tagged nucleic acid molecules or derivatives thereof, to provide sequence information of said tagged nucleic acid molecules or derivatives thereof.

In some cases, the information is obtained using the nucleotide sequences at the beginning and, optionally, the end of a sequence read. In some cases, the information mapped in (c) is selected from one or more of: (i) cleavage sites for the transposase; (ii) the sizes of the fragments produced in step (a); (iii) sequence read length; (iii) the positions of sequence reads of a defined range in length; and (iv) sequence read abundance. In some instances, the fragments of a defined size range are nucleosome-free fragments.

In some instances, the epigenetic map shows one or more of: (i) a profile of chromatin accessibility along the region; (ii) DNA binding protein occupancy for a binding site in the region; (iii) nucleosome-free DNA in the region; (iv) positioning of nucleosomes along the region; and/or (v) chromatin states. In some cases, the method can further comprise measuring global occupancy of a binding site for the DNA binding protein. The DNA binding protein can, for example, be a transcription factor.

In some cases, the population of cells can be composed of about 500 to 100,000 cells. The cells can be isolated from an individual, such as from the blood of the individual. In some examples, the cells can be of the same cell type. In some examples, the cells can be FACS-selected cells.

In some instances, the treating step (a) can comprise: isolating nuclei from the population of cells; and combining the isolated nuclei with the insertional enzyme complex,

wherein the combining results in both lysis of the nuclei to release the chromatin and production of the tagged fragments of genomic DNA. In some examples, the transposase can be derived from Tn5 transposase. In other examples, the transposase can be derived from MuA transposase. In further examples, the transposase can be derived from Vibhar
5 transposase (e.g. from *Vibrio harveyi*).

The present disclosure also provides a method for comparing two samples comprising: (a) analyzing a first population of cells to produce a first epigenetic map; and (b) analyzing a second population of cells to produce a second epigenetic map; and (c) comparing the first epigenetic map to the second epigenetic map. For example, the first

population of cells and the second population of cells can be collected from the same individual at different times. Alternatively, the first population of cells and the second population of cells can be different populations of cells collected from different individuals.

5 The present disclosure further provides a diagnostic method, comprising: analyzing chromatin from a patient to produce an epigenetic map; and providing a diagnosis or prognosis based on the epigenetic map.

10 The present disclosure provides a method for determining accessibility of a polynucleotide at a site, wherein the polynucleotide is from a cell sample, comprising: (a) inserting a plurality of molecular tags with an insertional enzyme into the polynucleotide; and (b) using the molecular tags to determine accessibility at the site. The method can further comprise using the determined accessibility to identify one or more proteins that are bound to the polynucleotide at the site. In some cases, at least one of the proteins is a transcription factor. The method can also comprise using the molecular tags to generate an accessibility map of the polynucleotide.

15 The present disclosure also provides a method for analyzing the three-dimensional structure of a polynucleotide from a cell sample, comprising: (a) inserting a plurality of molecular tags with an insertional enzyme into the polynucleotide; and (b) using the molecular tags to analyze the three-dimensional structure of the polynucleotide. In some cases, the insertional enzyme can comprise two or more enzymatic moieties wherein each of the enzymatic moieties inserts a common sequence into the polynucleotide. The enzymatic moieties can be linked together. The common sequence can comprise a common barcode. The enzymatic moieties can comprise transposases. The polynucleotide can be fragmented into a plurality of fragments during step (a), wherein the fragments comprising the common barcode are determined to be in proximity in the three-dimensional structure of the polynucleotide.

20 The polynucleotide can be fragmented into a plurality of fragments during the insertion. The method can further comprise amplifying the fragments. The accessibility can be determined by sequencing the fragments and thereby generating a plurality of sequencing reads. The fragments can, for example, be sequenced by a high-throughput sequencing technique. The method can further comprise normalizing the sequencing reads based on the sequence insertion preference of the insertional enzyme. The length of the sequenced reads can also be used to determine a chromatin state annotation.

30 The cell sample can be permeabilized to allow access for the insertional enzyme. In some cases, the nuclei in the cell sample can be minimally perturbed during the

a methylated or hydroxymethylated DNA. The affinity tag can also be a single-stranded nucleic acid, which may be optionally bound to a target nucleic acid. The insert element can be bound to the insertional enzyme and the insertional enzyme is bound to the polynucleotide. The polynucleotide can be further bound to a plurality of association molecules. The association molecules can comprise proteins such as, for example, histones.

The present disclosure further provides kits. The kit can comprise: (a) reagents for isolating nuclei from a population of cells; (b) an insertional enzyme complex, and (c) transposase reaction buffer. In some cases, the components of the kit can be configured such that, combining the reaction buffer, transposon tags and adaptors with nuclei *in vitro* results in both lysis of the nuclei to release chromatin and production of tagged fragments of genomic DNA. The kit can also comprise: a cell lysis buffer; an insertional enzyme comprising an affinity tag; and an insert element comprising a nucleic acid, wherein the nucleic acid comprises a predetermined sequence. The kit can further comprise: a cell lysis buffer; an insertional enzyme comprising two or more enzymatic moieties, wherein the enzymatic moieties are linked together; and (c) an insert element. The affinity tag can be an antibody, which can optionally bind to a transcription factor, a modified nucleosome, and/or a modified nucleic acid. The modified nucleic acid can be, for example, a methylated or hydroxymethylated DNA. The affinity can also be a single-stranded nucleic acid, which may be optionally bound to a target nucleic acid.

These and other features of the present teachings are set forth herein.

INCORPORATION BY REFERENCE

All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE FIGURES

The skilled artisan will understand that the drawings, described below, are for illustration purposes only. The drawings are not intended to limit the scope of the present teachings in any way.

FIGS. 1A-1C: ATAC-seq is a sensitive, accurate probe of open chromatin state. **(a)** ATAC-seq reaction schematic. Transposase (green), loaded with sequencing adapters (red

CLAIMS

What is claimed is:

1. A method for analyzing a biological sample, comprising:
 - (a) contacting chromatin of a genome region of said biological sample with an insertional enzyme complex to produce tagged nucleic acid molecules, wherein said insertional enzyme complex does not comprise an antibody specific to a protein that is part of said chromatin;
 - (b) performing a nucleic acid assay on said tagged nucleic acid molecules or derivatives thereof, to provide sequence information of said tagged nucleic acid molecules or derivatives thereof.
2. The method of claim 1, further comprising generating a representation of epigenetic features of said genome region at least in part by mapping said sequence information to said genome region.
3. The method of claim 1, wherein said biological sample is a cell or a nucleus.
4. The method of claim 3, further comprising lysing said cell or said nucleus.
5. The method of claim 3, further comprising permeabilizing said cell or said nucleus to permit said insertional enzyme complex to access said chromatin.
6. The method of claim 5, wherein said permeabilizing is configured to minimally perturb said chromatin of said cell or said nucleus.
7. The method of claim 5, wherein said cell or said nucleus is permeabilized with a reagent configured to permit said insertional enzyme complex to access said chromatin.
8. The method of claim 7, wherein said reagent comprises NP40, digitonin, tween, streptolysin, or a cationic lipid.

9. The method of claim 5, wherein said cell or said nucleus is permeabilized by hypotonic shock or ultrasonication.
10. The method of claim 1, wherein said insertional enzyme comprises a transposase.
11. The method of claim 10, wherein said transposase is a Tn5 transposase or derived from a Tn5 transposase.
12. The method of claim 11, wherein said transposase is a hyperactive Tn5 transposase.
13. The method of claim 1, wherein said insertional enzyme complex comprises a first nucleic acid insert element comprising a first adapter sequence.
14. The method of claim 13, wherein said first adapter sequence comprises a first sequencing adapter sequence.
15. The method of claim 13, wherein said first adapter sequence comprises a barcode sequence.
16. The method of claim 13, wherein said first adapter sequence comprises a first primer sequence.
17. The method of claim 13, wherein said insertional enzyme complex further comprises a second nucleic acid insert element comprising a second adapter sequence.
18. The method of claim 17, wherein said second adapter sequence comprises a second sequencing adapter sequence.
19. The method of claim 17, wherein said second adapter sequence comprises a barcode sequence.
20. The method of claim 17, wherein said second adapter sequence comprises a second primer sequence.

21. The method of claim 1, further comprising subjecting said tagged nucleic acid molecules to one or more reactions.

22. The method of claim 21, wherein said one or more reactions comprise a nucleic acid amplification reaction.

23. The method of claim 22, wherein said nucleic acid amplification reaction is configured to add one or more functional sequences to said tagged nucleic acid molecules or derivatives thereof, wherein said one or more functional sequences are compatible with a selected next generation sequencing platform.