Electronically Filed

PRELIMINARY AMENDMENT	Attorney Docket No.	STAN-1111CON
	Confirmation No.	TBD
	First Named Inventor	GIRESI Paul
	Application Number	TBD
	Filing Date	HEREWITH
Address to: Mail Stop Amendment Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450	Group Art Unit	TBD
	Examiner Name	TBD
	Title: *"TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS"*	

Sir:

Prior to the examination of the above-referenced application on the merits, please enter the amendments below.

Amendments TO THE SPECIFICATION

Please amend the paragraph starting on line 10 of page 1 as shown below:

This application is a continuation of U.S. Patent Application No. 14/784,250, filed on October 13, 2015, which application is a § 371 national phase filing of International Application No. PCT/US2014/038825, filed on May 20, 2014, which ~~This~~ application claims the benefit of U.S. Provisional application serial no. 61/826,728, filed on May 23, 2013, which **applications are** ~~application is~~ incorporated by reference herein in **their** ~~its~~ entirety.

PARSE EX.1007.002

Amendments to the Claims

Please amend the claims as shown below. A complete listing of the claims, including their current status identifiers, is set forth below.

1-77. (Cancelled)

78. (New) 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; and

(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.

79. (New) The method of claim 78, 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.

80. (New) The method of claim 78, wherein said biological sample is a cell or a nucleus.

81. (New) The method of claim 80, further comprising lysing said cell or said nucleus.

82. (New) The method of claim 80, further comprising permeabilizing said cell or said nucleus to permit said insertional enzyme complex to access said chromatin.

83. (New) The method of claim 82, wherein said permeabilizing is configured to minimally perturb said chromatin of said cell or said nucleus.

84. (New) The method of claim 82, wherein said cell or said nucleus is permeabilized with a reagent configured to permit said insertional enzyme complex to access said chromatin.

PARSE EX.1007.003

85. (New) The method of claim 84, wherein said reagent comprises NP40, digitonin, tween, streptolysin, or a cationic lipid.

86. (New) The method of claim 82, wherein said cell or said nucleus is permeabilized by hypotonic shock or ultrasonication.

87. (New) The method of claim 78, wherein said insertional enzyme comprises a transposase.

88. (New) The method of claim 87, wherein said transposase is a Tn5 transposase or derived from a Tn5 transposase.

89. (New) The method of claim 88, wherein said transposase is a hyperactive Tn5 transposase.

90. (New) The method of claim 78, wherein said insertional enzyme complex comprises a first nucleic acid insert element comprising a first adapter sequence.

91. (New) The method of claim 90, wherein said first adapter sequence comprises a first sequencing adapter sequence.

92. (New) The method of claim 90, wherein said first adapter sequence comprises a barcode sequence.

93. (New) The method of claim 90, wherein said first adapter sequence comprises a first primer sequence.

94. (New) The method of claim 90, wherein said insertional enzyme complex further comprises a second nucleic acid insert element comprising a second adapter sequence.

95. (New) The method of claim 94, wherein said second adapter sequence comprises a second sequencing adapter sequence.

PARSE EX.1007.004

96. (New) The method of claim 94, wherein said second adapter sequence comprises a barcode sequence.

97. (New) The method of claim 94, wherein said second adapter sequence comprises a second primer sequence.

98. (New) The method of claim 78, further comprising subjecting said tagged nucleic acid molecules to one or more reactions.

99. (New) The method of claim 98, wherein said one or more reactions comprise a nucleic acid amplification reaction.

100. (New) The method of claim 99, 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.

PARSE EX.1007.005

Remarks Under 37 CFR § 1.115

Formal Matters

Claims 78-100 are pending.

Claims 1-77 are canceled. Claims 78-100 are added

No new matter has been added.

Certification Regarding Sequence Listing

Applicant submits herewith the attached Sequence Listing in .txt format. As set out in MPEP §2422.03(a), the Office has advised that if the sequence listing text file submitted via EFS-Web complies with the requirements of 37 CFR 1.824(a)(2)-(6) and (b) (i.e., is a compliant sequence listing ASCII text file), the text file will serve as both the paper copy required by 37 CFR 1.821(c) and the computer readable form (CRF) required by 37 CFR 1.821(e). Further, per MPEP §2422.03(a), neither (1) a second copy of the sequence listing in a PDF file; nor (2) a statement under 37 CFR 1.821(f) (indicating that the paper copy and CRF copy of the sequence listing are identical) should be submitted.

The Sequence Listing was prepared with the software PatentIn version 3.5, and conforms to the Patent Office guidelines. Applicant respectfully submits that the subject application is in adherence to 37 CFR §§ 1.821-1.825. I hereby certify that the enclosed submission includes no new matter.
Applicant respectfully submits that the present patent application is now in compliance with 37 CFR §§ 1.821-1.825.

Applicant submits that all of the claims are in condition for allowance, which action is requested. If the Examiner finds that a telephone conference would expedite the prosecution of this application, please telephone the undersigned at the number provided.

Respectfully submitted,
BOZICEVIC, FIELD & FRANCIS LLP

Date: July 24, 2018

By: /James S. Keddie /
James S. Keddie
Registration No. 48,920

Bozicevic, Field & Francis LLP
201 Redwood Shores Parkway, Suite 200
Redwood City, California 94065
Telephone: (650) 327-3400
Facsimile: (650) 327-3231

PARSE EX.1007.006

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2014 (27.11.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/189957 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2014/038825

(22) International Filing Date:
20 May 2014 (20.05.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/826,728 23 May 2013 (23.05.2013) US

(71) Applicant: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; 1705 El Camino Real, Palo Alto, California 94306-1106 (US).

(72) Inventors: **GIRESI, Paul**; 264 Durand Way, Palo Alto, California 94304 (US). **BUENROSTRO, Jason D.**; 3523 Altamont Way, Redwood City, California 94062 (US). **CHANG, Howard Y.**; 946 Casanueva Place, Stanford, California 94305 (US). **GREENLEAF, William J.**; 2372 Branner Drive, Menlo Park, California 94025 (US).

(74) Agent: **KEDDIE, James S.**; Bozicevic, Field & Francis LLP, 1900 University Avenue, Suite 200, East Palo Alto, California 94303 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: TRANSPOSITION INTO NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

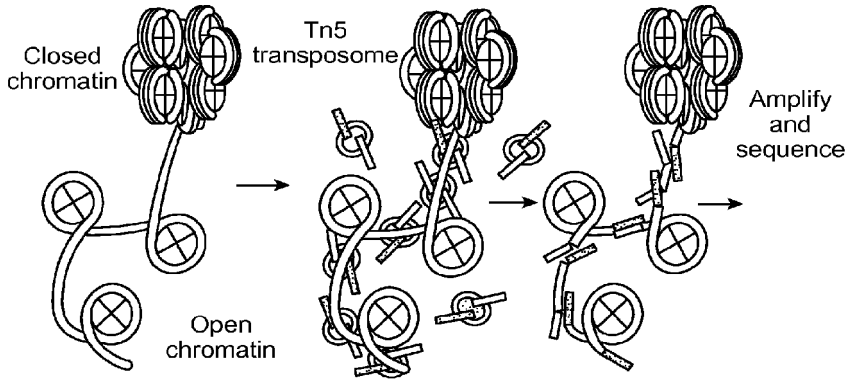


FIG. 1A

(57) **Abstract**: 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 an insertional enzyme complex to produce tagged fragments of genomic DNA; (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. A kit for performing the method is also provided.

WO 2014/189957 A2

PARSE EX.1007.007

TRANSPOSITION INTO NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

GOVERNMENT SUPPORT

This invention was made with Government support under contracts AI057229, HG000044, and NS073015 awarded by the National Institutes of Health. The Government has certain rights in the invention.

CROSS-REFERENCING

This application claims the benefit of U.S. Provisional application serial no. 61/826,728, filed on May 23, 2013, which application is incorporated by reference herein in its entirety.

BACKGROUND

Eukaryotic genomes are hierarchically packaged into chromatin, and the nature of this packaging plays a central role in gene regulation. Major insights into the epigenetic information encoded within the nucleoprotein structure of chromatin have come from high-throughput, genome-wide methods for separately assaying the chromatin accessibility ("open chromatin"), nucleosome positioning, and transcription factor (TF) occupancy. While published protocols exist, those methods require millions of cells as starting material, complex and time-consuming sample preparations, and cannot simultaneously probe the interplay of nucleosome positioning, chromatin accessibility, and TF binding. These limitations are problematic in three major ways: First, current methods can average over and "drown out" heterogeneity in cellular populations. Second, cells must often be grown *ex vivo* to obtain sufficient biomaterials, perturbing the *in vivo* context and modulating the epigenetic state in unknown ways. Third, input requirements often prevent application of these assays to well-defined clinical samples, precluding generation of "personal epigenomes" in diagnostic timescales. Provided herein are methods for analyzing polynucleotides, including their accessibility and their structure, that can overcome these limitation(s). Also provided are single-cell methods that can provide higher sensitivity and further information on chromatin accessibility, including cell-to-cell variability, to potentially enable its use as a biomarker.

PARSE EX.1007.008

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 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 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

PARSE EX.1007.009

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.

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.

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.

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 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.

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.

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

PARSE EX.1007.010

permeabilization. The cell sample can be permeabilized using a permeabilization agent including, but not limited to, NP40, digitonin, tween, streptolysin, and/or cationic lipids. The cell sample can also be permeabilized using hypotonic shock and/or ultrasonication.

The method can further comprise analyzing a disease state in a subject based on the accessibility of the specific site, wherein the cell sample is obtained from the subject. The cell sample and/or the polynucleotides can also be divided into a plurality of portions, which may be optionally divided based on the molecular tags. The method can further comprise analyzing a phenotype of the cell sample. In some cases, the phenotype can be correlated to the accessibility of the site.

The insertion can be facilitated by addition of one or more divalent cations. In some cases, the one or more divalent cations can comprise magnesium. In some cases, the one or more divalent cations can comprise manganese.

The cell sample can be obtained from a primary source. The cell sample can consist of less than about 500,000 cells, or even a single cell. The polynucleotide can be bound to a plurality of association molecules. The association molecules can comprise proteins, such as histones. The insertional enzyme can be a transposase. In some cases, the transposase can be derived from a Tn5 transposase. In other cases, the transposase can be derived from a MuA transposase. In further cases, the transposase can be derived from a Vibhar transposase (e.g. from *Vibrio harveyi*). In some cases, the molecular tags can comprise sequencing adaptors, which may further comprise a barcode label. The barcode label can comprise a unique sequence. In other cases, the molecular tags can comprise fluorescence tags. The insertional enzyme can further comprise an affinity tag, which may optionally be an antibody that binds to a transcription factor, a modified nucleosome, and/or a modified nucleic acid. The modified nucleic acid can, for example be a methylated or hydroxymethylated DNA. The affinity tag can also be a single-stranded nucleic acid, which may optionally bind to a target nucleic acid. The insertional enzyme can further comprise a nuclear localization signal.

The present disclosure also provides compositions. The composition can comprise a polynucleotide, an insertional enzyme and an insert element, wherein: the insert element comprises a nucleic acid comprising a predetermined sequence; and the insertional enzyme further comprises an affinity tag. The composition can also comprise a polynucleotide, an insertional enzyme and an insert element, wherein: the insertional enzyme comprises two or more enzymatic moieties; and the enzymatic moieties are linked together. The affinity tag can be an antibody, which may optionally be bound to a transcription factor, a modified nucleosome, and/or a modified nucleic acid. The modified nucleic acid can be, for example,

PARSE EX.1007.011

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

PARSE EX.1007.012

and blue), inserts only in regions of open chromatin (nucleosomes in grey) and generates sequencing library fragments that can be PCR amplified. (**b**) Approximate reported input material and sample preparation time requirements for genome-wide methods of open chromatin analysis. (**c**) A comparison of ATAC-seq to other open chromatin assays at a locus in GM12878 lymphoblastoid cells displaying high concordance. Lower ATAC-seq track was generated from 500 FACS-sorted cells.

FIGS. 2A-2B: ATAC-seq provides genome-wide information on chromatin compaction. (**a**) ATAC-seq fragment sizes generated from GM12878 nuclei (red) indicate chromatin-dependent periodicity with a spatial frequency consistent with nucleosomes, as well as a high frequency periodicity consistent with the pitch of the DNA helix for fragments less than 200 bp. (Inset) log-transformed histogram shows clear periodicity persists to 6 nucleosomes. (**b**) Normalized read enrichments for 7 classes of chromatin state previously defined.

FIGS. 3A-3E: ATAC-seq provides genome-wide information on nucleosome positioning in regulatory regions. (**a**) An example locus containing two transcription start sites (TSSs) showing nucleosome free read track, calculated nucleosome track (Methods), as well as DNase, MNase, and H3K27ac, H3K4me3, and H2A.Z tracks for comparison. (**b**) ATAC-seq (198 million paired reads) and MNase-seq (4 billion single-end reads from ref 23) nucleosome signal shown for all active TSSs (n=64,836), TSSs are sorted by CAGE expression. (**c**) TSSs are enriched for nucleosome free fragments, and show phased nucleosomes similar to those seen by MNase-seq at the -2, -1, +1, +2, +3 and +4 positions. (**d**) Relative fraction of nucleosome associated vs. nucleosome free (NFR) bases in TSS and distal sites (see Methods). (**e**) Hierarchical clustering of DNA binding factor position with respect to the nearest nucleosome dyad within accessible chromatin reveals distinct classes of DNA binding factors. Factors strongly associated with nucleosomes are enriched for chromatin remodelers.

FIGS 4A-4C: ATAC-seq assays genome-wide factor occupancy. (**a**) CTCF footprints observed in ATAC-seq and DNase-seq data, at a specific locus on chr1. (**b**) Aggregate ATAC-seq footprint for CTCF (motif shown) generated over binding sites within the genome (**c**) CTCF predicted binding probability inferred from ATAC-seq data, position weight matrix (PWM) scores for the CTCF motif, and evolutionary conservation (PhyloP). Right-most column is the CTCF ChIP-seq data (ENCODE) for this GM12878 cell line, demonstrating high concordance with predicted binding probability.

PARSE EX.1007.013

FIGS. 5A-5D: ATAC-seq enables real-time personal epigenomics. (**a**) Work flow from standard blood draws. (**b**) Serial ATAC-seq data from proband T-cells over three days. (**c**) Example of application of ATAC-seq data (green track) to prioritize candidate TF drug targets. Among identified TF binding sites proximal to cytokine gene *IL2* that can be targeted by FDA-approved drugs, only NFAT is engaged in proband T-cells. ATAC-seq footprint prediction is confirmed by alignment with published NFAT ChIP-seq data (blue track, data from ref [35]). (**d**) Cell type-specific regulatory network from proband T cells compared with GM12878 B-cell line. Each row or column is the footprint profile of a TF versus that of all other TFs in the same cell type. Color indicates relative similarity (yellow) or distinctiveness (blue) in T versus B cells. NFAT is one of the most highly differentially regulated TFs (red box) whereas canonical CTCF binding is essentially similar in T and B cells.

FIG. 6: ATAC-seq peak intensity correlates well with DNase-seq peak intensity. Peaks in Duke DNase-seq (down sampled to 60 x 10^6 reads), UW DNase-seq (40 x 10^6 reads), and ATAC-seq data (60 x 10^6 paired-end reads) were called using ZINBA (Rashid et al Genome Biol. 2011 12: R67). Because each data set has different read lengths we chose to filter for peaks within mappable regions (Duke DNase-seq = 20 bp reads, UW DNase-Seq = 36bp reads and ATAC-Seq = paired-end 50 bp reads). The log10(read intensity) was compared for (A) Duke DNase-seq and ATAC-seq, (B) UW DNase-seq and ATAC-seq, and (C) UW DNAse-seq and Duke DNase-seq. Technical reproducibility of ATAC-seq data is shown in D.

FIG. 7: ATAC-seq captures a large fraction of DNase identified peaks. Peaks were called for all data sets using ZINBA. The venn-diagram shows overlap of the peak calls between each method. Below: The majority of ATAC-seq reads are in intense peaks that intersect with Duke and UW DNase-seq peaks. The total fraction of reads within peaks called from ATAC-seq, UW DNase-seq, and Duke DNase-seq, as well as the intersections of these data are shown. More than 65% of reads from all three methods are found in the intersection of the three methods' peaks, suggesting that strong well-stereotyped peaks are detected by all methods. Table cell color is proportional to fraction of reads.

FIG. 8: Graphs of the number of reads overlapping the set of open chromatin regions identified by Duke DNase, UW DNase and FAIRE in GM12878 cells compared to a set of background regions, wherein to determine the read depth required for detecting open chromatin sites sensitivity and specificity was assessed at varying read depths, including

PARSE EX.1007.014

50k, 100k, 500k, 10 million and 50 million reads. The bottom graph shows the performance of ATAC-seq in GM12878 cells was assessed using 500, 5,000 or 50,000 cells as starting material.

FIG. 9: Tn5 insertion preferences in genomic DNA and chromatin. Nucleotide frequency scores represent the observed nucleotide frequency of each base, nucleotide frequencies are normalized to 1. The x=0 position represents the read start, and the dotted line represents the symmetry axis of the Tn5 dimer. We see no substantial differences between Tn5 insertion preferences between purified genomic DNA and human chromatin, suggesting that the local insertion preference into chromatin is identical to that found in naked genomic DNA. These reported sequence preferences are similar to those previously reported (main text ref. 11).

FIG. 10: Graph of the average intensity per base of each feature at every ATAC-seq peak; all ENCODE ChIP data was normalized to input; data has been processed using a sliding window of 200 peaks.

FIG. 11: ATAC-seq of various cell numbers. A representative UCSC genome browser track of data from different starting numbers of cells for ATAC-seq. This same locus is also shown in Fig. 1b of the main text. In order: 500 cells were isolated using FACS and two replicates of 500 cells and 5,000 cells were done by a simple dilution from cell culture. For comparison, the bottom track represents 50,000 cells, also show in Fig. 1b. This figure demonstrates that we are able to capture open chromatin sites from as few as 500 cells.

FIG. 12: Fitting nucleosome peaks in ATAC-seq fragment size distribution to enable nucleosome occupancy measurements. The observed fragment distribution was partitioned into four populations of reads – reads expected to originate from open DNA, and reads that span 1, 2 or 3 putative nucleosomes. To enable this partitioning of the data, the ATAC-seq fragment distribution was fit to the sum of 1) an exponential function for fragment distribution pattern at insert sizes below one nucleosome, and 2) 5 Gaussians to the distributions arising from protection from one, two, three, four and five nucleosomes. The sum of these fits is shown (black dotted line) is similar to the observed fragment distribution (blue line). Vertical dotted lines are boundaries for identification of fragments as originating from the nucleosome-free (<100bps), 1-nucleosome, 2-nucleosome and 3-nucleosome regions. Dotted lines were set to ensure that <10% of fragments originate from neighboring, as defined by our fit.

PARSE EX.1007.015

FIG. 13: Select set of transcription factor footprints detected by ATAC-seq in GM12878 cells. For the indicated transcription factors the aggregate signal of ATAC-seq reads were computed using CENTIPEDE on the genome-wide sets of sites matching the corresponding motif. Reads were calculated in the region +/-100 bp of the motif boundary. The vertical dashed lines indicate the boundaries of the motifs.

FIG. 14: Prediction of CTCF binding sites using ATAC-seq and DNase footprinting with CENTIPEDE. Prediction of CTCF binding sites was assessed using the genome-wide set of CTCF motifs sorted by the posterior probability reported by CENTIPEDE. Those overlapping CTCF ChIP-seq peaks were used as the positive set and all others were considered as the negative set. This yielded an area under the curve (AUC) of 0.92, which suggests specific and sensitive binding inference for CTCF. Duke DNase and UW DNase data were used with the same settings of CENTIPEDE, and ROC plots are shown. ATAC-seq data consisted of 198 x 10^6 paired reads, Duke DNase-comprised 245 x 10^6 reads, and UW DNase comprised 48 x 10^6 reads.

FIG. 15: T-cell specific NFAT regulation: Examples of T-cell-specific NFAT target genes predicted by ATAC-seq and confirmed by alignment with NFAT ChIP-seq (data from main text ref 35).

FIG. 16: ATAC-seq of FACS-purified cell populations from human blood. (A) From a standard blood draw, we used Fluorescence-Activated Cell Sorting (FACS) to purify CD4+ T-cells, CD8+ T-cells, and CD14+ monocytes. Each population generated successful ATAC-seq data (B) and revealed cell-type specific open chromatin sites at known lineage-specific genes.

FIG. 17: Detection of allele specific open chromatin in GM12878 cells with ATAC-seq. Using publicly available variant data, we measured the allele frequency in open chromatin regions at putative heterozygous loci. Because of potential for spurious heterozygous sites, we required more than two reads to validate the heterozygosity of the allele. Red points (n=167) are candidate allele specific open chromatin sites at $p<10^{-5}$, while grey (n=900) represent candidates at $p<0.01$. P-values were calculated using a Bayesian model developed by Audic et al (Genome Research 1997 7, 986–995).

FIG. 18: Transposases can serve as an open-chromatin stain. By loading Tn5 transposes with fluorescently labeled DNA adapters, transposition events, shown in green, are primarily localized to the nucleus, and exhibit a punctate pattern consistent with higher order organization.

PARSE EX.1007.016

FIG. 19: Single-cell ATAC-seq data from a single nucleus (blue) show clear peak at the expected positions of open-chromatin genome wide compared to 50,000 cells.

FIG. 20: Single cell insert length distribution matches that from 50,000 cells showing periodicity due to the presence of nucleosomes.

DEFINITIONS

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described.

All patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

The headings provided herein are not limitations of the various aspects or embodiments of the invention. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Markham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, N.Y. (1991) provide one of skill with the general meaning of many of the terms used herein. Still, certain terms are defined below for the sake of clarity and ease of reference.

The term "sample" as used herein relates to a material or mixture of materials, typically containing one or more analytes of interest. In one embodiment, the term as used in its broadest sense, refers to any plant, animal or viral material containing DNA or RNA, such as, for example, tissue or fluid isolated from an individual (including without limitation plasma, serum, cerebrospinal fluid, lymph, tears, saliva and tissue sections) or from in vitro cell culture constituents, as well as samples from the environment.

PARSE EX.1007.017

The term "nucleic acid sample," as used herein, denotes a sample containing nucleic acids. Nucleic acid samples used herein may be complex in that they contain multiple different molecules that contain sequences. Genomic DNA samples from a mammal (e.g., mouse or human) are types of complex samples. Complex samples may have more than about 10^4, 10^5, 10^6 or 10^7, 10^8, 10^9 or 10^{10} different nucleic acid molecules. A DNA target may originate from any source such as genomic DNA, or an artificial DNA construct. Any sample containing nucleic acid, e.g., genomic DNA from tissue culture cells or a sample of tissue, may be employed herein.

The term "mixture," as used herein, refers to a combination of elements, that are interspersed and not in any particular order. A mixture is heterogeneous and not spatially separable into its different constituents. Examples of mixtures of elements include a number of different elements that are dissolved in the same aqueous solution and a number of different elements attached to a solid support at random positions (i.e., in no particular order). A mixture is not addressable. To illustrate by example, an array of spatially separated surface-bound polynucleotides, as is commonly known in the art, is not a mixture of surface-bound polynucleotides because the species of surface-bound polynucleotides are spatially distinct and the array is addressable.

The term "nucleotide" is intended to include those moieties that contain not only the known purine and pyrimidine bases, but also other heterocyclic bases that have been modified. Such modifications include methylated purines or pyrimidines, acylated purines or pyrimidines, alkylated riboses or other heterocycles. In addition, the term "nucleotide" includes those moieties that contain hapten or fluorescent labels and may contain not only conventional ribose and deoxyribose sugars, but other sugars as well. Modified nucleosides or nucleotides also include modifications on the sugar moiety, e.g., wherein one or more of the hydroxyl groups are replaced with halogen atoms or aliphatic groups, or are functionalized as ethers, amines, or the like.

The term "nucleic acid" and "polynucleotide" are used interchangeably herein to describe a polymer of any length, e.g., greater than about 2 bases, greater than about 10 bases, greater than about 100 bases, greater than about 500 bases, greater than 1000 bases, greater than 10,000 bases, greater than 100,000 bases, greater than about 1,000,000, up to about 10^{10} or more bases composed of nucleotides, e.g., deoxyribonucleotides or ribonucleotides, and may be produced enzymatically or synthetically (e.g., PNA as described in U.S. Patent No. 5,948,902 and the references cited therein) which can hybridize with naturally occurring nucleic acids in a sequence specific manner analogous to that of two

PARSE EX.1007.018

naturally occurring nucleic acids, e.g., can participate in Watson-Crick base pairing interactions. Naturally-occurring nucleotides include guanine, cytosine, adenine, thymine, uracil (G, C, A, T and U respectively). DNA and RNA have a deoxyribose and ribose sugar backbone, respectively, whereas PNA's backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds. In PNA various purine and pyrimidine bases are linked to the backbone by methylenecarbonyl bonds. A locked nucleic acid (LNA), often referred to as inaccessible RNA, is a modified RNA nucleotide. The ribose moiety of an LNA nucleotide is modified with an extra bridge connecting the 2' oxygen and 4' carbon. The bridge "locks" the ribose in the 3'-endo (North) conformation, which is often found in the A-form duplexes. LNA nucleotides can be mixed with DNA or RNA residues in the oligonucleotide whenever desired. The term "unstructured nucleic acid," or "UNA," is a nucleic acid containing non-natural nucleotides that bind to each other with reduced stability. For example, an unstructured nucleic acid may contain a G' residue and a C' residue, where these residues correspond to non-naturally occurring forms, i.e., analogs, of G and C that base pair with each other with reduced stability, but retain an ability to base pair with naturally occurring C and G residues, respectively. Unstructured nucleic acid is described in US20050233340, which is incorporated by reference herein for disclosure of UNA.

The term "oligonucleotide" as used herein denotes a single-stranded multimer of nucleotide of from about 2 to 200 nucleotides, up to 500 nucleotides in length. Oligonucleotides may be synthetic or may be made enzymatically, and, in some embodiments, are 30 to 150 nucleotides in length. Oligonucleotides may contain ribonucleotide monomers (i.e., may be oligoribonucleotides) or deoxyribonucleotide monomers, or both ribonucleotide monomers and deoxyribonucleotide monomers. An oligonucleotide may be 10 to 20, 21 to 30, 31 to 40, 41 to 50, 51 to 60, 61 to 70, 71 to 80, 80 to 100, 100 to 150 or 150 to 200 nucleotides in length, for example.

"Primer" means an oligonucleotide, either natural or synthetic, that is capable, upon forming a duplex with a polynucleotide template, of acting as a point of initiation of nucleic acid synthesis and being extended from its 3' end along the template so that an extended duplex is formed. The sequence of nucleotides added during the extension process is determined by the sequence of the template polynucleotide. Usually primers are extended by a DNA polymerase. Primers are generally of a length compatible with their use in synthesis of primer extension products, and are usually in the range of between 8 to 100 nucleotides in length, such as 10 to 75, 15 to 60, 15 to 40, 18 to 30, 20 to 40, 21 to 50, 22 to 45, 25 to 40, and so on. Typical primers can be in the range of between 10-50 nucleotides long, such as

PARSE EX.1007.019

15-45, 18-40, 20-30, 21-25 and so on, and any length between the stated ranges. In some embodiments, the primers are usually not more than about 10, 12, 15, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, or 70 nucleotides in length.

Primers are usually single-stranded for maximum efficiency in amplification, but may alternatively be double-stranded. If double-stranded, the primer is usually first treated to separate its strands before being used to prepare extension products. This denaturation step is typically effected by heat, but may alternatively be carried out using alkali, followed by neutralization. Thus, a "primer" is complementary to a template, and complexes by hydrogen bonding or hybridization with the template to give a primer/template complex for initiation of synthesis by a polymerase, which is extended by the addition of covalently bonded bases linked at its 3' end complementary to the template in the process of DNA synthesis.

The term "hybridization" or "hybridizes" refers to a process in which a region of nucleic acid strand anneals to and forms a stable duplex, either a homoduplex or a heteroduplex, under normal hybridization conditions with a second complementary nucleic acid strand, and does not form a stable duplex with unrelated nucleic acid molecules under the same normal hybridization conditions. The formation of a duplex is accomplished by annealing two complementary nucleic acid strand region in a hybridization reaction. The hybridization reaction can be made to be highly specific by adjustment of the hybridization conditions (often referred to as hybridization stringency) under which the hybridization reaction takes place, such that two nucleic acid strands will not form a stable duplex, e.g., a duplex that retains a region of double-strandedness under normal stringency conditions, unless the two nucleic acid strands contain a certain number of nucleotides in specific sequences which are substantially or completely complementary. "Normal hybridization or normal stringency conditions" are readily determined for any given hybridization reaction. See, for example, Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, Inc., New York, or Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press. As used herein, the term "hybridizing" or "hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

A nucleic acid is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Moderate and high stringency hybridization conditions are known (see, e.g., Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons 1995 and Sambrook et al., Molecular Cloning: A

PARSE EX.1007.020

Laboratory Manual, Third Edition, 2001 Cold Spring Harbor, N.Y.). One example of high stringency conditions include hybridization at about 42 °C in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1 X SSC and 0.5% SDS at 42 °C.

The term "duplex," or "duplexed," as used herein, describes two complementary polynucleotide region that are base-paired, i.e., hybridized together.

The term "amplifying" as used herein refers to the process of synthesizing nucleic acid molecules that are complementary to one or both strands of a template nucleic acid. Amplifying a nucleic acid molecule may include denaturing the template nucleic acid, annealing primers to the template nucleic acid at a temperature that is below the melting temperatures of the primers, and enzymatically elongating from the primers to generate an amplification product. The denaturing, annealing and elongating steps each can be performed one or more times. In certain cases, the denaturing, annealing and elongating steps are performed multiple times such that the amount of amplification product is increasing, often times exponentially, although exponential amplification is not required by the present methods. Amplification typically requires the presence of deoxyribonucleoside triphosphates, a DNA polymerase enzyme and an appropriate buffer and/or co-factors for optimal activity of the polymerase enzyme. The term "amplification product" refers to the nucleic acids, which are produced from the amplifying process as defined herein.

The terms "determining," "measuring," "evaluating," "assessing," "assaying," and "analyzing" are used interchangeably herein to refer to any form of measurement, and include determining if an element is present or not. These terms include both quantitative and/or qualitative determinations. Assessing may be relative or absolute. "Assessing the presence of" includes determining the amount of something present, as well as determining whether it is present or absent.

The term "using" has its conventional meaning, and, as such, means employing, e.g., putting into service, a method or composition to attain an end. For example, if a program is used to create a file, a program is executed to make a file, the file usually being the output of the program. In another example, if a computer file is used, it is usually accessed, read, and the information stored in the file employed to attain an end. Similarly if a unique identifier, e.g., a barcode is used, the unique identifier is usually read to identify, for example, an object or file associated with the unique identifier.

PARSE EX.1007.021

The term "ligating," as used herein, refers to the enzymatically catalyzed joining of the terminal nucleotide at the 5' end of a first DNA molecule to the terminal nucleotide at the 3' end of a second DNA molecule.

A "plurality" contains at least 2 members. In certain cases, a plurality may have at least 2, at least 5, at least 10, at least 100, at least 100, at least 10,000, at least 100,000, at least 10^6, at least 10^7, at least 10^8 or at least 10^9 or more members.

If two nucleic acids are "complementary," they hybridize with one another under high stringency conditions. The term "perfectly complementary" is used to describe a duplex in which each base of one of the nucleic acids base pairs with a complementary nucleotide in the other nucleic acid. In many cases, two sequences that are complementary have at least 10, e.g., at least 12 or 15 nucleotides of complementarity.

An "oligonucleotide binding site" refers to a site to which an oligonucleotide hybridizes in a target polynucleotide. If an oligonucleotide "provides" a binding site for a primer, then the primer may hybridize to that oligonucleotide or its complement.

The term "strand" as used herein refers to a nucleic acid made up of nucleotides covalently linked together by covalent bonds, e.g., phosphodiester bonds. In a cell, DNA usually exists in a double-stranded form, and as such, has two complementary strands of nucleic acid referred to herein as the "top" and "bottom" strands. In certain cases, complementary strands of a chromosomal region may be referred to as "plus" and "minus" strands, the "first" and "second" strands, the "coding" and "noncoding" strands, the "Watson" and "Crick" strands or the "sense" and "antisense" strands. The assignment of a strand as being a top or bottom strand is arbitrary and does not imply any particular orientation, function or structure. The nucleotide sequences of the first strand of several exemplary mammalian chromosomal regions (e.g., BACs, assemblies, chromosomes, etc.) is known, and may be found in NCBI's Genbank database, for example.

The term "top strand," as used herein, refers to either strand of a nucleic acid but not both strands of a nucleic acid. When an oligonucleotide or a primer binds or anneals "only to a top strand," it binds to only one strand but not the other. The term "bottom strand," as used herein, refers to the strand that is complementary to the "top strand." When an oligonucleotide binds or anneals "only to one strand," it binds to only one strand, e.g., the first or second strand, but not the other strand.

The term "sequencing," as used herein, refers to a method by which the identity of at least 10 consecutive nucleotides (e.g., the identity of at least 20, at least 50, at least 100 or at least 200 or more consecutive nucleotides) of a polynucleotide is obtained.

PARSE EX.1007.022

The terms "next-generation sequencing" or "high-throughput sequencing" refer to the so-called parallelized sequencing-by-synthesis or sequencing-by-ligation platforms currently employed by Illumina, Life Technologies, and Roche, etc. Next-generation sequencing methods may also include nanopore sequencing methods or electronic-detection based methods such as Ion Torrent technology commercialized by Life Technologies or single-molecule fluorescence-based method commercialized by Pacific Biosciences.

The term "barcode sequence" or "molecular barcode," as used herein, refers to a unique sequence of nucleotides used to a) identify and/or track the source of a polynucleotide in a reaction and/or b) count how many times an initial molecule is sequenced (e.g., in cases where substantially every molecule in a sample is tagged with a different sequence, and then the sample is amplified). A barcode sequence may be at the 5′-end, the 3'-end or in the middle of an oligonucleotide. Barcode sequences may vary widely in size and composition; the following references provide guidance for selecting sets of barcode sequences appropriate for particular embodiments: Brenner, U.S. Pat. No. 5,635,400; Brenner et al, Proc. Natl. Acad. Sci., 97: 1665-1670 (2000); Shoemaker et al, Nature Genetics, 14: 450-456 (1996); Morris et al, European patent publication 0799897A1; Wallace, U.S. Pat. No. 5,981,179; and the like. In particular embodiments, a barcode sequence may have a length in range of from 4 to 36 nucleotides, or from 6 to 30 nucleotides, or from 8 to 20 nucleotides.

The term "in vitro" refers to a reaction that occurs in a vessel with isolated components, not in cells.

The term "distributed" in the context of cleavage sites that are distributed along the length of a target nucleic acid molecule, refers to insertions that are spaced from another along the length of the target nucleic acid molecule. There is no requirement that all of the insertions are spaced by the same amount. Rather, spacing between insertions may be random, semi-random, or not random.

The term "chromatin," as used herein, refers to a complex of molecules including proteins and polynucleotides (e.g. DNA, RNA), as found in a nucleus of a eukaryotic cell. Chromatin is composed in part of histone proteins that form nucleosomes, genomic DNA, and other DNA binding proteins (e.g., transcription factors) that are generally bound to the genomic DNA.

The term "treating," as used herein, refers to combining under conditions (e.g., a suitable temperature, time and conditions) that result in a reaction, e.g., cleavage.

PARSE EX.1007.023

The term "chromatin isolated from a population of cells," as used herein, refers to a source of chromatin that is caused to be made available. Isolated nuclei (which can be lysed to produce chromatin) as well as isolated chromatin (i.e., the product of lysed nuclei) are both considered types of chromatin isolated from a population of cells.

The term "transcription factor", as used herein, refers to any polypeptide that may act by itself or in combination with at least one other polypeptide to regulate gene expression levels. The term includes, but is not limited to, polypeptides that directly bind DNA sequences. Transcription factors can either increases or suppress expression levels. Examples of transcription factors include, but are not limited to Myc/Max, AP-1 (Jun, Fos, ATF), CREB, SMAD, HIF, ETS, ERG, ELK, STAT, estrogen receptor (ER), androgen receptor (AR), glucocorticoid receptor (GR), progesterone receptor (PR), NFκB, p53, OCT, SOX and PAX. The transcription factor may be a transcription factor identified by sequence analysis or a naturally-occurring reading frame sequence that has not been previously characterized as a transcription factor. The polypeptide may also be an artificially generated or chemically or enzymatically modified polypeptide.

The term "insertional enzyme complex," as used herein, refers to a complex comprising an insertional enzyme and two adaptor molecules (the "transposon tags") that are combined with polynucleotides to fragment and add adaptors to the polynucleotides. Such a system is described in a variety of publications, including Caruccio (Methods Mol. Biol. 2011 733: 241-55) and US20100120098, which are incorporated by reference herein.

The term "tagged fragments," as used herein, refers to polynucleotide fragments that are attached to tags.

The term "region," as used herein, refers to a contiguous length of nucleotides in a genome of an organism. A chromosomal region may be in the range of 1 bp to the length of an entire chromosome. In some instances, a region may have a length of at least 200 bp, at least 500 bp, at least 1 kb, at least 10 kb or at least 100 kb or more (e.g., up to 1 Mb or 10 Mb or more). The genome may be from any eukaryotic organism, e.g., an animal or plant genome such as the genome of a human, monkey, rat, fish or insect.

The term "epigenetic map," as used herein, refers to any representation of epigenetic features, e.g., sites of nucleosomes, nucleosome-free regions, binding sites for transcription factors, etc. A map can be physically displayed, e.g., on a computer monitor. Exemplary epigenetic maps are shown in Fig. 1C, 3A, 4A, 4B, 5B and 5C.

The term "mapping information," as used herein, refers to assembling experimentally-obtained information about an area to a physical map of the area.

PARSE EX.1007.024

The term "sequence read abundance," as used herein, refers to the number of times a particular sequence or nucleotide is observed in a collection of sequence reads.

The term "nucleosome-free fragments," as used herein, refers to fragments of genomic DNA that are relatively depleted or devoid of nucleosomes, i.e., between nucleosomes.

The term "chromatin accessibility," as used herein, refers to how accessible a nucleic acid site is within a polynucleotide, such as in genomic DNA, i.e., how "open" the chromatin is. A nucleic acid site associated with a polypeptide, such as with genomic DNA in nucleosomes, is usually inaccessible. A nucleic acid site not complexed with a polypeptide is generally accessible, such as with genomic DNA between nucleosomes (with the exception of nucleic acid sites complexed with transcription factors and other DNA binding proteins).

The term "DNA binding protein occupancy," as used herein, refers to whether a binding site for a sequence specific DNA binding protein (e.g., a binding site for a transcription factor) is occupied by the DNA binding protein. DNA binding protein occupancy can be measured quantitatively or qualitatively.

The term "global occupancy," as used herein, refers to whether a plurality of different binding sites for a DNA binding protein that are distributed throughout the genome (e.g., a binding sites for a transcription factor) are bound by the DNA binding protein . DNA binding protein occupancy can be measured quantitatively or qualitatively.

The term "diagnosis," as used herein, refers to a determination of whether a subject has a particular disease or condition.

The term "prognosis," as used herein, refers to prediction of a clinical outcome, e.g., disease recurrence, recovery from a disease, death, as well as a prediction of how a subject that has a particular disease or condition will respond to a particular treatment.

Other definitions of terms may appear throughout the specification.

DESCRIPTION OF EXEMPLARY EMBODIMENTS

In one aspect, a method for analyzing chromatin is provided. In certain embodiments, the method comprises: (a) treating chromatin isolated from a population of cells with an insertional enzyme complex to produce tagged fragments of genomic DNA. In this step, the chromatin is tagmented (i.e., cleaved and tagged in the same reaction) using an insertional enzyme such as Tn5 or MuA that cleaves the genomic DNA in open regions in the chromatin and adds adaptors to both ends of the fragments. Methods for tagmenting isolated genomic DNA are known in the art (see, e.g., Caruccio Methods Mol. Biol. 2011 733: 241-

PARSE EX.1007.025

55; Kaper et al, Proc. Natl. Acad. Sci. 2013 110: 5552-7; Marine et al, Appl. Environ. Microbiol. 2011 77: 8071-9 and US20100120098) and are commercially available from Illumina (San Diego, CA) and other vendors. Such systems may be readily adapted for use herein. In some cases, the conditions may be adjusted to obtain a desirable level of insertion in the chromatin (e.g., an insertion that occurs, on average, every 50 to 200 base pairs in open regions). The chromatin used in the method may be made by any suitable method. In some embodiments, nuclei may be isolated, lysed, and the chromatin may be further purified, e.g., from the nuclear envelope. In other embodiments, the chromatin may be isolated by contacting isolated nuclei with the reaction buffer. In these embodiments, the isolated nuclei may lyse when it makes contact with the reaction buffer (which comprises insertional enzyme complexes and other necessary reagents), which allows the insertional enzyme complexes access to the chromatin. In these embodiments, the method may comprise isolating nuclei from a population of cells; and combining the isolated nuclei with the transposase and adaptors, wherein the combining results in both lysis of the nuclei to release said chromatin and production of the adaptor-tagged fragments of genomic DNA. The chromatin does not require cross-linking as in other methods (e.g., ChIP-SEQ methods).

After the chromatin has been fragmented and tagged to produce tagged fragments of genomic DNA, at least some of the adaptor tagged fragments are sequenced to produce a plurality of sequence reads. The fragments may be sequenced using any convenient method. For example, the fragments may be sequenced using Illumina's reversible terminator method, Roche's pyrosequencing method (454), Life Technologies' sequencing by ligation (the SOLiD platform) or Life Technologies' Ion Torrent platform. Examples of such methods are described in the following references: Margulies et al (Nature 2005 437: 376–80); Ronaghi et al (Analytical Biochemistry 1996 242: 84–9); Shendure et al (Science 2005 309: 1728-32); Imelfort et al (Brief Bioinform. 2009 10:609-18); Fox et al (Methods Mol Biol. 2009;553:79-108); Appleby et al (Methods Mol Biol. 2009;513:19-39) and Morozova et al (Genomics. 2008 92:255-64), which are incorporated by reference for the general descriptions of the methods and the particular steps of the methods, including all starting products, methods for library preparation, reagents, and final products for each of the steps. As would be apparent, forward and reverse sequencing primer sites that are compatible with a selected next generation sequencing platform can be added to the ends of the fragments during the amplification step. In certain embodiments, the fragments may be amplified using PCR primers that hybridize to the tags that have been added to the fragments, where the primer used for PCR have 5' tails that are compatible with a particular sequencing platform.

PARSE EX.1007.026

In certain cases, the primers used may contain a molecular barcode (an "index") so that different pools can be pooled together before sequencing, and the sequence reads can be traced to a particular sample using the barcode sequence.

In another aspect, the present disclosure provides a method for determining accessibility of a polynucleotide at a site, wherein the polynucleotide is from a cell sample, said method comprising: inserting a plurality of molecular tags with an insertional enzyme into the polynucleotide and using the molecular tags to determine accessibility at the site. The cell sample can be from a primary source. The cell sample may consist of a single cell. The cell sample may consist of a finite number of cells (e.g. less than about 500,000 cells).

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 instances, at least one of the proteins is a transcription factor. Additionally, the method can comprise using the molecular tags to generate an accessibility map of the polynucleotide.

The polynucleotide may be fragmented into a plurality of fragments during the insertion of the molecular tags. In some cases, the fragments may be amplified. In some cases, the fragments can be sequenced to generate a plurality of sequencing reads. This may be used to determine the accessibility of the polynucleotide at any given site. The fragments may be sequenced using a high-throughput sequencing technique. In some cases, the sequencing reads can be normalized based on the sequence insertion preference of the insertional enzyme. The length of the sequenced reads can be used to determine a chromatin state annotation.

The polynucleotide can be bound to a plurality of association molecules. The association molecules can be, for example, proteins, nucleic acids or saccharides. In some cases, the association molecules can comprise histones. In other cases, the association molecules can comprise aptamers.

The insertional enzyme can be any enzyme capable of inserting a nucleic acid sequence into a polynucleotide. In some cases, the insertional enzyme can insert the nucleic acid sequence into the polynucleotide in a substantially sequence-independent manner. The insertional enzyme can be prokaryotic or eukaryotic. Examples of insertional enzymes include, but are not limited to, transposases, HERMES, and HIV integrase. The transposase can be a Tn transposase (e.g. Tn3, Tn5, Tn7, Tn10, Tn552, Tn903), a MuA transposase, a Vibhar transposase (e.g. from *Vibrio harveyi*), Ac-Ds, Ascot-1, Bs1, Cin4, Copia, En/Spm, F element, hobo, Hsmar1, Hsmar2, IN (HIV), IS1, IS2, IS3, IS4, IS5, IS6, IS10, IS21, IS30, IS50, IS51, IS150, IS256, IS407, IS427, IS630, IS903, IS911, IS982, IS1031, ISL2, L1,

PARSE EX.1007.027

Mariner, P element, Tam3, Tc1, Tc3, Tel, THE-1, Tn/O, TnA, Tn3, Tn5, Tn7, Tn10, Tn552, Tn903, Tol1, Tol2, TnlO, Tyl, any prokaryotic transposase, or any transposase related to and/or derived from those listed above. In certain instances, a transposase related to and/or derived from a parent transposase can comprise a peptide fragment with at least about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% amino acid sequence homology to a corresponding peptide fragment of the parent transposase. The peptide fragment can be at least about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 60, about 70, about 80, about 90, about 100, about 150, about 200, about 250, about 300, about 400, or about 500 amino acids in length. For example, a transposase derived from Tn5 can comprise a peptide fragment that is 50 amino acids in length and about 80% homologous to a corresponding fragment in a parent Tn5 transposase. In some cases, the insertion can be facilitated and/or triggered by addition of one or more cations. The cations can be divalent cations such as, for example, Ca^{2+}, Mg^{2+} and Mn^{2+}.

The molecular tags can comprise sequencing adaptors, locked nucleic acids (LNAs), zip nucleic acids (ZNAs), RNAs, affinity reactive molecules (e.g. biotin, dig), self-complementary molecules, phosphorothioate modifications, azide or alkyne groups. In some cases, the sequencing adaptors can further comprise a barcode label. Further, the barcode labels can comprises a unique sequence. The unique sequences can be used to identify the individual insertion events. Any of the tags can further comprise fluorescence tags (e.g. fluorescein, rhodamine, Cy3, Cy5, thiazole orange, etc.).

Additionally, the insertional enzyme can further comprise an affinity tag. In some cases, the affinity tag can be an antibody. The antibody can bind to, for example, a transcription factor, a modified nucleosome or a modified nucleic acid. Examples of modified nucleic acids include, but are not limited to, methylated or hydroxymethylated DNA. In other cases, the affinity tag can be a single-stranded nucleic acid (e.g. ssDNA, ssRNA). In some examples, the single-stranded nucleic acid can bind to a target nucleic acid. In further cases, the insertional enzyme can further comprise a nuclear localization signal.

In some cases, the cell sample can be permeabilized to allow access for the insertional enzyme. The permeabilization can be performed in a way to minimally perturb the nuclei in the cell sample. In some instances, the cell sample can be permeabilized using a permeabilization agent. Examples of permeabilization agents include, but are not limited to, NP40, digitonin, tween, streptolysin, and cationic lipids. In other instances, the cell sample

PARSE EX.1007.028

can be permeabilized using hypotonic shock and/or ultrasonication. In other cases, the insertional enzyme can be highly charged, which may allow it to permeabilize through cell membranes.

In yet another aspect, the present disclosure provides a method for analyzing the three-dimensional structure of a polynucleotide from a cell sample, comprising: inserting a plurality of molecular tags with an insertional enzyme into the polynucleotide; and using the molecular tags to analyze the three-dimensional structure of the polynucleotide. The insertional enzyme can comprise two or more enzymatic moieties, which may be optionally linked together. The enzymatic moieties can be linked by using any suitable chemical synthesis or bioconjugation methods. For example, the enzymatic moieties can be linked via an ester/amide bond, a thiol addition into a maleimide, Native Chemical Ligation (NCL) techniques, Click Chemistry (i.e. an alkyne-azide pair), or a biotin-streptavidin pair. In some cases, each of the enzymatic moieties can insert a common sequence into the polynucleotide. The common sequence can comprise a common barcode. The enzymatic moieties can comprise transposases or derivatives thereof. In some embodiments, the polynucleotide may be fragmented into a plurality of fragments during the insertion. The fragments comprising the common barcode can be determined to be in proximity in the three-dimensional structure of the polynucleotide.

The polynucleotide can be genomic DNA. The polynucleotide can be further bound to proteins, such as histones, and may be optionally packaged in the form of chromatin. In particular cases, DNA fragments corresponding to one or more regions of a genome (e.g., 2 or more, 10 or more, 50 or more, 100 or more, up to 1,000 or more regions) may be enriched, i.e., selected, by hybridization prior to sequencing. In these embodiments, the entire library does not need to be sequenced. Depending on the desired result and length of the selected region (if a selection step has been performed), this step of the method may result in at least 1,000 sequencing (e.g., at least 10,000, at least 100,000, at least 500,000, at least 10^6, at least $5x10^6$, up to 10^7 or more sequencing reads). The sequence reads are generally stored in computer memory.

Some embodiments of the methods involve making an epigenetic map of a region of the genome of the cells. This step may be done by mapping information obtained from the sequence reads to the region. In these embodiments, the sequence reads are analyzed computationally to produce a number of numerical outputs that are mapped to a representation (e.g., a graphical representation) of a region of interest. As will be explained in greater detail below, many types of information may be mapped, including, but not

PARSE EX.1007.029

limited to: (i) cleavage sites for the transposase; (ii) the sizes of the fragments produced in step a); (iii) fragment length; (iii) the positions of sequence reads of a defined range in length; and (iv) sequence read abundance.

For example, the sequence reads may be analyzed computationally to identify the ends of the fragments (from which the transposon cleavage sites can be inferred). In these embodiments, one end of a fragment can be defined by sequence that is at the beginning of a sequencing read and the other end of the fragment can be defined by sequence that is at the beginning of a second sequencing read, where the first and second sequencing reads were obtained by paired end sequencing (e.g., using Illumina's sequencing platform). The same information can be obtained from examining the beginning and end of longer sequence reads (which should, in theory, have the sequence of both adaptors; one at one end and the other at the other end). In these embodiments, a single sequence read may contain both adaptor sequences, in which case both ends of a fragment (which correspond to two cleavage sites for the two separate transposases) can be inferred from a single sequence read. The lengths of the fragments can be calculated by, e.g., mapping the fragment ends onto the nucleotide sequence of the region of interest, and counting the number of base pairs between those positions. The information used may be obtained using the nucleotide sequences at the beginning and/or the end of a sequence read.

In certain cases, the sequence reads can be placed into groups by length. In some embodiments, some sequences can be annotated as being a nucleosome-free sequence (i.e., a sequence from a fragment that is predicted to be between nucleosomes) based on its size. Reads that are associated with mononucleosomes, dinucleosomes and trinucleosomes can also be identified. These cutoffs can be determined using the data shown in **Fig. 12**. Fragment lengths (which provide the same information as sequence read lengths) can also be processed in the same way. In certain cases, sequence read abundance, i.e., the number of times a particular sequence in a genomic region is represented in the sequence reads, may be calculated.

The resultant epigenetic map can provide an analysis of the chromatin in the region of interest. For example, depending on which information is mapped, the map can show one or more of the following: a profile of chromatin accessibility along the region; DNA binding protein (e.g., transcription factor) occupancy for a site in the region; nucleosome-free DNA in the region; positioning of nucleosomes along the region; and a profile of chromatin states along the region. In some embodiments, the method may further comprise measuring global occupancy of a binding site for the DNA binding protein by, e.g., aggregating data for one

PARSE EX.1007.030

DNA binding protein over a plurality of sites to which that protein binds. In certain instances, the map can also be annotated with sequence information, and information about the sequence (e.g., the positions of promoters, introns, exons, known enhancers, transcriptional start sites, untranslated regions, terminators, etc.) so that the epigenetic information can be viewed in context with the annotation.

In certain embodiments, the epigenetic map can provide information regarding active regulatory regions and/or the transcription factors that are bound to the regulatory regions. For example, nucleosome positions can be inferred from the lengths of sequencing reads generated. Alternatively, transcription factor binding sites can be inferred from the size, distribution and/or position of the sequencing reads generated. In some cases, novel transcription factor binding sites can be inferred from sequencing reads generated. In other cases, novel transcription factors can be inferred from sequencing reads generated.

The population of cells used in the assay may be composed of any number of cells, e.g., about 500 to about 10^6 or more cells, about 500 to about 100,000 cells, about 500 to about 50,000 cells, about 500 to about 10,000 cells, about 50 to 1000 cells, about 1 to 500 cells, about 1 to 100 cells, about 1 to 50 cells, or a single cell. In some cases, the cell sample can consist of less than about 1000, about 2000, about 3000, about 4000, about 5000, about 6000, about 7000, about 8000, about 9000, about 10,000, about 15,000, about 20,000, about 25,000, about 30,000, about 40,000, about 50,000, about 60,000, about 70,000, about 80,000, about 90,000, about 100,000, about 120,000, about 140,000, about 160,000, about 180,000, about 200,000, about 250,000, about 300,000, about 350,000, about 400,000, about 450,000, about 500,000, about 600,000, about 700,000, about 800,000, about 900,000, or about 1,000,000 cells. In other cases, the cell sample can consist of more than about 1000, about 2000, about 3000, about 4000, about 5000, about 6000, about 7000, about 8000, about 9000, about 10,000, about 15,000, about 20,000, about 25,000, about 30,000, about 40,000, about 50,000, about 60,000, about 70,000, about 80,000, about 90,000, about 100,000, about 120,000, about 140,000, about 160,000, about 180,000, about 200,000, about 250,000, about 300,000, about 350,000, about 400,000, about 450,000, about 500,000, about 600,000, about 700,000, about 800,000, about 900,000, or about 1,000,000 cells.

The cells can be from any source. In certain cases, the cells may be obtained from a culture of cells, e.g., a cell line. In other cases, the cells may be isolated from an individual (e.g., a patient or the like). The cells may be isolated from a soft tissue or from a bodily fluid, or from a cell culture that is grown *in vitro*. In particular embodiments, the chromatin may be isolated from a soft tissue such as brain, adrenal gland, skin, lung, spleen, kidney,

PARSE EX.1007.031

liver, spleen, lymph node, bone marrow, bladder stomach, small intestine, large intestine or muscle, etc. Bodily fluids include blood, plasma, saliva, mucous, phlegm, cerebral spinal fluid, pleural fluid, tears, lactal duct fluid, lymph, sputum, cerebrospinal fluid, synovial fluid, urine, amniotic fluid, and semen, etc.

In some embodiments, the polynucleotide (e.g. genomic DNA, chromosomal DNA) used in the method may be from blood cells, wherein blood cells refers to a sample of whole blood or a sub-population of cells in whole blood. Sub-populations of cells in whole blood include platelets, red blood cells (erythrocytes), platelets and white blood cells (i.e., peripheral blood leukocytes, which are made up of neutrophils, lymphocytes, eosinophils, basophils and monocytes). These five types of white blood cells can be further divided into two groups, granulocytes (which are also known as polymorphonuclear leukocytes and include neutrophils, eosinophils and basophils) and mononuclear leukocytes (which include monocytes and lymphocytes). Lymphocytes can be further divided into T cells, B cells and NK cells. Peripheral blood cells are found in the circulating pool of blood and not sequestered within the lymphatic system, spleen, liver, or bone marrow. Other cells are present in blood that can be isolated. If blood is first contacted with an agent and then a sample of the blood is used in an assay, then a portion or all of the contacted blood may be used in the assay.

In certain embodiments, the cell sample can be isolated directly from a primary source. For example, the cell sample can be isolated directly from fresh tissues. In other cases, the cell sample can be isolated directly from frozen tissues. In yet other cases, the cell sample can be isolated directly from fixed tissues. Further examples of primary sources of cell samples include, but are not limited to, cells dissociated from tissues, blood cells, FFPE tissues, bacterial, viral, mitochondria, chloroplast, *in vitro* assembled protein DNA complexes, neutrophil extracellular traps.

Using the methods provided in the present disclosure, the disease state in a subject can be analyzed based on the accessibility of a polynucleotide site in a cell sample obtained from the subject. For example, transcription factor occupancy at any given site can result in the lack of accessibility at the site. Based on the transcription factor occupancy, the subject can then be treated with a suitable agent (e.g. a transcription factor inhibitor).

In certain cases, the cell samples can be further analyzed phenotypically. For example, the cell samples can be analyzed using fluorescence activated cell sorting (FACS) and/or laser capture microdissection (LCM). In some cases, the cell sample and/or polynucleotides may be divided into a plurality of portions. The portions can be divided

PARSE EX.1007.032

based on the molecular tags (e.g. fluorescence tags). In some cases, the cell sample and/or polynucleotides can be sorted. The sorting can be performed after the molecular tags are inserted into the polynucleotide. The sorting can be performed before the fragments are sequenced. The gene transcription of the cell samples can also be analyzed using techniques such as fluorescence *in situ* hybridization (FISH). The chromatin accessibility can be correlated with the phenotypical, transcriptional or translational analysis.

In some embodiments, the cells are of the same cell type. In these embodiments, the population of cells may be selected by MACS or FACS from a heterogeneous population of cells, e.g., blood, by known methods using labeled antibodies to cells surface markers. A wide variety of cells can be isolated using these methods, including stem cells, cancer stem cells and subsets of blood cells. In particular embodiments the following cells may be isolated from blood by FACS or MACS; T cells ($CD3^+$ $CD4^+$ $CD8^+$), B cells ($CD19^+$ $CD20^+$), dendritic cells ($CD11c^+$ $CD20^+$), NK Cell ($CD56^+$), stem cells/precursor cells ($CD34^+$; hematopoietic stem cells only), macrophage/monocytes ($CD14^+$ $CD33^+$), granulocytes ($CD66b^+$), platelet ($CD41^+$ $CD61^+$ $CD62^+$), erythrocytes ($CD235a^+$), endothelial cells ($CD146^+$) and epithelial cells ($CD326^+$). Subsets of these cells can be isolated using antibodies to further cell surface markers.

In some embodiments, the method can be used to compare two samples. In these embodiments, the method may comprise analyzing a first population of cells using the above-described method to produce a first epigenetic map; and analyzing a second population of cells using the above-described method to produce a second epigenetic map; and comparing the first epigenetic map to the second epigenetic map, e.g., to see if there are any changes in chromatin openness or transcription factor occupancy, for example.

In some embodiments, the first population of cells and the second population of cells are collected from the same individual at different times. In other embodiments, the first population of cells and the second population of cells are different populations of cells collected from tissues or different individuals.

Exemplary cell types that can be used in the method include, for example, cells isolated from a tissue biopsy (e.g., from a tissue having a disease such as colon, breast, prostate, lung, skin cancer, or infected with a pathogen etc.) and normal cells from the same tissue, e.g., from the same patient; cells grown in tissue culture that are immortal (e.g., cells with a proliferative mutation or an immortalizing transgene), infected with a pathogen, or treated (e.g., with environmental or chemical agents such as peptides, hormones, altered temperature, growth condition, physical stress, cellular transformation, etc.), and normal

PARSE EX.1007.033

cells (e.g., cells that are otherwise identical to the experimental cells except that they are not immortalized, infected, or treated, etc.); cells isolated from a mammal with a cancer, a disease, a geriatric mammal, or a mammal exposed to a condition, and cells from a mammal of the same species, e.g., from the same family, that is healthy or young; and differentiated cells and non-differentiated cells from the same mammal (e.g., one cell being the progenitor of the other in a mammal, for example). In one embodiment, cells of different types, e.g., neuronal and non-neuronal cells, or cells of different status (e.g., before and after a stimulus on the cells) may be compared. In another embodiment, the experimental material is cells susceptible to infection by a pathogen such as a virus, e.g., human immunodeficiency virus (HIV), etc., and the control material is cells resistant to infection by the pathogen. In another embodiment of the invention, the sample pair is represented by undifferentiated cells, e.g., stem cells, and differentiated cells. Cells from yeast, plants and animals, such as fish, birds, reptiles, amphibians and mammals may be used in the subject methods. In certain embodiments, mammalian cells, i.e., cells from mice, rabbits, primates, or humans, or cultured derivatives thereof, may be used.

In some exemplary embodiments, the method may be used to identify the effect of a test agent, e.g., a drug, or to determine if there are differences in the effect of two or more different test agents. In these embodiments, two or more identical populations of cells may be prepared and, depending on how the experiment is to be performed, one or more of the populations of cells may be incubated with the test agent for a defined period of time. After incubation with the test agent, the chromatin of the populations of cells can be analyzed using the methods set forth above, and the results can be compared. In a particular embodiment, the cells may be blood cells, and the cells can be incubated with the test agent ex vivo. These methods can be used to determine the mode of action of a test agent, to identify changes in chromatin structure or transcription factor occupancy in response to the drug, for example.

The method described above may also be used as a diagnostic (which term is intended to include methods that provide a diagnosis as well as methods that provide a prognosis). These methods may comprise, e.g., analyzing chromatin from a patient using the method described above to produce an epigenetic map; and providing a diagnosis or prognosis based on the epigenetic map.

The method set forth herein may be used to provide a reliable diagnostic to any condition associated with altered chromatin or DNA binding protein occupancy. The method can be applied to the characterization, classification, differentiation, grading, staging,

PARSE EX.1007.034

diagnosis, or prognosis of a condition characterized by an epigenetic pattern (e.g., a pattern of chromatin accessibility or DNA binding protein occupancy). For example, the method can be used to determine whether the epigenetic map of a sample from an individual suspected of being affected by a disease or condition is the same or different compared to a sample that is considered "normal" with respect to the disease or condition. In particular embodiments, the method can be directed to diagnosing an individual with a condition that is characterized by an epigenetic pattern at a particular locus in a test sample, where the pattern is correlated with the condition. The methods can also be used for predicting the susceptibility of an individual to a condition.

Exemplary conditions that are suitable for analysis using the methods set forth herein can be, for example, cell proliferative disorder or predisposition to cell proliferative disorder; metabolic malfunction or disorder; immune malfunction, damage or disorder; CNS malfunction, damage or disease; symptoms of aggression or behavioral disturbance; clinical, psychological and social consequences of brain damage; psychotic disturbance and personality disorder; dementia or associated syndrome; cardiovascular disease, malfunction and damage; malfunction, damage or disease of the gastrointestinal tract; malfunction, damage or disease of the respiratory system; lesion, inflammation, infection, immunity and/or convalescence; malfunction, damage or disease of the body as an abnormality in the development process; malfunction, damage or disease of the skin, the muscles, the connective tissue or the bones; endocrine and metabolic malfunction, damage or disease; headache or sexual malfunction, and combinations thereof.

In some embodiments, the method can provide a prognosis, e.g., to determine if a patient is at risk for recurrence. Cancer recurrence is a concern relating to a variety of types of cancer. The prognostic method can be used to identify surgically treated patients likely to experience cancer recurrence so that they can be offered additional therapeutic options, including preoperative or postoperative adjuncts such as chemotherapy, radiation, biological modifiers and other suitable therapies. The methods are especially effective for determining the risk of metastasis in patients who demonstrate no measurable metastasis at the time of examination or surgery.

The method can also be used to determining a proper course of treatment for a patient having a disease or condition, e.g., a patient that has cancer. A course of treatment refers to the therapeutic measures taken for a patient after diagnosis or after treatment. For example, a determination of the likelihood for recurrence, spread, or patient survival, can assist in determining whether a more conservative or more radical approach to therapy should be

PARSE EX.1007.035

taken, or whether treatment modalities should be combined. For example, when cancer recurrence is likely, it can be advantageous to precede or follow surgical treatment with chemotherapy, radiation, immunotherapy, biological modifier therapy, gene therapy, vaccines, and the like, or adjust the span of time during which the patient is treated.

In a particular embodiment, a lab will receive a sample (e.g., blood) from a remote location (e.g., a physician's office or hospital), the lab will analyze cells in the sample as described above to produce data, and the data may be forwarded to the remote location for analysis.

Compositions

In one aspect, the present disclosure provides compositions related to the methods provided herein. The composition can comprise a polynucleotide, an insertional enzyme and an insert element, wherein: the insert element can comprise a nucleic acid comprising a predetermined sequence and the insertional enzyme can further comprise an affinity tag. The polynucleotide can be further bound to a plurality of association molecules. The association molecules can be proteins (e.g. histones) or nucleic acids (e.g. aptamers). The affinity tag can be an antibody. In some cases, the antibody can be bound to a transcription factor. In other cases, the antibody can be bound to a modified nucleosome. In further cases, the antibody can be bound to a modified nucleic acid. Examples of modified nucleic acids include, but are not limited to, methylated or hydroxymethylated DNA. The affinity tag can also be a single-stranded nucleic acid (e.g. ssDNA, ssRNA). In some cases, the single-stranded nucleic acid can be bound to a target nucleic acid. In some instances, the insertional enzyme can further comprise a nuclear localization signal.

The composition can comprise a polynucleotide, an insertional enzyme and an insert element, wherein: the insertional enzyme comprises two or more enzymatic moieties and the enzymatic moieties are linked together. The insert element can be bound to the insertional enzyme. The insertional enzyme can also be bound to the polynucleotide. In some cases, the polynucleotide can be further bound to a plurality of association molecules. The association molecules can be proteins (e.g. histones) or nucleic acids (e.g. aptamers).

Kits

In yet another aspect, the present disclosure provides kits that contain reagents for practicing the subject methods, as described above. The subject kits can comprise: (a) reagents for isolating nuclei from a population of cells; (b) transposase and transposon tags,

PARSE EX.1007.036

and (c) transposase reaction buffer, wherein the components of the kit are configured such that, combining the reaction buffer, transposase and adaptors with nuclei *in vitro* results in both lysis of the nuclei to release chromatin and production of adaptor-tagged fragments of genomic DNA.

In some cases, the kit can comprise: (a) a cell lysis buffer; (b) an insertional enzyme comprising an affinity tag; and (c) an insert element comprising a nucleic acid, wherein said nucleic acid comprises a predetermined sequence. The insertional enzyme can be, for example, a transposase. The insertional enzyme can also comprise two or more enzymatic moieties that are linked together. In some cases, the affinity tag can be an antibody. The antibody can bind to a transcription factor, a modified nucleosome, or a modified nucleic acid. Examples of modified nucleic acids include, but are not limited to, methylated or hydroxymethylated DNA. In other cases, the affinity tag can be a single-stranded nucleic acid (e.g. ssDNA, ssRNA).

The kit may optionally contain other components, for example: PCR primers, PCR reagents such as polymerase, buffer, nucleotides etc., as described above. The various components of the kit may be present in separate containers or certain compatible components may be precombined into a single container, as desired.

In addition to above-mentioned components, the subject kits may further include instructions for using the components of the kit to practice the subject methods, i.e., instructions for sample analysis. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g., CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g., via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

PARSE EX.1007.037

Embodiments

A method of mapping chromatin is provided. In some embodiments, this method comprises the steps of: fragmenting the chromatin of rare or abundant cells with a transposase which inserts sequencing adapters into the polynucleotides within the chromatin, and amplifying and sequencing the fragments to generate a cell-specific map.

In certain embodiments, the cell-specific map provides information regarding active regulatory regions and the transcription factors that are bound to said regulatory regions.

In certain embodiments, the number of said rare cells is between 1 and 100,000.

In certain embodiments, the transposase is derived from Tn5 transposase.

In certain embodiments, the transposase is derived from MuA transposase.

In certain embodiments, nucleosome positions are inferred from the lengths of sequencing reads generated.

In certain embodiments, transcription factor binding sites are inferred from sequencing reads generated.

In certain embodiments, chromatin is isolated directly from fresh tissues.

In certain embodiments, chromatin is isolated directly from frozen tissues.

In certain embodiments, chromatin is isolated directly from fixed tissues.

In certain embodiments, sequences are added to the sequencing adapter to uniquely identify the fragments for multiplexing (barcoding).

In certain embodiments, an affinity tag is used to target the transposase to a specific macromolecule of interest.

In certain embodiments, sequences are added to the sequencing adapter to uniquely identify the fragments for multiplexing (barcoding), and an affinity tag is used to target the transposase to a specific macromolecule of interest.

In certain embodiments the affinity tag is an antibody targeted to a transcription factor.

In certain embodiments the affinity tag is an antibody targeted to a modified nucleosome.

In certain embodiments, the insert size distribution at a specific genomic locus is used to infer chromatin openness.

In certain embodiments, the insert size distribution and positions of insertion are used to infer transcription factor binding.

In certain embodiments, the number of sequencing reads obtained is normalized by measured sequence insertion preference of the transposase.

PARSE EX.1007.038

In certain embodiments, novel transcription factor binding sites are inferred from sequencing reads generated.

In certain embodiments, novel transcription factors are inferred from sequencing reads generated.

In certain embodiments, causal variants can be inferred by looking at allele specific generation of sequencing reads.

In certain embodiments, chromatin state annotations are inferred from the distribution of lengths of sequencing reads.

EXAMPLES

Aspects of the present teachings can be further understood in light of the following examples, which should not be construed as limiting the scope of the present teachings in any way.

Example 1. Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq)

Described herein is an Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq) - based on direct *in vitro* transposition of sequencing adapters into native chromatin – as a rapid and sensitive method for integrative epigenomic analysis. ATAC-seq captures open chromatin sites using a simple 2-step protocol from 500 to 50,000 cells, and reveals the interplay between genomic locations of open chromatin, DNA binding proteins, individual nucleosomes, and higher-order compaction at regulatory regions with nucleotide resolution. Classes of DNA binding factor that strictly avoid, can tolerate, or tend to overlap with nucleosomes have been discovered. Using ATAC-seq, the serial daily epigenomes of resting human T cells was measured and evaluated from a proband via standard blood draws, demonstrating the feasibility of reading personal epigenomes in clinical timescales for monitoring health and disease.

Materials and Methods

An exemplary implementation of ATAC-seq protocol has three major steps:

1) Prepare nuclei: To prepare nuclei, 50,000 cells were spun at 500 x g for 5 minutes, followed by a wash using 50 μL of cold 1x PBS and centrifugation at 500 x g for 5 minutes. Cells were lysed using cold lysis buffer (10 mM Tris-Cl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500 x g for 10 minutes using a refrigerated centrifuge. To avoid losing cells during the nuclei

PARSE EX.1007.039

prep, a fixed angle centrifuge was used and they were carefully pipetted away from the pellet after centrifugations.

2) Transpose and purify: Immediately following the nuclei prep, the pellet was resuspended in the transposase reaction mix (25 μL 2x TD buffer, 2.5 μL Transposase (Illumina) and 22.5 μL of nuclease free water). The transposition reaction was carried out for 30 minutes at 37 °C. Directly following transposition the sample was purified using a Qiagen Minelute kit.

3) PCR: Following purification, we amplified library fragments using 1x NEBnext PCR master mix and 1.25 μM of custom Nextera PCR primers 1 and 2 (see table below), using the following PCR conditions: 72°C for 5 minutes, 98°C for 30 seconds, followed by thermocycling at 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 1 minute. To reduce GC and size bias in PCR, the PCR reactions were monitored using qPCR in order to stop amplification prior to saturation. To do this, the full libraries were amplified for 5 cycles, after 5 cycles an aliquot of the PCR reaction was taken and added to 10 μl of the PCR cocktail with Sybr Green at a final concentration of 0.6x. We ran this reaction for 20 cycles, to determine the additional number of cycles needed for the remaining 45 μL reaction. The libraries were purified using a Qiagen PCR cleanup kit yielding a final library concentration of ~30 nM in 20 μL. Libraries were amplified for a total of 10-12 cycles.

Ad1_noMX:	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG (SEQ ID NO:1)
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:2)
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:3)
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:4)
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:5)
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:6)
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:7)
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:8)
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:9)
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:10)
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:11)
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:12)
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:13)

PARSE EX.1007.040

Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:14)
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:15)
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:16)
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:17)
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:18)
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:19)
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:20)
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:21)
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:22)
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:23)
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:24)
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT (SEQ ID NO:25)

Low cell number protocol: To prepare the 500 and 5,000 cell reactions the same protocol was used with some notable exceptions: The transposition reaction was done in a 5 μL instead of 50 μL reaction. Also, the Qiagen Minelute purification was eliminated prior to PCR and instead took the 5 μL reaction immediately after transposition directly into the 50 μL PCR.

Library QC and quantitation: During the ATAC-seq protocol, the size selection step was avoided to maximize the library complexity. The sequenced insert size is a distribution between 40 bp to 1 kb with a mean of ~120 bps. From bioanalyzer and gels we observed fragments >2 kb, which would make Qubit and other mass-based quantitation methods hard to interpret. For this reason we quantified our libraries using qPCR based methods.

$CD4^+$ enrichment from peripheral blood: One green-top tube of whole blood was obtained from 1 normal volunteer three times over a 72-hour period, under a Stanford University IRB-approved protocol. Informed consent was obtained. 5mL of blood at each timepoint was negatively selected for CD4+ cells, using RosetteSep Human $CD4^+$ T Cell Enrichment Cocktail (StemCell Technology). RosetteSep cocktail was incubated with the blood at 50μL/mL for 20 min, diluted in an equal volume of PBS with 2% FBS, and underlayed with 15 mL Ficol-Paque Plus (GE). Blood was centrifuged for 20 minutes at

PARSE EX.1007.041

1200 x g without break, negatively selected cells were removed from the density medium: plasma interface, and cells were washed X2 in PBS with 2% FBS.

FACS sorting peripheral blood leukocytes and GM cells: GM 12878 cells were stained with DAPI NucBlue Fixed Cell Stain (molecular probes) and live cells were sorted using a FACSAria (BD Biosciences) using a 100 μm nozzle. One peripheral blood sample (buffy coat) was stained with BD Bioscience antibodies CD14-A-488 (M5E2, 1:20), CD3-PE-Cy7 (SK7, 1:20), CD4-APC-Cy7 (RPA-T4, 1:20), and CD8 (RPA-T8, 1:20) for 20 minutes in the dark at RT. Cells were lysed using BDpharmLyse 1:10 dil in diH20 (BD) for 15 min, centrifuged for 5 minutes, washed with PBS 2% FBS X 2, and resuspended in PBS with 2% FBS. 50,000 $CD3^+CD8^+$, $CD3^+CD4^+$, and $CD14^+$ cell populations were sorted into PBS with 10% FBS.

Data Analysis

Primary data processing: Data was collected using either 34 x 8 x 34 reads from a MiSeq or 50 x 8 x 50 reads on a HiSeq. Reads were aligned to hg19 using BOWTIE (Langmead et al *Genome Biol.* 2009 **10,** R25) using the parameters –X2000 and –m1. These parameters ensured that fragments up to 2 kb were allowed to align (-X2000) and that only unique aligning reads were collected (-m1). For all data files duplicates were removed using Picard.

For peak calling and footprinting, the read start sites were adjusted to represent the center of the transposon binding event. Previous descriptions of the Tn5 transposase show that the transposon binds as a dimer and inserts two adapters separated by 9 bps (Adey, A. *et al. Genome Biol* 2010 **11:** R119). Therefore, all reads aligning to the + strand were offset by +4 bps, and all reads aligning to the – strand were offset -5 bps.

ATAC-seq peak calling: We used ZINBA to call all reported ATAC-seq peaks in this manuscript. ZINBA was run using a window size of 300 bp and an offset 75 bp. Alignability was used to model the zero-inflated component and the ATAC-seq read count for the background and enriched components. Enriched regions were identified as those with a posterior probability >0.8.

ATAC-seq insertion size enrichment analysis within chromatin annotations: First, the distribution of paired-end sequencing fragment sizes overlapping each chromatin state (see the ensemble.org website) were computed. The distributions were then normalized to the percent maximal within each state and enrichment was computed relative to the genome-wide set of fragment sizes.

PARSE EX.1007.042

Nucleosome positioning: To generate the nucleosome position data track, we chose to split reads into various bins. Reads below 100 bps were considered nucleosome free, reads between 180 and 247 bps were considered to be mononucleosomes, reads between 315 and 473 bps were considered to be dinucleosomes and reads between 558 and 615 were considered to be trinucleosomes (for determining cutoffs see Fig. 12). Dinucleosome reads were split into two reads and Trinucleosome reads were split into three reads. Reads were analyzed using Danpos and Dantools using the parameters -p 1, -a 1, -d 20, -clonalcut 0. The background used was nucleosome free reads (reads less than 100 bps), allowing an effective negative weighting of these reads. This analysis allows calling multiple overlapping nucleosomes. Although generating nucleosome tracks using simple insert size cutoffs may yield false positives due to other nucleosome sized features, i.e. enhanaceosomes, we observed that we faithfully recapitulated global features on nucleosome position genome-wide (**Fig 2c,d** main text).

ChIP-seq peak calling and clustering: ChIP-seq data was downloaded from the UCSC ENCODE repository. Peaks where called using GEM (Guo et al, *PLoS Comput. Biol.* 2012 **8:** e1002638), the parameters used where -k_min 6 -k_max 20. Inputs where used as a control for peak calling. Binding events were annotated by distance to the nearest dyad in bins of 10 bps. Factors were then hierarchically clustered using Euclidean distance and normalized by gene and centered by mean. (Eisen et al. *Proc. Natl. Acad. Sci.* 1998 **95:** 14863–14868).

Footprinting using CENTIPEDE: The genome-wide set of motifs were obtained from the ENCODE motif repository (at the website of broadinstitute.org). The input for CENTIPEDE included the PWM score, conservation (PhyloP) and ATAC-seq counts within +/-100bp of each genomic region matching a motif. ChIP-seq data was obtained from the UCSC ENCODE repository.

Comparison of transcription factor regulatory networks: Transcription factor regulatory networks were constructed by comparing the GENCODE v14 genes with the genome-wide set of posterior probabilities estimated by CENTIPEDE for the respective cell-types. The extent of a transcription factor regulating each gene was determined by taking the sum of the weighed posterior probabilities for a given transcription factor mapping to the same chromosome. For each mapped motif the posterior probability was weighted based on the distance to the transcription start site for each gene. Comparison of transcription factor regulatory networks was computed as the correlation of each transcription factor in a given

PARSE EX.1007.043

cell type with all transcription factors in the other cell type. The resulting correlation matrix was hierarchically clustered using the Pearson correlation coefficient and complete linkage.

Candidate IL2 enhancer analysis: ENCODE data on the UCSC genome browser was inspected to identify putative IL2 enhancers in one or more cell types that may be responsive to FDA approved immomodulatory drugs. We scanned the intergenic region upstream of IL2 in hg19 for (i) enhancer-associated histone marks (H3K4me1 and H3K27ac), (ii) binding by one or more TFs as confirmed by ChIP-seq, and (iii) the TF pathway can be targeted by a human therapeutic. This analysis identified IRF4 and STAT3 binding sites in addition to the known NFAT-responsive elements.

Results

ATAC-seq probes chromatin accessibility with transposons

Hyperactive Tn5 transposase (Goryshin, J Biol Chem. 1998 **273:** 7367–7374; Adey, A. *et al. Genome Biol* 2010 **11:** R119, loaded *in vitro* with adapters for high-throughput DNA sequencing, can simultaneously fragment and tag a genome with sequencing adapters (previously described as "tagmentation"). It was hypothesized that transposition by purified Tn5, a prokaryotic transposase, on small numbers of unfixed eukaryotic nuclei would interrogate regions of accessible chromatin. An Assay for Transposase Accessible Chromatin followed by high-throughput sequencing (ATAC-seq) is described. ATAC-seq uses Tn5 transposase to integrate its adapter payload into regions of accessible chromatin, whereas steric hindrance less accessible chromatin makes transposition less probable. Therefore, amplifiable DNA fragments suitable for high-throughput sequencing are preferentially generated at locations of open chromatin (**Fig 1a**). The entire assay and library construction can be carried out in a simple two-step process involving Tn5 insertion and PCR. In contrast, published DNase- and FAIRE-seq protocols for assaying chromatin accessibility involve multi-step protocols and many potentially loss-prone steps, such as adapter ligation, gel purification, and crosslink reversal. For instance, a published DNase-seq protocol calls for approximately 44 steps, and two overnight incubations, while published FAIRE-seq protocols require two overnight incubations carried out over at least 3 days. Furthermore, these protocols require 1-50 million cells (FAIRE) or 50 million cells (DNase-seq), perhaps because of these complex workflows (**Fig 1b**). In comparison to established methods, ATAC-seq enables rapid and efficient library generation because assay and library preparation are carried out in a single enzymatic step.

PARSE EX.1007.044

Extensive analyses show that ATAC-seq provides accurate and sensitive measure of chromatin accessibility genome-wide. ATAC-seq was carried out on 50,000 and 500 unfixed nuclei isolated from GM12878 lymphoblastoid cell line (ENCODE Tier 1) for comparison and validation with chromatin accessibility data sets, including DNase-seq and FAIRE-seq. At a locus previously highlighted by others, (**Fig. 1c**), ATAC-seq has a signal-to-noise ratio similar to DNase-seq, which was generated from approximately 3 to 5 orders-of-magnitude more cells. Peak intensities were highly reproducible between technical replicates (R=0.98), and highly correlated between ATAC-seq and DNase-seq (R=0.79 and R=0.83, **Fig. 6**), and it is noted that the majority of reads within peaks come from intersections of DNase and ATAC-seq peaks (**Fig. 7**). Comparing our data to DHSs identified in ENCODE DNase-seq data, receiver operating characteristic (ROC) curves demonstrate a similar sensitivity and specificity as DNase-seq (**Fig. 8**). It is also noted that ATAC-seq peak intensities correlate well with markers of active chromatin and not with transposase sequence preference (**Figs. 9** and **10**). Highly sensitive open chromatin detection is maintained even when using 5,000 or 500 human nuclei as starting material (**Figs. 8** and **11**), although, under the conditions used, sensitivity is diminished for smaller numbers of input material, as can be seen in Fig 1c.

ATAC-seq insert sizes disclose nucleosome positions

It was found that ATAC-seq paired-end reads produce detailed information about nucleosome packing and positioning. The insert size distribution of sequenced fragments from human chromatin has clear periodicity of approximately 200 base pairs, suggesting many fragments are protected by integer multiples of nucleosomes (**Fig 2a**). This fragment size distribution also shows clear periodicity equal to the helical pitch of DNA. By partitioning insert size distribution according to functional classes of chromatin as defined by previous models (Hoffman et al. Nucleic Acids Res. 2013 41: 827–841), and normalizing to the global insert distribution we observe clear class-specific enrichments across this insert size distribution (**Fig. 2b**), demonstrating that these functional states of chromatin have an accessibility "fingerprint" that can be read out with ATAC-seq. These differential fragmentation patterns are consistent with the putative functional state of these classes, as CTCF-bound regions are enriched for short fragments of DNA, while transcription start sites are differentially depleted for mono-, di- and tri-nucleosome associated fragments. Transcribed and promoter flanking regions are enriched for longer multi-nucleosomal fragments, suggesting they may represent more compacted forms of chromatin. Finally, prior studies have shown that certain DNA sequences are refractory to nuclease digestion and

PARSE EX.1007.045

released as large, multi-nucleosome-sized fragments; subsequent studies showed that such fragments are condensed heterochromatin. Indeed repressed regions were found to be depleted for short fragments and enriched for phased multi-nucleosomal inserts, consistent with their expected inaccessible state. These data suggest that ATAC-seq reveals differentially accessible forms of chromatin, which have been long hypothesized to exist *in vivo*.

To explore nucleosome positioning within accessible chromatin in the GM12878 cell line, data was partitioned into reads generated from putative nucleosome free regions of DNA, and reads likely derived from nucleosome associated DNA (see **Fig. 12**). Using a simple heuristic that positively weights nucleosome associated fragments and negatively weights nucleosome free fragments (see Methods), we calculated a data track used to call nucleosome positions within regions of accessible chromatin (Chen, K. et al. Genome Research 2013 23, 341–351). An example locus (**Fig. 3a**) contains a putative bidirectional promoter with CAGE data showing two transcription start sites (TSS) separated by ~700bps. ATAC-seq reveals in fact two distinct nucleosome free regions, separated by a single well-positioned mononucleosome (**Fig. 3a**). Compared to MNase-seq, ATAC-seq data is more amenable to detecting nucleosomes within putative regulatory regions, as the majority of reads are concentrated within accessible regions of chromatin (**Fig. 3b**). By averaging signal across all active TSSs, it is noted that nucleosome-free fragments are enriched at a canonical nucleosome-free promoter region overlapping the TSS, while the nucleosome signal is enriched both upstream and downstream of the active TSS, and displays characteristic phasing of upstream and downstream nucleosomes (**Fig. 3c**). Because ATAC-seq reads are concentrated at regions of open chromatin, a strong nucleosome signal is seen at the +1 nucleosome, which decreases at the +2, +3 and +4 nucleosomes, in contrast, MNase-seq nucleosome signal increases at larger distances from the TSS likely due to over digestion of more accessible nucleosomes. Additionally, MNase-seq (4 billion reads) assays all nucleosomes, whereas reads generated from ATAC-seq (198 million paired reads) are concentrated at regulatory nucleosomes (**Fig. 3b,c**). Using the nucleosome calls, putative distal regulatory regions and TSSs were further partititioned into regions that were nucleosome free and regions that were predicted to be nucleosome bound. It is noted that TSSs were enriched for nucleosome free regions when compared to distal elements, which tend to remain nucleosome rich (**Fig. 3d**). These data suggest ATAC-seq can provide high-resolution readout of nucleosome associated and nucleosome free regions in regulatory elements genome wide.

PARSE EX.1007.046

ATAC-seq reveals patterns of nucleosome-TF spacing

ATAC-seq high-resolution regulatory nucleosome maps can be used to understand the relationship between nucleosomes and DNA binding factors. Using ChIP-seq data, we plotted the position of a variety of DNA binding factors with respect to the dyad of the nearest nucleosome. Unsupervised hierarchical clustering (**Figure 3e**) revealed major classes of binding with respect to the proximal nucleosome, including 1) a strongly nucleosome avoiding group of factors with binding events stereotyped at ~180 bases from the nearest nucleosome dyad (comprising C-FOS, NFYA and IRF3), 2) a class of factors that "nestle up" precisely to the expected end of nucleosome DNA contacts, which notably includes chromatin looping factors CTCF and cohesion complex subunits RAD21 and SMC3; 3) a large class of primarily TFs that have gradations of nucleosome avoiding or nucleosome-overlapping binding behavior, and 4) a class whose binding sites tend to overlap nucleosome-associated DNA. Interestingly, this final class includes chromatin remodeling factors such as CHD1 and SIN3A as well as RNA polymerase II, which appears to be enriched at the nucleosome boundary. The interplay between precise nucleosome positioning and locations of DNA binding factor immediately suggests specific hypotheses for mechanistic studies, a potential advantage of ATAC-seq.

ATAC-seq footprints infer factor occupancy genome-wide

ATAC-seq enables accurate inference of DNA binding factor occupancy genome-wide. DNA sequences directly occupied by DNA-binding proteins should be protected from transposition; the resulting sequence "footprint" reveals the presence of the DNA-binding protein at each site, analogous to DNase digestion footprints. At a specific CTCF binding site on chromosome 1, we observed a clear footprint (a deep notch of ATAC-seq signal), similar to footprints seen by DNase-seq, at the precise location of the CTCF motif that coincides with the summit of the CTCF ChIP-seq signal in GM12878 cells (**Fig 4a**). The ATAC-seq signal was averaged over all expected locations of CTCF within the genome and observed a well-stereotyped "footprint" (**Fig. 4b**). Similar results were obtained for a variety of common TFs (for examples see **Fig. 13**). We inferred the CTCF binding probability from motif consensus score, evolutionary conservation, and ATAC-seq insertion data to generate a posterior probability of CTCF binding at all loci (**Fig. 4c**) (Pique-Regi et al. Genome Research 2011 21 447–455). Results using ATAC-seq closely recapitulate ChIP-seq binding data in this cell line and compare favorably to DNase-based factor occupancy inference (see **Fig. 14**), suggesting that factor occupancy data can be extracted from these ATAC-seq data

PARSE EX.1007.047

allowing reconstruction of regulatory networks.

ATAC-seq enables epigenomic analysis on clinical timescales

ATAC-seq is rapid, information rich, and compatible with small numbers of cells and may serve as a powerful tool for personalized epigenomics in the clinic. Specifically, one can envision "personal epigenomics" as genome-scale information about chromatin generated from an individual from a standard clinical sample in a clinical timescale. ATAC-seq was applised to assay the personal T-cell epigenome of a healthy volunteer via standard serial blood draws, to demonstrate a workflow capable of generating ATAC-seq libraries in clinical timescales. Using rapid T-cell enrichment and sample handling protocols, the total required time from blood draw to sequencing was approximately 275 minutes (**Fig. 5a**). When coupled with ongoing improvements to sequencing and analysis turn-around times, ATAC-seq can offer the possibility of a daily turn-around time for a personal epigenomic map. To explore this possibility, ATAC-seq was performed on three consecutive days via standard blood draws from a single individual (**Fig. 5b**). As an exercise to consider how personal epigenomic maps may contain personalized regulatory information, we investigated ATAC-seq profile at the *IL2* locus. IL-2 is a key cytokine that drives T-cell growth and functions in inflammatory and autoimmune diseases. Furthermore, distinct drugs inhibit the activities of different transcription factors that bind putative *IL2* enhancers in a context-dependent manner. In principle, one might wish to identify the causal transcription factor pathway in order to rationally target inhibition without exposing the patient to drugs unlikely to serve the therapeutic goal of IL-2 blockade. ATAC-seq shows that in the proband's T-cells, only NFAT, but not two other drug targets, is engaging *IL2* (**Fig. 5c**), providing clinically relevant information on the regulatory state of this individual.

Using ATAC-seq footprints the occupancy profiles of 89 transcription factors in proband T-cells were generated, enabling systematic reconstruction of regulatory networks. With this personalized regulatory map, we compared the genomic distribution of the same 89 transcription factors between GM12878 and proband CD4+ T-cells. Transcription factors that exhibit large variation in distribution between T-cells and B-cells are enriched for T-cell specific factors (**Fig. 5d**). This analysis shows NFAT is differentially regulating, while canonical CTCF occupancy is highly correlated within these two cell types (**Fig. 5d**). Supporting this interpretation, it is noted that specific loci where NFAT is localized nearby to known T-cell specific genes such as CD28 and a novel lincRNA RP11-229C3.2 (**Fig. 15**). Additionally, ATAC-seq of $CD4^+$ and $CD8^+$ T-cells, and monocytes isolated by

PARSE EX.1007.048

fluorescence-activated cell sorting (FACS) from a single blood draw created an interpretative framework for the personal epigenomes, and demonstrated that ATAC-seq is compatible with cellular enrichment using surface markers (**Fig. 16**). Separately, allele-specific chromatin accessibility has been shown to be particularly relevant to our understanding of human disease. As a proof of principle we also used ATAC-seq to identify candidate allele-specific open chromatin regions within the GM12878 cell line (**Fig. 17**). These results demonstrate the feasibility of generating detailed personalized gene regulatory networks from clinical samples, opening the door for future diagnostic applications.

Epigenomic studies of chromatin accessibility have yielded tremendous biological insights, but are currently limited in application by their complex workflows and large cell number requirements. While, improvements of existing methods may enable them to reach the similar performance, ATAC-seq in certain cases can offer substantial advantages over existing technologies due to its speed, simplicity, and low input cell number requirement. ATAC-seq is an information rich assay, allowing simultaneous interrogation of factor occupancy, nucleosome positions in regulatory sites, and chromatin accessibility genome-wide. These insights are derived from both the position of insertion and the distribution of insert lengths captured during the transposition reaction. While extant methods such as DNase- and MNase-seq can provide some subsets of the information in ATAC-seq, they each require separate assays with large cell numbers, which increases the time, cost, and limits applicability to many systems. ATAC-seq also provides insert size "fingerprints" of biologically relevant genomic regions, suggesting that it capture information on chromatin compaction. ATAC-seq may have broad applicability, significantly add to the genomics toolkit, and improve our understanding of gene regulation, particularly when integrated with other powerful rare cell techniques, such as FACS, laser capture microdissection (LCM) and recent advancements in RNA-seq.

ATAC-seq may be used to generate "personal epigenomic" profiles on a timescale compatible with clinical decision-making. Optimized procedures can transform a clinical blood sample to completed sequencing library in 275 minutes. The reduced input requirements and rapid workflows, when coupled with the recent introduction of rapid-turnaround high-throughput sequencing instruments, such as the MiSeq and HiSeq2500, should enable investigation of personalized epigenetic landscapes of selected tissues both in the lab and the clinic. ATAC-seq is compatible with FACS, enabling studies on carefully sorted and rare subpopulations from primary tissues. Cellular subpopulations selected at different points in development and aging, and human diseases, including cancer,

PARSE EX.1007.049

autoimmunity, and neuropsychiatric disorders are viable applications.

Example 2. Single-Cell ATAC-seq

Single-cell chromatin accessibility datasets were obtained using the ATAC-seq protocol. To ensure that the ratio of transposase molecules to open chromatin sites was kept nearly constant, the single-cell ATAC-seq assay was carried out by manipulating individual cells after an initial transposition reaction.

Transposases can serve as an open-chromatin stain

It was observed that after *in vitro* insertion of sequencing adapter, the Tn5 transposase remained tightly bound to the DNA and formed a high-affinity macromolecular complex that prevented dissociation of the generated ATAC-seq DNA fragments. To support this observation, Tn5 transposase was loaded with fluorescently labeled DNA adapters, and allowed for visualization of regions of open chromatin within the nucleus of individual cells (**Fig. 18**). Additional electrophoretic mobility shift assays also indicated that the transposase remained associated to DNA after transposition.

Single-Cell ATAC-seq provides unique reads characteristic of chromosomal DNA

Since this fluorescence signal localized to the nucleus and was detectable even after transposition, the single-cell ATAC-seq experiment was performed by keeping the transposed fragments in the nucleus during subsequent sorting and cell selection steps. A group of cells were permeabilized, and the chromosomal DNA was transposed with Tn5 transposase. The cells were kept under conditions that prevented the resulting ATAC-seq fragments from leaving the cell nucleus, (i.e. divalent cation was not chelated), and the individual cells were sorted into independent PCR reactions for library preparation, as described above. This workflow significantly simplified the workflow for single-cell analysis and provided two additional advantages. First, this abrogated any effect of the sorting process on the chromatin state because transposition preceded sorting. Second, it provided more robust ATAC-seq signal, as cells were sorted directly into a PCR master mix and amplified. Using this workflow, ~2,000-5,000 unique ATAC-seq reads per cell were generated. These reads were enriched for known open chromatin sites in GM12878 cells (**Fig. 19**) and displayed characteristic periodic enrichments indicative of nucleosomes (**Fig. 20**).

PARSE EX.1007.050

Example 3. Quality control

Assay for **T**ransposase **A**ccessible **C**hromatin (ATAC-seq) has been shown to be compatible with many methods for cell collection and has also worked effectively across many cell types and species. However, the following protocol has been optimized for human lymphoblastoid cells. Minor variations (i.e. cell number, centrifugation speeds, and lysis conditions) may optimized for particular applications.

I. Cell Preparation

1. Harvest cells (no fixation), protocol to be defined by the user.
2. Spin down 50,000 cells at 500 ×g for 5 min, 4°C.
3. Wash once with 50 µL of cold 1x PBS buffer. Spin down at 500 ×g for 5 min, 4°C.
4. Gently pipette to resuspend the cell pellet in 50 µL of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630). Spin down immediately at 500 ×g for 10 min, 4°C.
5. Discard the supernatant, and immediately continue to transposition reaction.

II. Transposition Reaction and Purification

1. Make sure the cell pellet is set on ice.
2. To make the transposition reaction mix, combine the following:

   25 µL 2x TD Buffer (Illumina Cat #FC-121-1030)
   2.5 µL Tn5 Transposes (Illumina Cat #FC-121-1030)
   22.5 µL Nuclease Free H_2O
   50 µl Total

3. Gently pipette to resuspend nuclei in the transposition reaction mix.
4. Incubate the transposition reaction at 37°C for 30 min.
5. Immediately following transposition, purify using a Qiagen MinElute Kit.
6. Elute transposed DNA in 10 µL Elution Buffer (10mM Tris buffer, pH 8).
7. Purified DNA can be stored at -20°C.

III. PCR Amplification

1. To amplify transposed DNA fragments, combine the following in a PCR tube:

   10 µL Transposed DNA
   9.7 µL Nuclease Free H_2O
   2.5 µL 25µM Customized Nextera PCR Primer 1*

PARSE EX.1007.051

2.5 μL 25μM Customized Nextera PCR Primer 2* [Barcode]
0.3 μL 100x SYBR Green I** (Invitrogen Cat #S-7563)
25 μL NEBNext High-Fidelity 2x PCR Master Mix (New England Labs Cat #M0541)
50 μL Total

* Complete list of primers are shown above.

**10,000x SYBR Green I is diluted in 10mM Tris buffer, pH 8 to make a 100x working solution.

2. Cycle as follows:

(1) 72°C, 5 min
(2) 98°C, 30 sec
(3) 98°C, 10 sec
(4) 63°C, 30 sec
(5) 72°C, 1 min
(6) Repeat steps 3-5, 4x
(7) Hold at 4°C

3. In order to reduce GC and size bias in PCR, the PCR reaction is monitored using qPCR to stop amplification prior to saturation. To run a qPCR side reaction, combine the following:

5 μL 5 cycles PCR amplified DNA
4.44 μL Nuclease Free H_2O
0.25 μL 25μM Customized Nextera PCR Primer 1*
0.25 μL 25μM Customized Nextera PCR Primer 2*
0.06 μL 100x SYBR Green I
5 μL NEBNext High-Fidelity 2x PCR Master Mix
15 μL Total

* Complete list of primers available in Section VI of this protocol

4. qPCR cycle as follows:

(1) 98°C, 30 sec
(2) 98°C, 10 sec
(3) 63°C, 30 sec
(4) 72°C, 1 min
(5) Repeat steps 2-4, 19x
(6) Hold at 4°C

PARSE EX.1007.052

5. The additional number of cycles needed for the remaining 45 μL PCR reaction is determined as following:

(1) Plot linear Rn vs. Cycle
(2) Set 5000 RF threshold
(3) Calculate the # of cycle that is corresponded to ¼ of maximum fluorescent intensity

If the # of cycle to be added lies in between two cycles, the # is determined by taking the smaller integer as the # of cycle to be added (i.e., blue and pink samples)
If two samples have similar Ct values but differs in the fluorescent intensities, calculate the # of cycle using the sample with lower fluorescent intensity (i.e., red and blue samples)

6. Run the remaining 45 μL PCR reaction to the correct # of cycle. Cycle as follows:

(1) 98°C, 30 sec (2) 98°C, 10 sec (3) 63°C, 30 sec (4) 72°C, 1 min
(5) Repeat steps 2-4, x times
(6) Hold at 4°C

7. Purify amplified library using Qiagen PCR Cleanup Kit. Elute the purified library in 20 μL Elution Buffer (10mM Tris Buffer, pH 8). Be sure to dry the column before adding elution buffer.

IV. Library QC using Gel Electrophoresis

1. Dilute 1:20 100bp NEB DNA ladder with 10mM Tris Buffer, pH 8.
2. Add 0.6 μL 10x SYBR Green I to every 5 μL of diluted ladder.
3. Mix 1:1 of the diluted ladder with 2x DNA loading dye.
4. Mix 1:1 of amplified library with 2x DNA loading dye.
5. Run amplified library on 5% Bio-Rad Mini-Protean TBE Precast Gel (stored at 4°C). Load 5 μL diluted ladder/DNA loading dye mixture. Load 10 μL amplified library/DNA loading dye mixture.
6. Run at ~100 mV for 45 min.

PARSE EX.1007.053

7. SYBR Green I dye has an excitation maximum at ~488 nm and has an emission maximum at ~520 nm. DNA stained with SYBR Green I dye can be visualized using a blue-light source or imaging systems equipped with laser that emits at 488 nm. We typically use Typhoon TRIO Variable Mode Imager from Amersham Biosciences for visualization. Images are best obtained by digitizing at 100 microns pixel size resolution with a 520 nm band-pass emission filter to screen out reflected and scattered excitation light and background fluorescence.

V. Library Quantitation

We use qPCR based methods to quantify our ATAC-seq libraries. We have found that other methods, such as Bioanalyzer and Qubit, can give misleading and inaccurate results due to the large distribution of insert sizes. We recommend quantifying libraries using the KAPA Library Quant Kit for Illumina Sequencing Platforms (KAPABiosystems).

Although the foregoing embodiments have been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the above teachings that certain changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

PARSE EX.1007.054

CLAIMS

What is claimed is:

1. A method for analyzing chromatin, comprising:

(a) treating chromatin isolated from a population of cells with an insertional enzyme complex to produce tagged fragments of genomic DNA;

(b) sequencing at least some of the tagged fragments to produce a plurality of sequence reads; and

(c) making an epigenetic map of a region of the genome of said cells by mapping information obtained from the sequence reads to the region.

2. The method of claim 1, wherein the information is obtained using the nucleotide sequences at the beginning and, optionally, the end of a sequence read.

3. The method of any of claims 1-2, wherein 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.

4. The method of claim 3, wherein the fragments of a defined size range are nucleosome-free fragments.

5. The method of any of claims 1-4, wherein 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;

(v) chromatin states.

6. The method of claim 5, further comprising measuring global occupancy of a binding site for said DNA binding protein.

PARSE EX.1007.055

7. The method of claim 5, wherein the DNA binding protein is a transcription factor.

8. The method of any of claims 1-7, wherein the population of cells is composed of 500 to 100,000 cells.

9. The method of any of claims 1-8, wherein the cells are isolated from an individual.

10. The method of any of claims 1-9, wherein the cells are isolated from blood of said individual.

11. The method of claims 1-10, wherein the cells are of the same cell type.

12. The method of claim 11, wherein the cells are FACS–selected cells.

13. The method of any of claims 1-12, wherein the treating step (a) comprises:

isolating nuclei from the a population of cells; and

combining the isolated nuclei with said insertional enzyme complex, wherein the combining results in both lysis of the nuclei to release said chromatin and production of said tagged fragments of genomic DNA.

14. The method of any of claims 1-13, wherein the transposase is derived from Tn5 transposase.

15. The method of any of claims 1-14, wherein the transposase is derived from MuA transposase.

16. A method for comparing two samples comprising:

(a) analyzing a first population of cells using the method of claim 1 to produce a first epigenetic map; and

(b) analyzing a second population of cells using the method of claim 1 to produce a second epigenetic map; and

(c) comparing the first epigenetic map to the second epigenetic map.

PARSE EX.1007.056

17. The method of claim 16, wherein the first population of cells and the second population of cells are collected from the same individual at different times.

18. The method of claim 16, wherein the first population of cells and the second population of cells are different populations of cells collected from different individuals.

19. A diagnostic method, comprising:

analyzing chromatin from a patient using the method of claim 1, to produce an epigenetic map; and

providing a diagnosis or prognosis based on the epigenetic map.

20. A kit comprising:

(a) reagents for isolating nuclei from a population of cells;

(b) an insertional enzyme complex, and

(c) transposase reaction buffer,

wherein the components of the kit are 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.

21. A method for determining accessibility of a polynucleotide at a site, wherein said polynucleotide is from a cell sample, comprising:

(a) inserting a plurality of molecular tags with an insertional enzyme into said polynucleotide; and

(b) using said molecular tags to determine accessibility at said site.

22. The method of claim 21, further comprising using said determined accessibility to identify one or more proteins that are bound to said polynucleotide at said site.

23. The method of claim 22, wherein at least one of said proteins is a transcription factor.

24. The method of claim 21, further comprising using said molecular tags to generate an accessibility map of said polynucleotide.

25. A method for analyzing the three-dimensional structure of a polynucleotide from a cell sample, comprising:

PARSE EX.1007.057

(a) inserting a plurality of molecular tags with an insertional enzyme into said polynucleotide; and

(b) using said molecular tags to analyze the three-dimensional structure of said polynucleotide.

26. The method of claim 21 or 25, wherein said cell sample is obtained from a primary source.

27. The method of claim 21 or 25, wherein said cell sample consists of less than about 500,000 cells.

28. The method of claim 27, wherein said cell sample is a single cell.

29. The method of claim 21 or 25, wherein said polynucleotide is fragmented into a plurality of fragments during step (a).

30. The method of claim 29, further comprising amplifying said fragments.

31. The method of claim 29, wherein said accessibility is determined or said three-dimensional structure is analyzed by sequencing said fragments and thereby generating a plurality of sequencing reads.

32. The method of claim 31, wherein said fragments are sequenced by a high-throughput sequencing technique.

33. The method of claim 31, further comprising normalizing said sequencing reads based on the sequence insertion preference of said insertional enzyme.

34. The method of claim 31, wherein the length of said sequenced reads is used to determine a chromatin state annotation.

35. The method of claim 21 or 25, further comprising permeabilizing said cell sample to allow access for said insertional enzyme.

PARSE EX.1007.058

36. The method of claim 35, wherein the nuclei in said cell sample is minimally perturbed during said permeabilization.

37. The method of claim 35, wherein said cell sample is permeabilized using a permeabilization agent.

38. The method of claim 37, wherein said permeabilization agent is selected from the group consisting of NP40, digitonin, tween, streptolysin, and cationic lipids.

39. The method of claim 35, wherein said cell sample is permeabilized using hypotonic shock and/or ultrasonication.

40. The method of claim 21 or 25, wherein said insertional enzyme further comprises a nuclear localization signal.

41. The method of claim 21 or 25, wherein said insertion is facilitated by addition of one or more divalent cations.

42. The method of claim 41, wherein said one or more divalent cations comprises magnesium.

43. The method of claim 41, wherein said one or more divalent cations comprises manganese.

44. The method of claim 21 or 25, further comprising analyzing a disease state in a subject based on said accessibility of said specific site or said three-dimensional structure of said polynucleotide, wherein said cell sample is obtained from said subject.

45. The method of claim 21 or 25, further comprising dividing said cell sample or said polynucleotides into a plurality of portions.

46. The method of claim 45, wherein said portions are divided based on said molecular tags.

PARSE EX.1007.059

47. The method of claim 21 or 25, further comprising analyzing a phenotype of said cell sample.

48. The method of claim 47, wherein said phenotype is correlated to said accessibility of said site or said three-dimensional structure of said polynucleotide.

49. The method of claim 21 or 25, wherein said insertional enzyme comprises two or more enzymatic moieties.

50. The method of claim 49, wherein each of said enzymatic moieties inserts a common sequence into said polynucleotide.

51. The method of claim 50, wherein said common sequence comprises a common barcode.

52. The method of claim 49, wherein said enzymatic moieties are linked together.

53. The method of claim 49, wherein said enzymatic moieties comprise transposases.

54. The method of claim 21 or 25, wherein said polynucleotide is fragmented into a plurality of fragments during step (a), and wherein said fragments comprising said common barcode are determined to be in proximity in the three-dimensional structure of said polynucleotide.

55. The method of claim 21 or 25, wherein said molecular tags comprise sequencing adaptors.

56. The method of claim 55, wherein said sequencing adaptors further comprise a barcode label.

57. The method of claim 55, wherein said barcode labels comprises a unique sequence.

58. The method of claim 21 or 25, wherein said molecular tags comprise fluorescence tags.

PARSE EX.1007.060

59. A composition comprising a polynucleotide, an insertional enzyme and an insert element, wherein:

(a) said insert element comprises a nucleic acid comprising a predetermined sequence; and

(b) said insertional enzyme further comprises an affinity tag.

60. A composition comprising a polynucleotide, an insertional enzyme and an insert element, wherein:

(a) said insertional enzyme comprises two or more enzymatic moieties; and

(b) said enzymatic moieties are linked together.

61. A kit comprising:

(a) a cell lysis buffer;

(b) an insertional enzyme comprising an affinity tag; and

(c) an insert element comprising a nucleic acid, wherein said nucleic acid comprises a predetermined sequence.

62. A kit comprising:

(a) a cell lysis buffer;

(b) an insertional enzyme comprising two or more enzymatic moieties, wherein said enzymatic moieties are linked together; and

(c) an insert element.

63. The method, composition or kit of claim 21, 25, 59, 60, 61 or 62, wherein said polynucleotide is bound to a plurality of association molecules.

64. The method, composition or kit of claim 63, wherein said association molecules are proteins.

65. The method, composition or kit of claim 64, wherein said proteins comprise histones.

66. The method, composition or kit of claim 21, 25, 59, 60, 61 or 62, wherein said insertional enzyme is a transposase.

PARSE EX.1007.061

67. The method, composition or kit of claim 66, wherein said transposase is derived from a Tn5 transposase.

68. The method, composition or kit of claim 66, wherein said transposase is derived from a MuA transposase.

69. The method, composition or kit of claim 66, wherein said transposase is derived from a Vibhar transposase.

70. The method, composition or kit of claim 21, 25, 59, 60, 61 or 62, wherein said insertional enzyme further comprises an affinity tag.

71. The method, composition or kit of claim 70, wherein said affinity tag is an antibody.

72. The method, composition or kit of claim 71, wherein said antibody binds to a transcription factor.

73. The method, composition or kit of claim 71, wherein said antibody binds to a modified nucleosome.

74. The method, composition or kit of claim 71, wherein said antibody binds to a modified nucleic acid.

75. The method, composition or kit of claim 74, wherein said modified nucleic acid is a methylated or hydroxymethylated DNA.

76. The method, composition or kit of claim 70, wherein said affinity tag is a single-stranded nucleic acid.

77. The method, composition or kit of claim 76, wherein said single-stranded nucleic acid binds to a target nucleic acid.

PARSE EX.1007.062

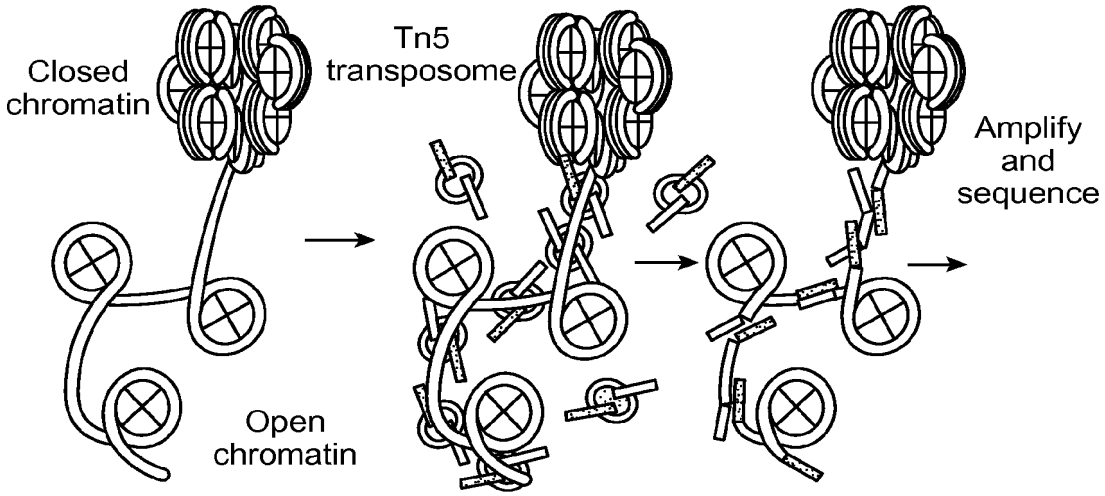


FIG. 1A

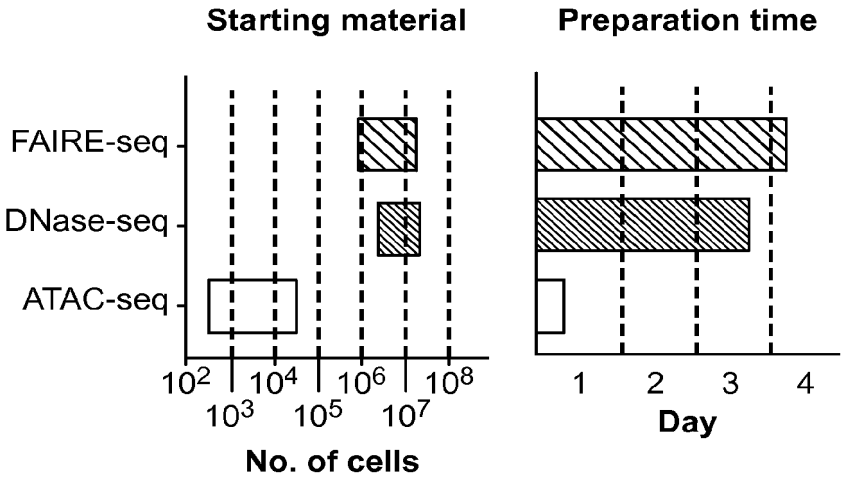


FIG. 1B

PARSE EX.1007.063

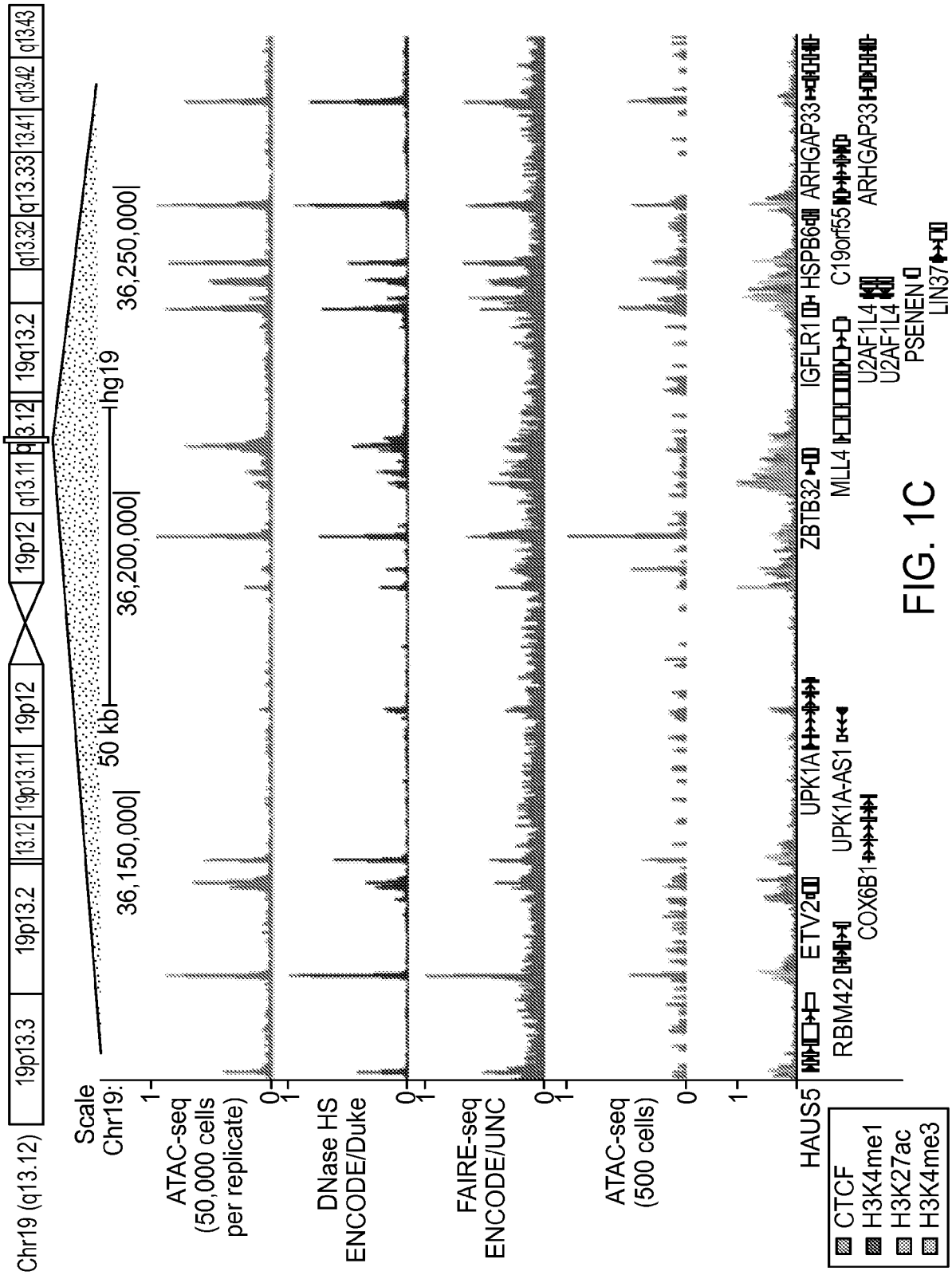


FIG. 1C

PARSE EX.1007.064

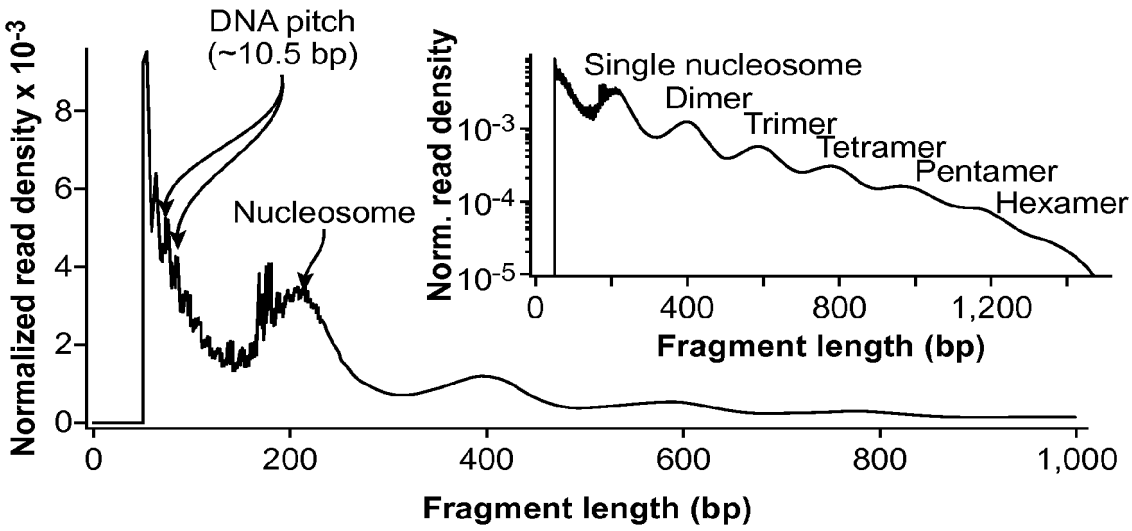


FIG. 2A

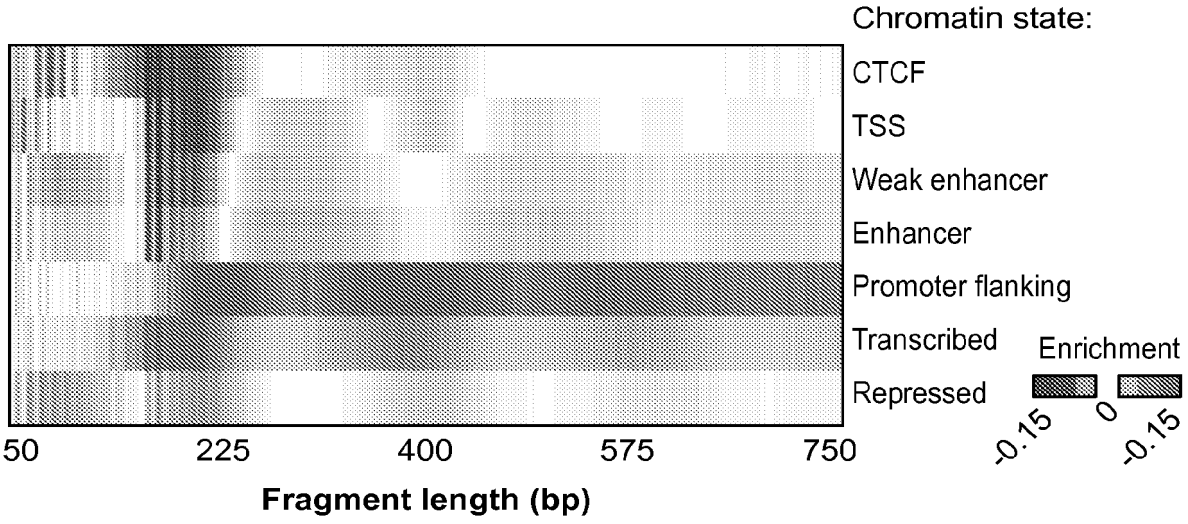


FIG. 2B

PARSE EX.1007.065

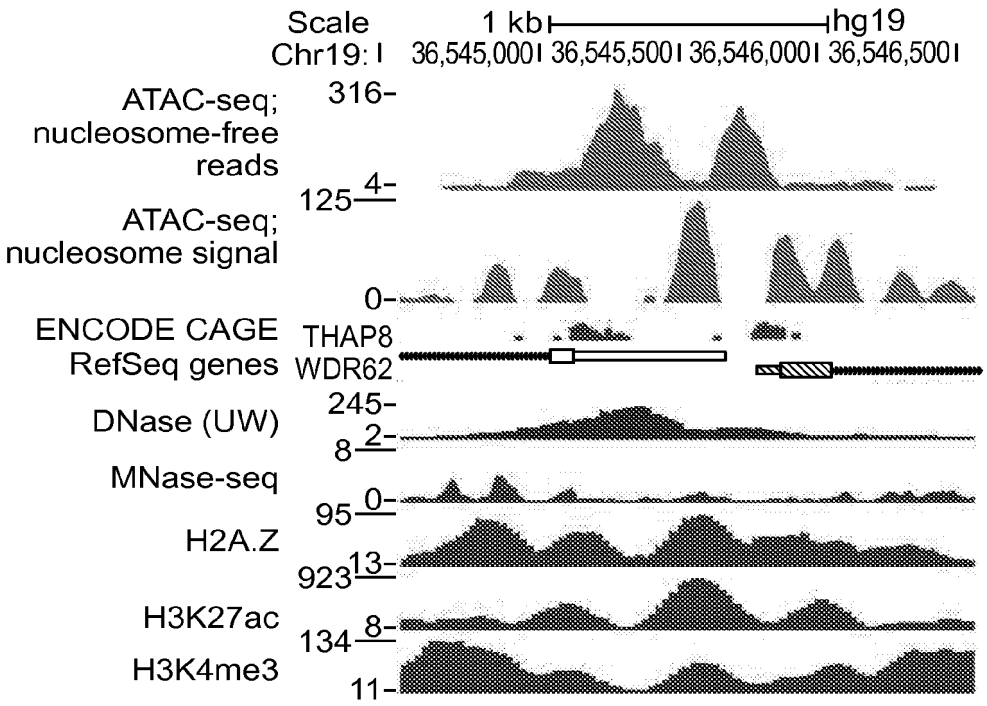


FIG. 3A

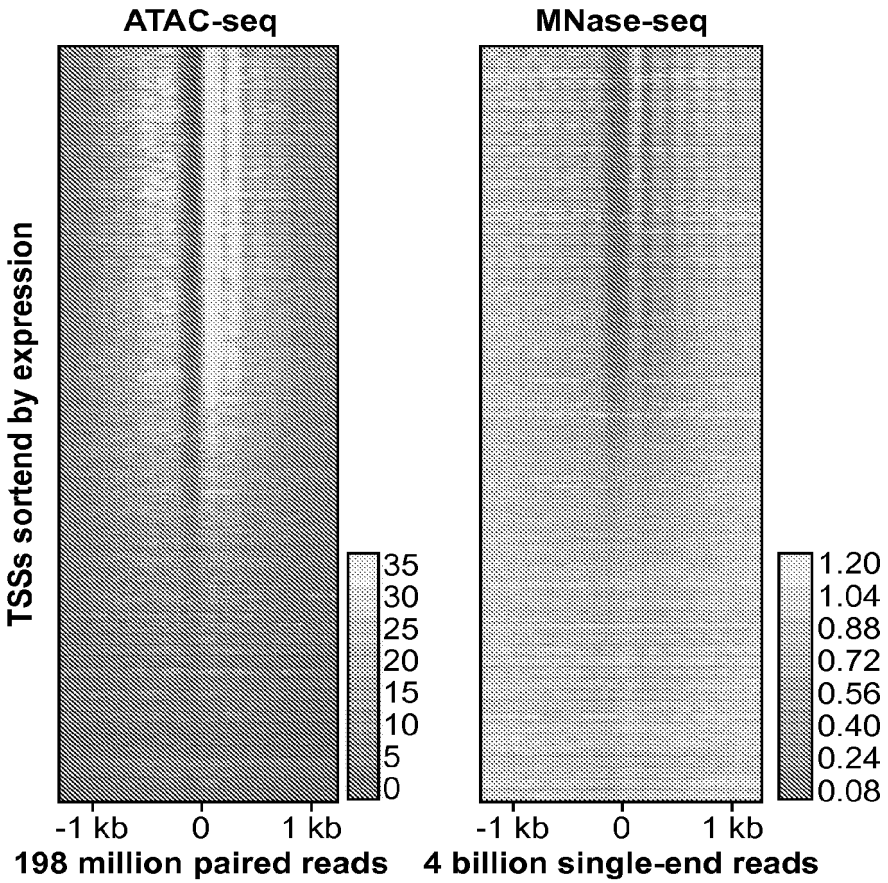


FIG. 3B

PARSE EX.1007.066

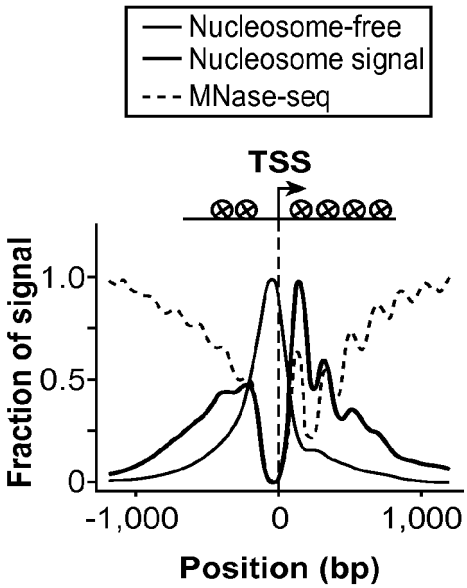


FIG. 3C

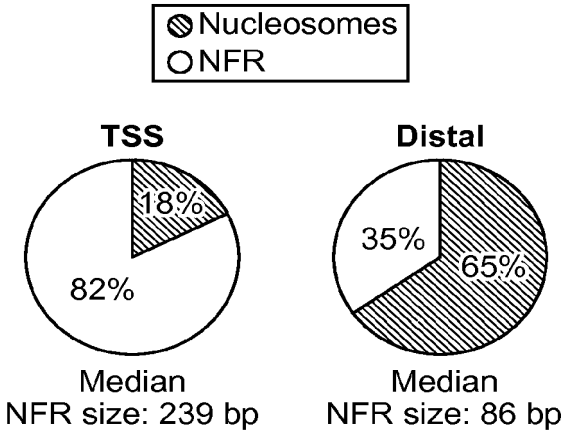


FIG. 3D

PARSE EX.1007.067

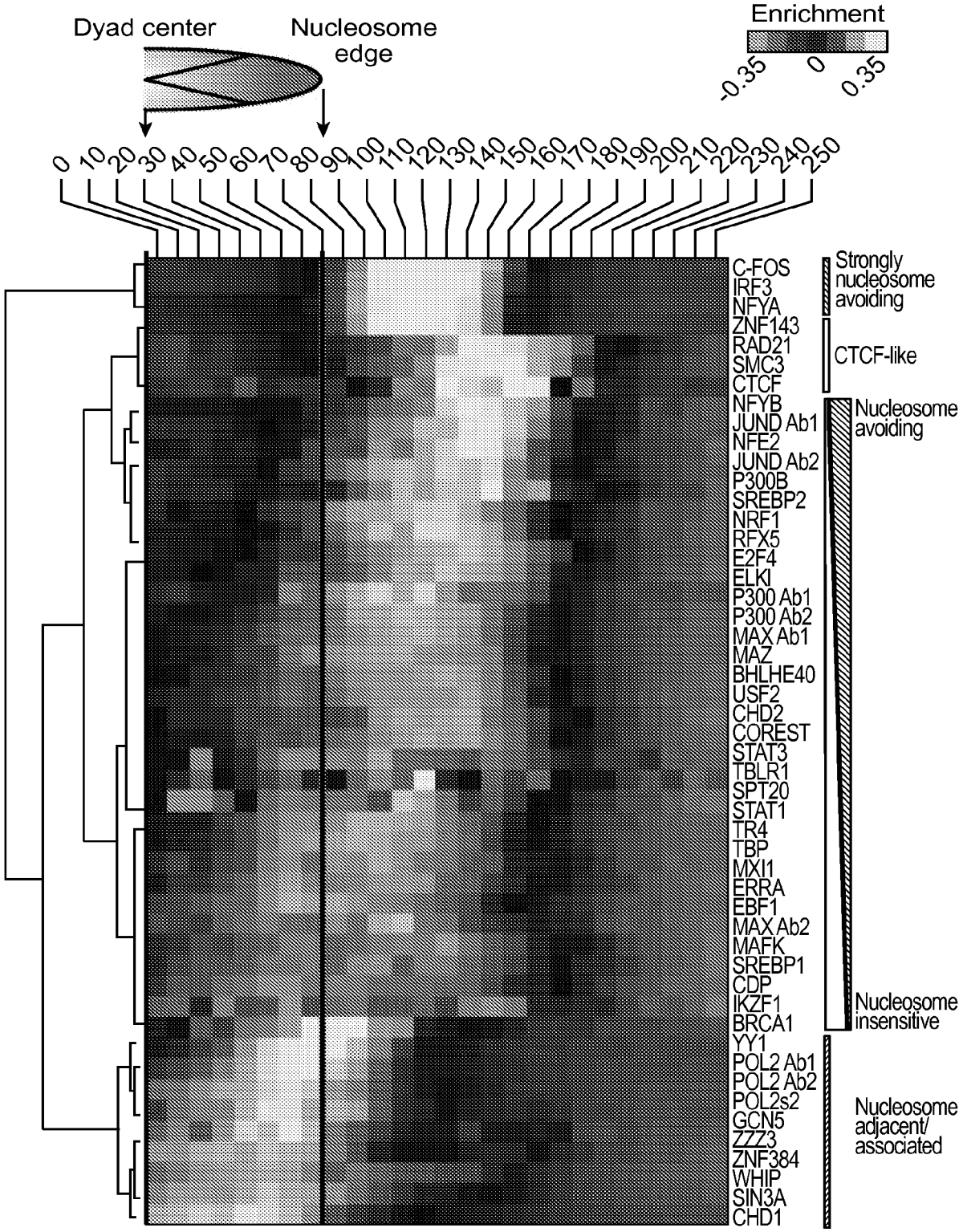


FIG. 3E

PARSE EX.1007.068

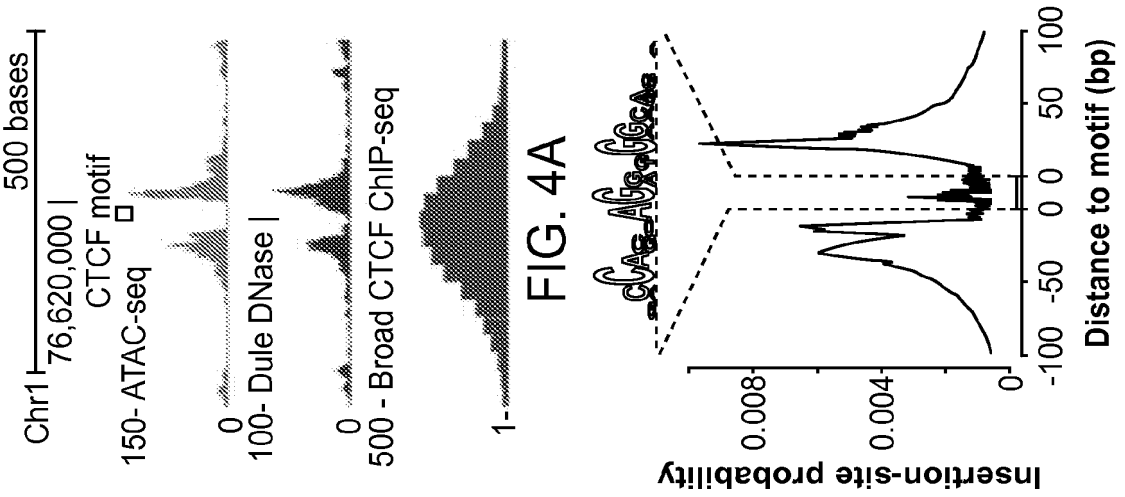


FIG. 4A

FIG. 4B

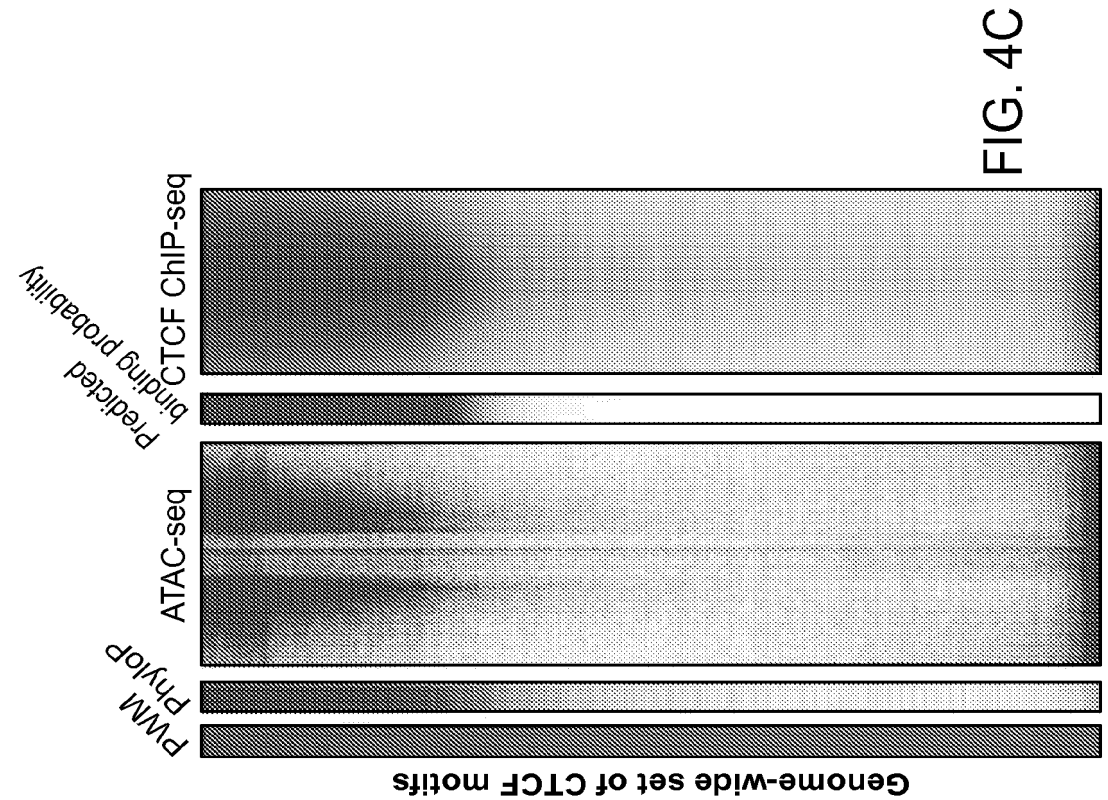


FIG. 4C

PARSE EX.1007.069

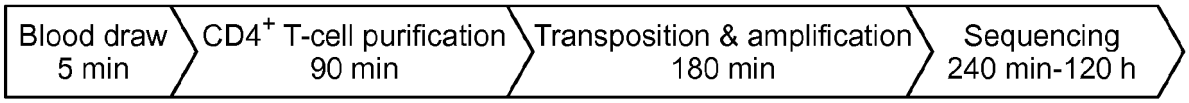


FIG. 5A

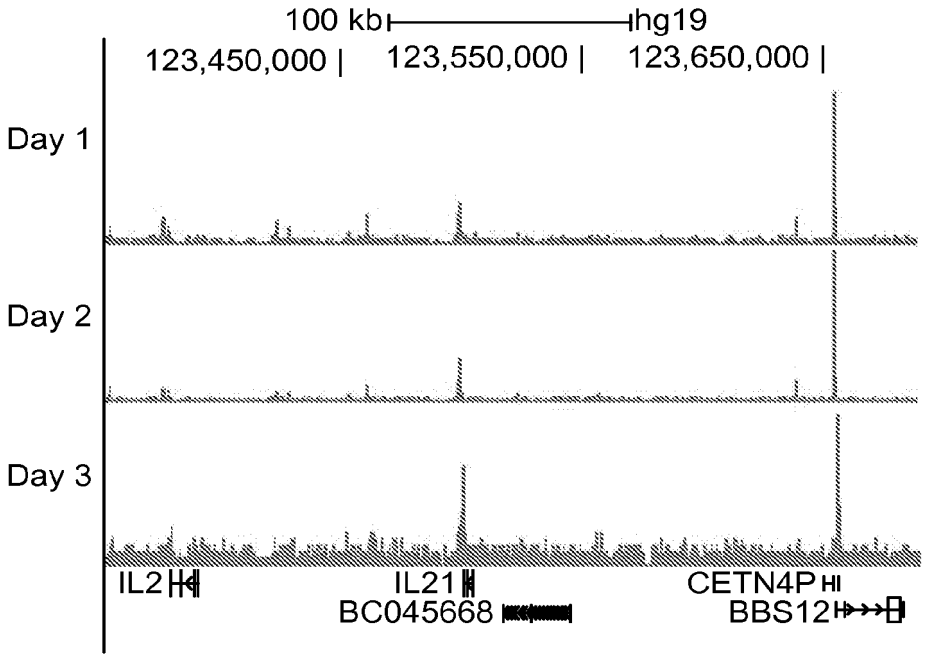


FIG. 5B

PARSE EX.1007.070

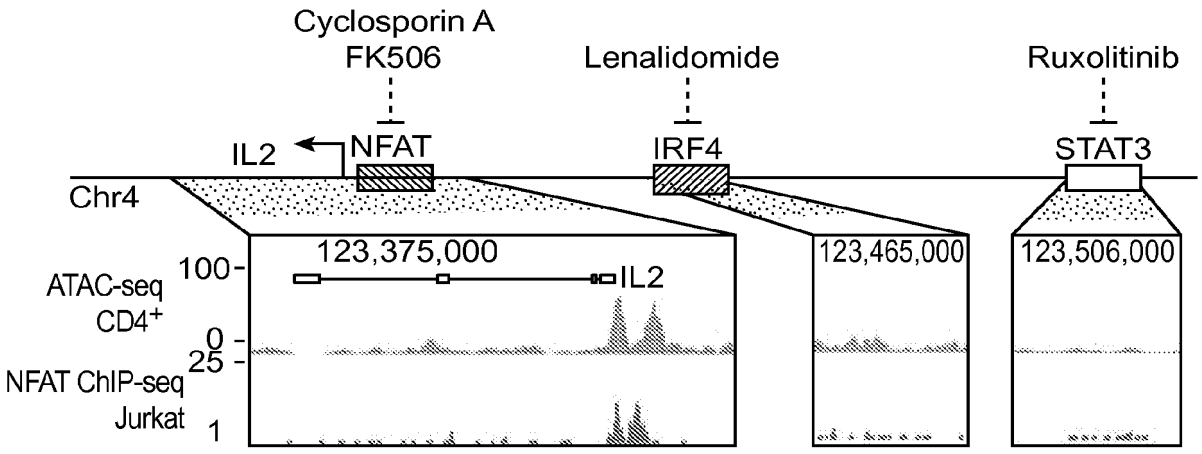


FIG. 5C

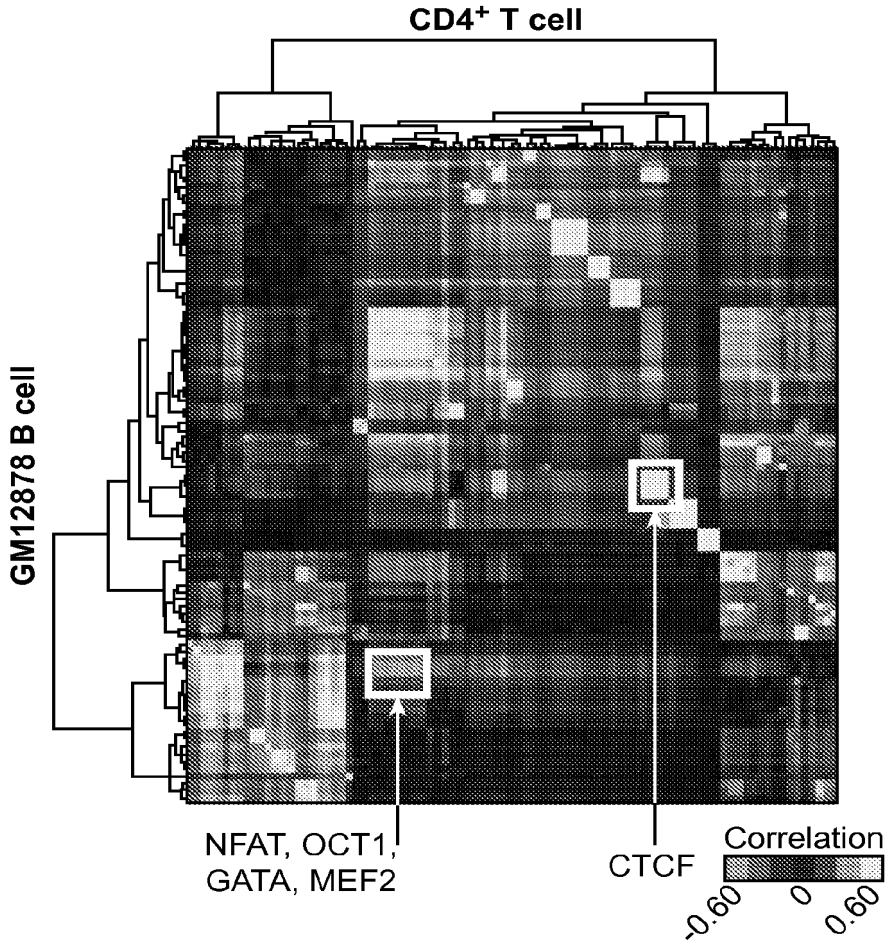


FIG. 5D

PARSE EX.1007.071

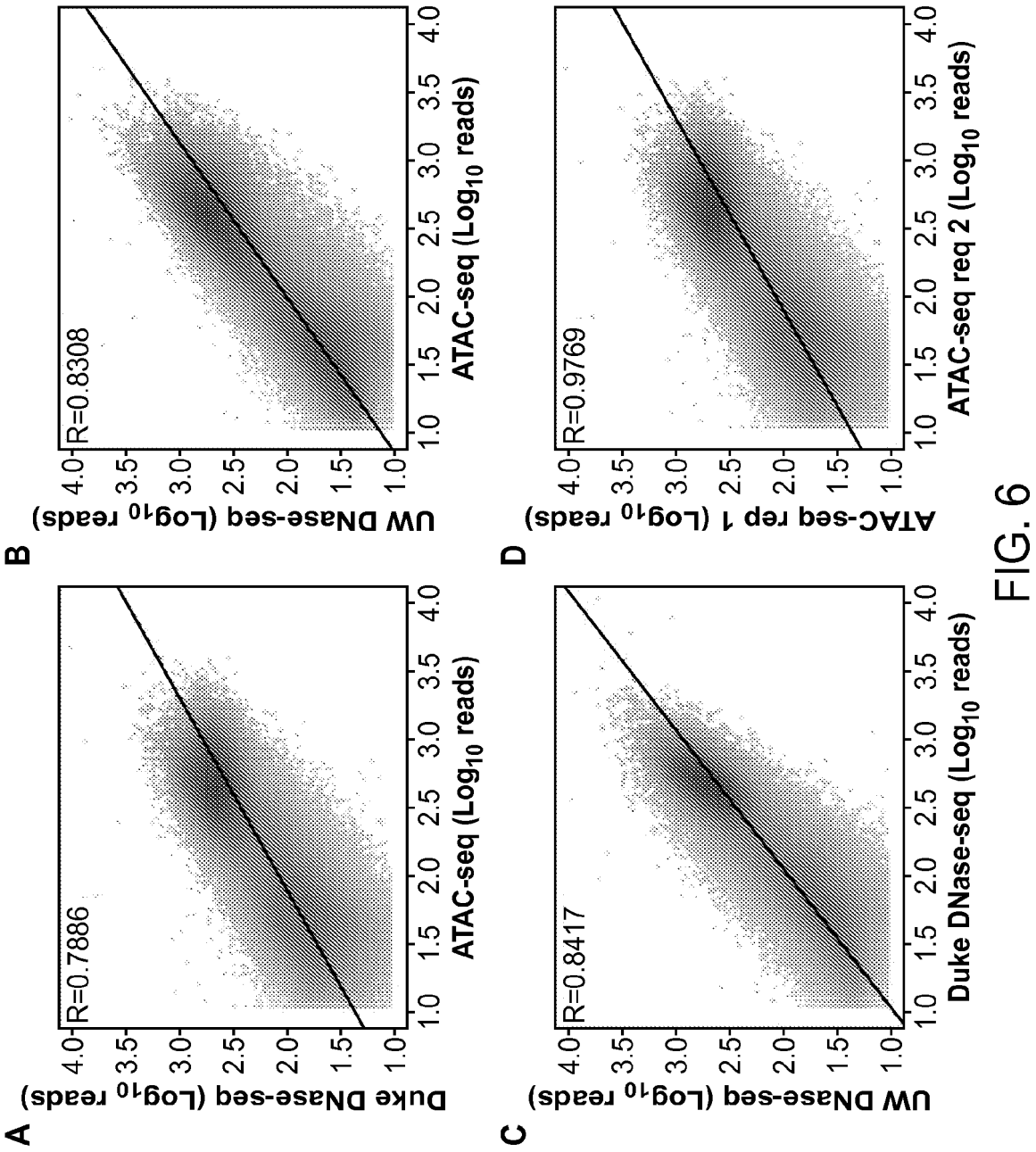
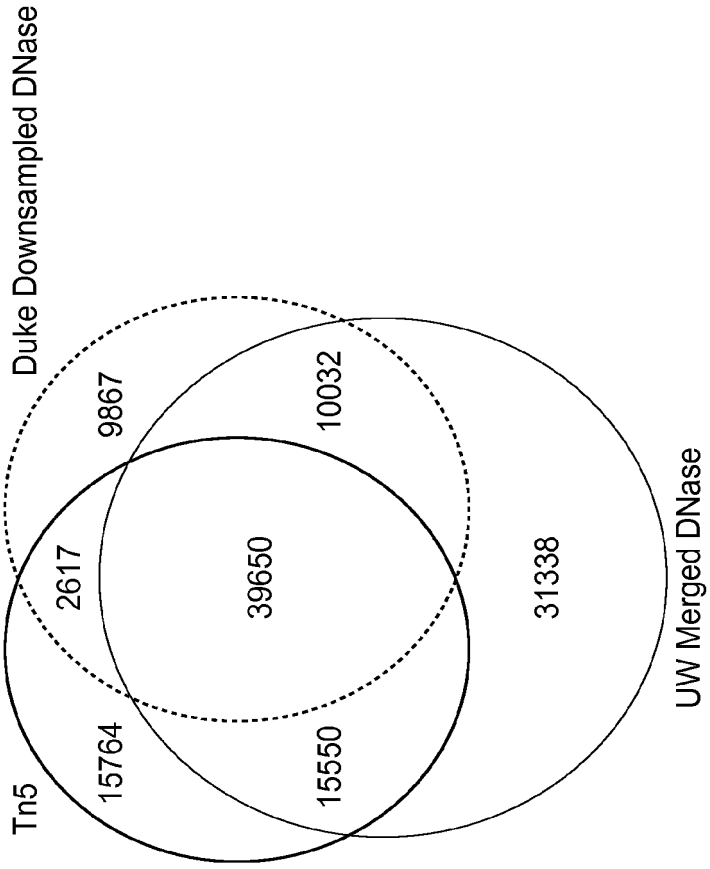


FIG. 6

PARSE EX.1007.072



	ATAC Unique	UW Unique	Duke Unique	ATAC and UW	ATAC and Duke	UW and Duke	Intersect
ATAC-Seq	6.30%	4.36%	1.08%	10.83%	1.53%	2.09%	73.80%
UW DNase	1.32%	8.13%	0.68%	7.57%	0.33%	4.34%	77.62%
Duke DNase	2.10%	5.27%	14.80%	4.02%	1.44%	6.91%	65.46%

FIG. 7

PARSE EX.1007.073

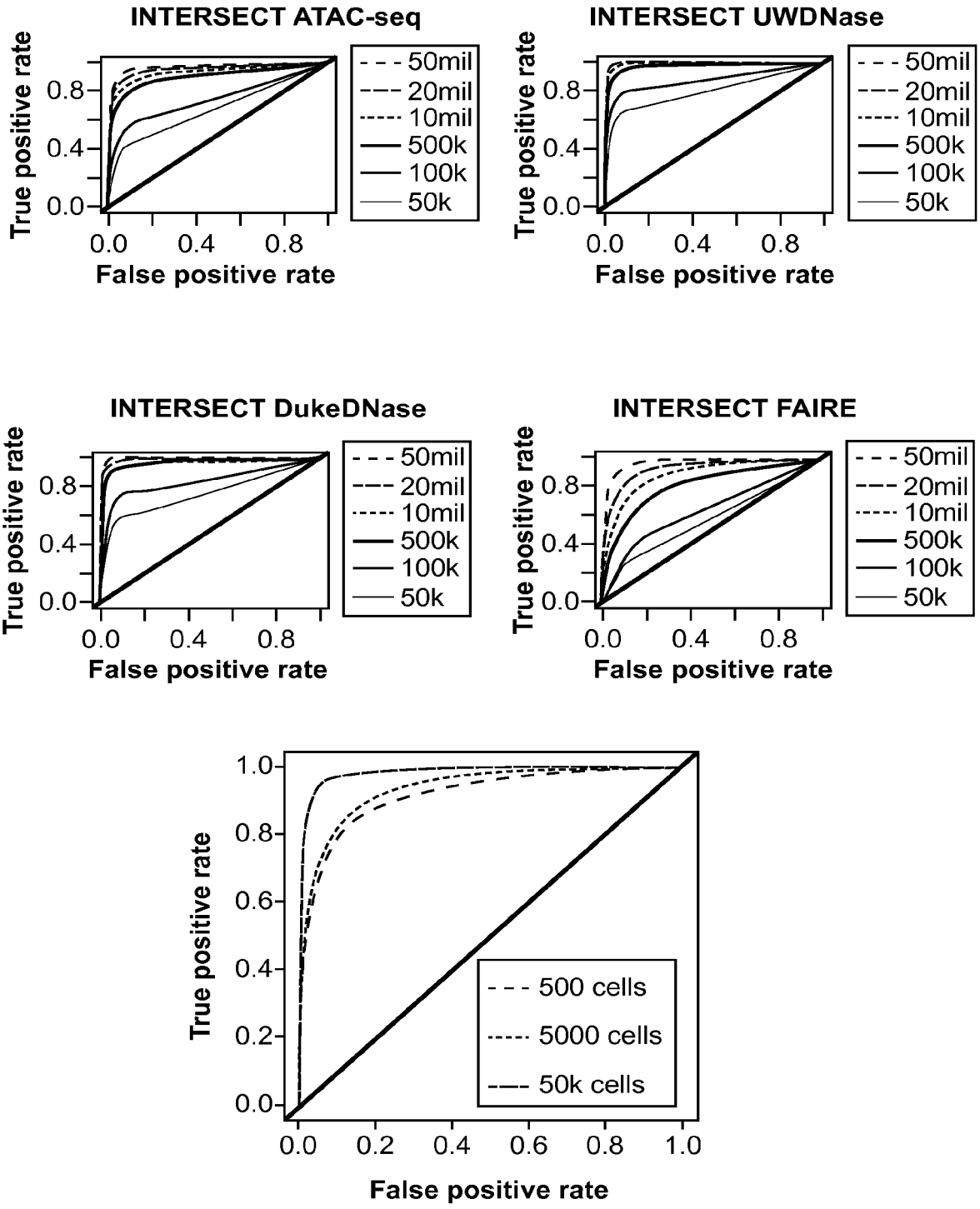


FIG. 8

PARSE EX.1007.074

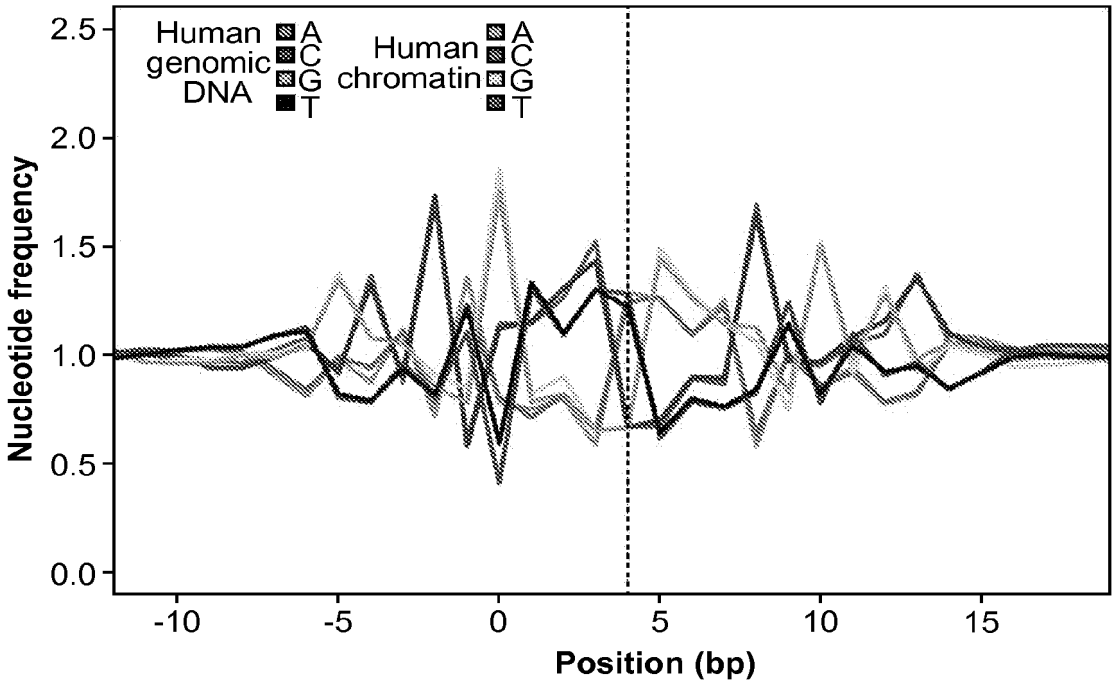


FIG. 9

PARSE EX.1007.075

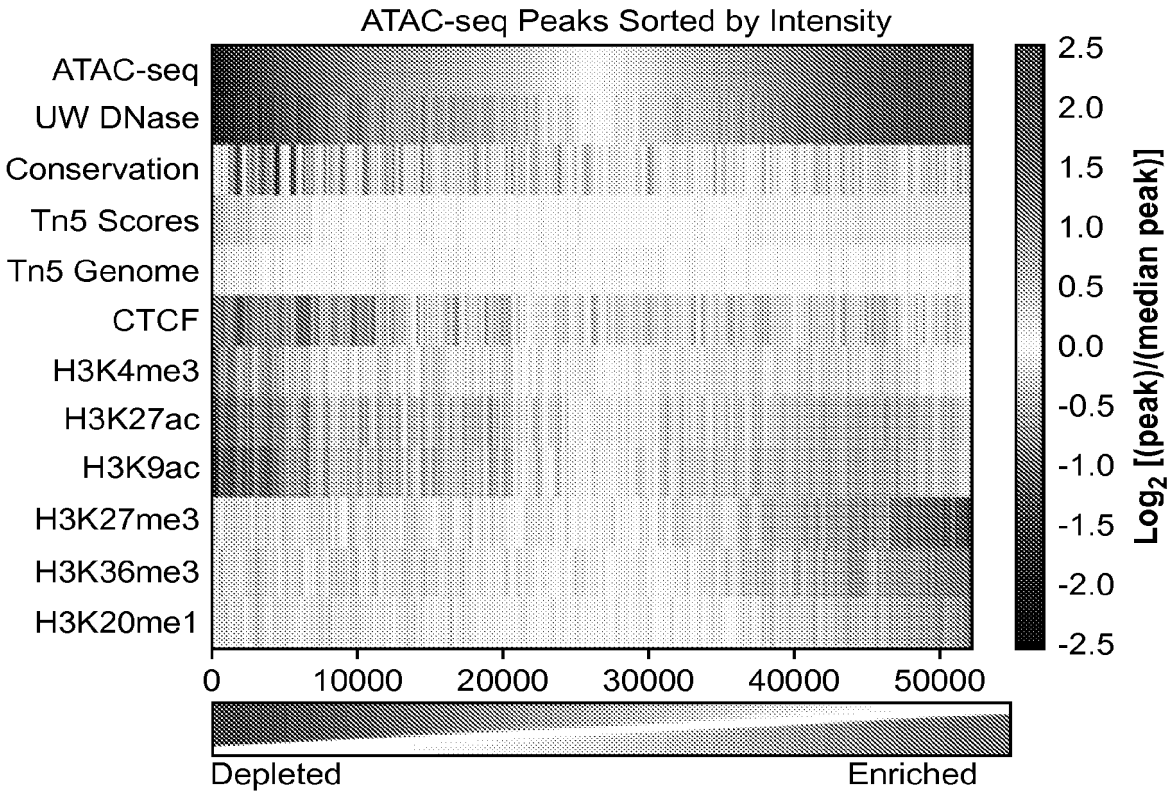


FIG. 10

PARSE EX.1007.076

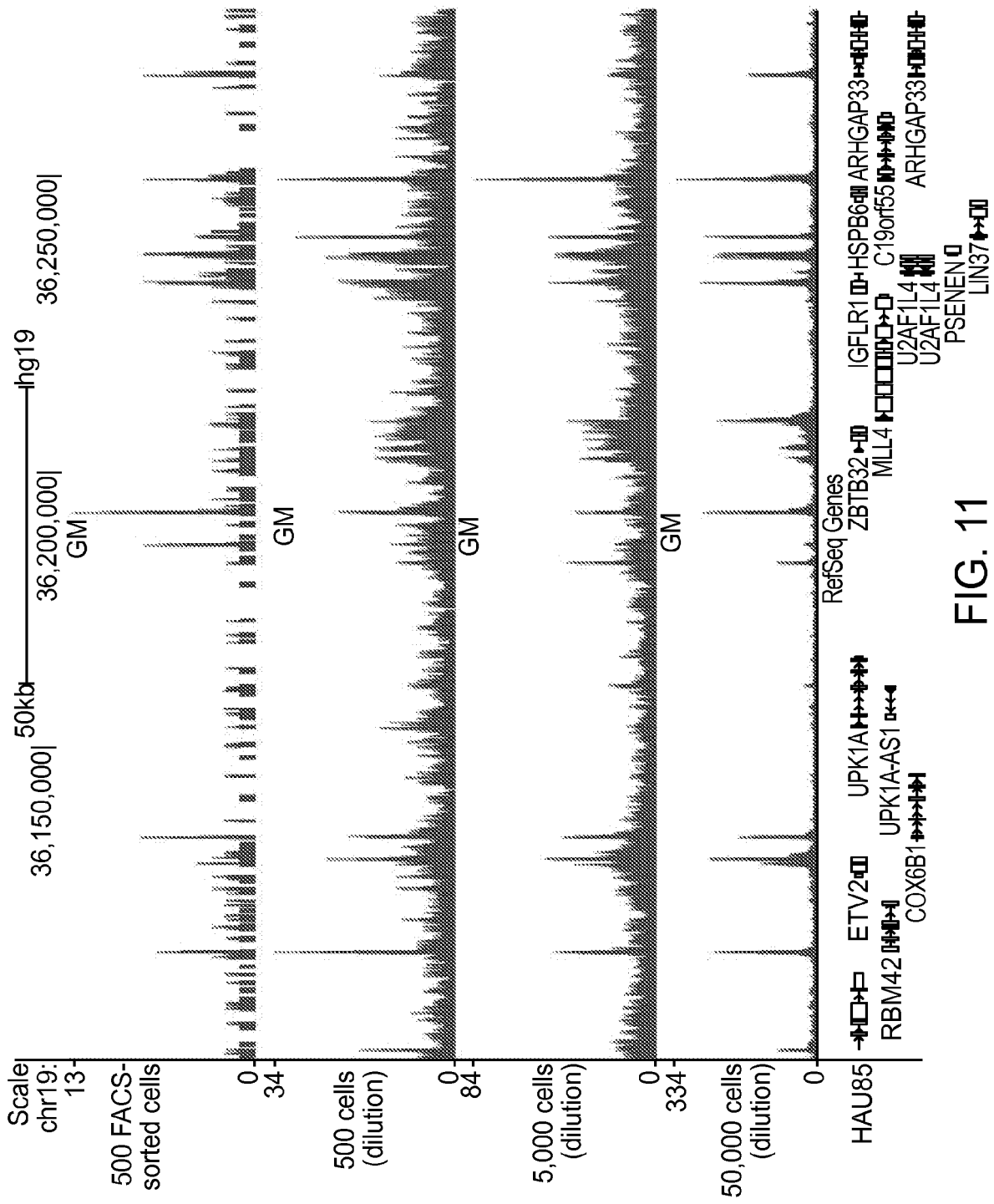


FIG. 11

PARSE EX.1007.077

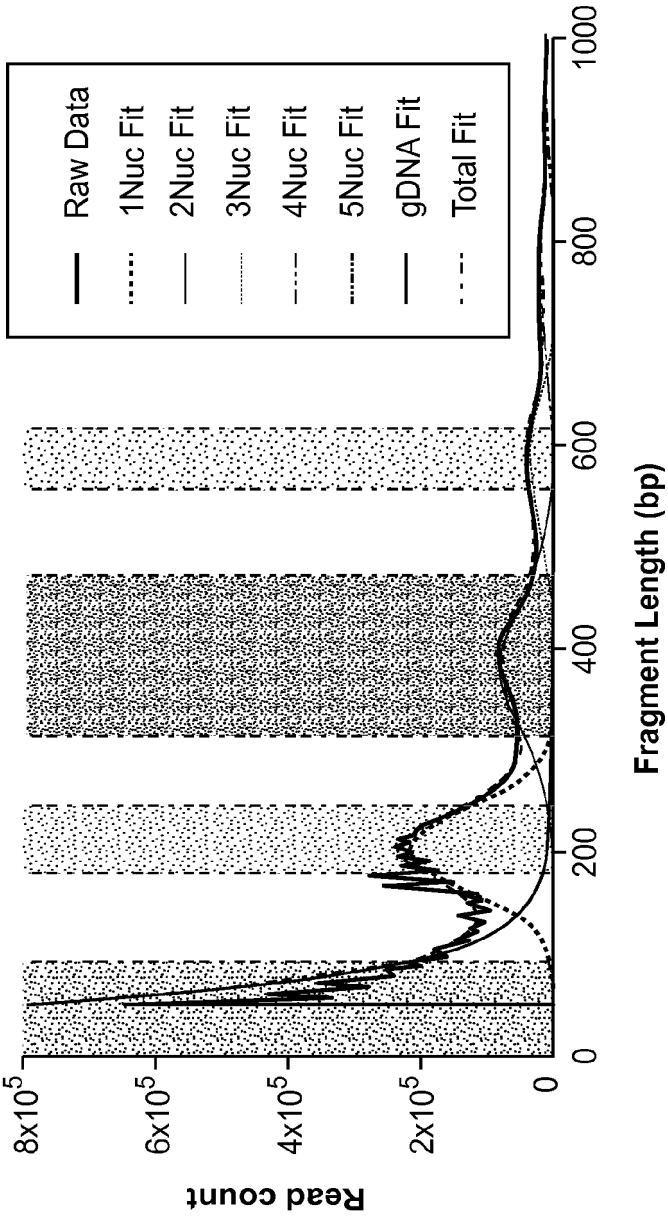


FIG. 12

PARSE EX.1007.078

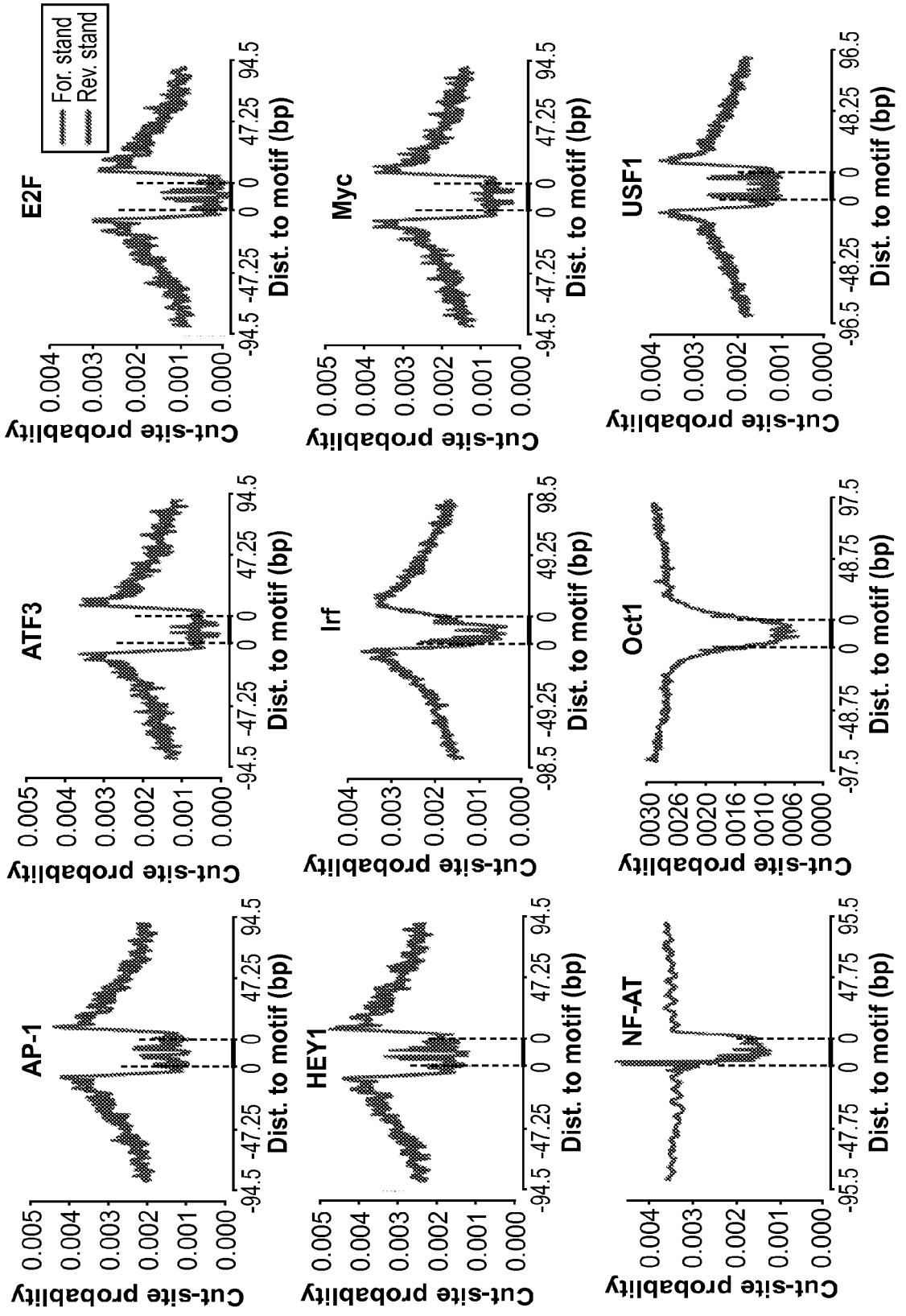


FIG. 13

PARSE EX.1007.079

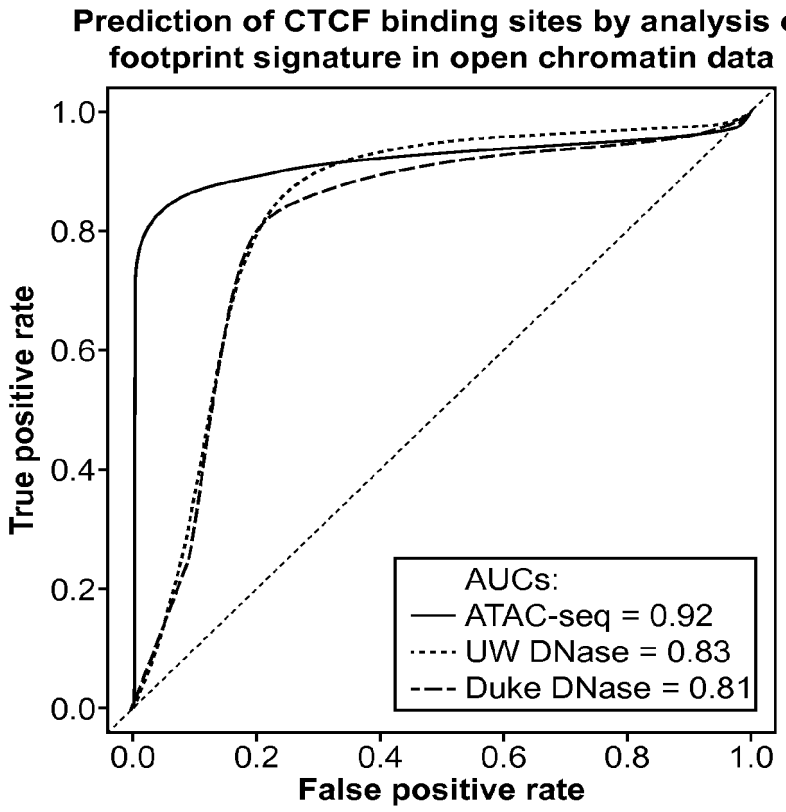


FIG. 14

PARSE EX.1007.080

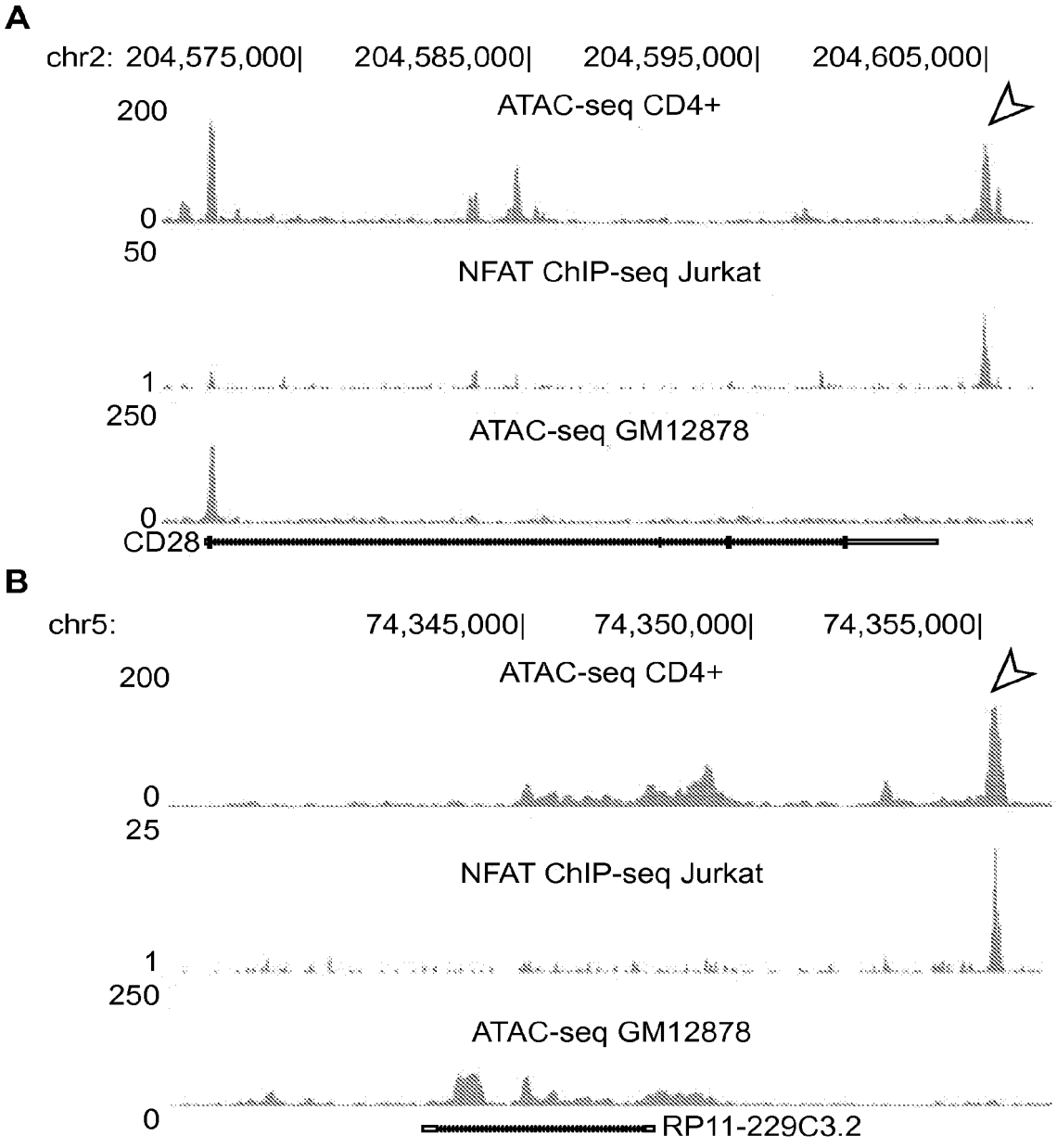


FIG. 15

PARSE EX.1007.081

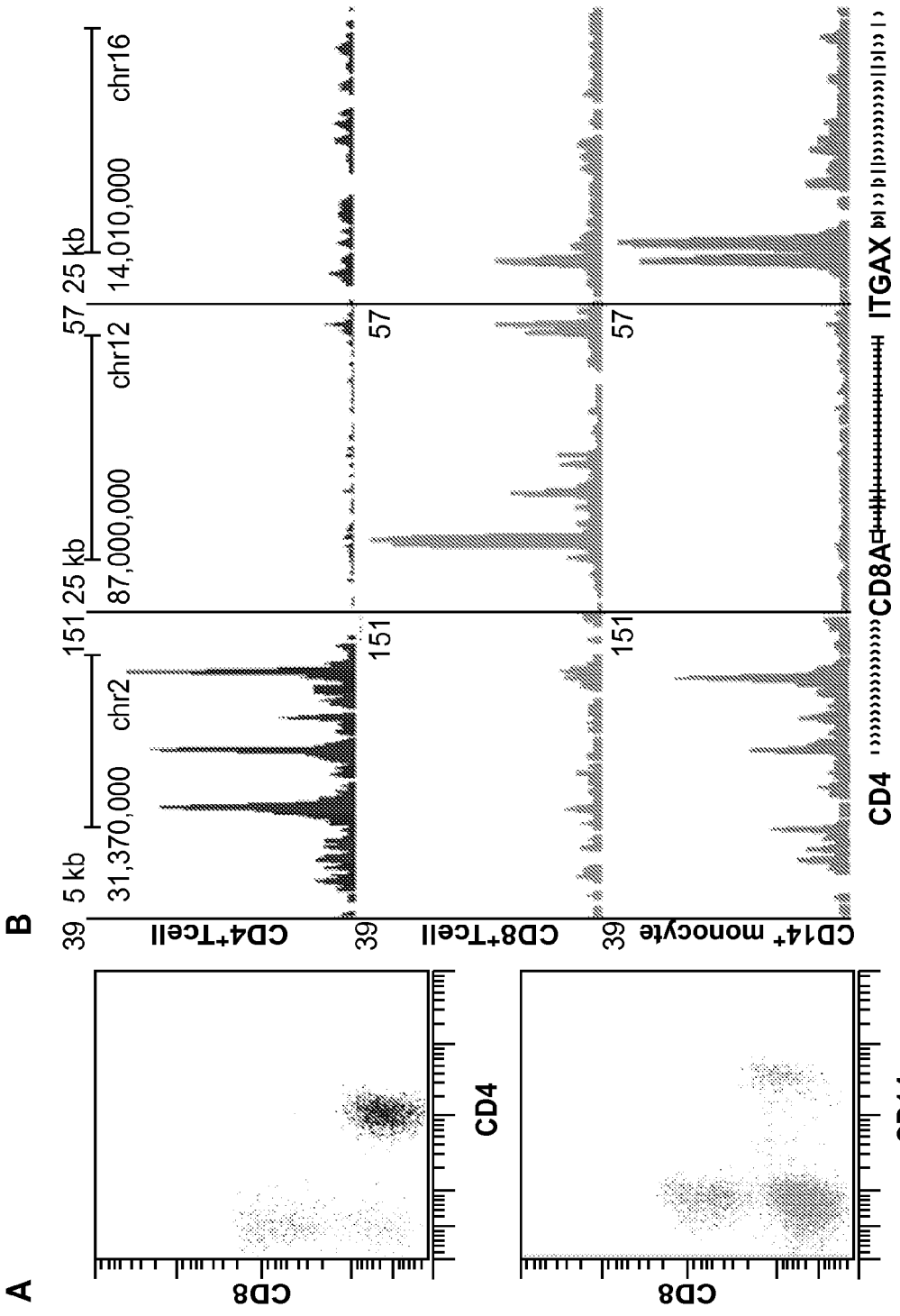


FIG. 16

PARSE EX.1007.082

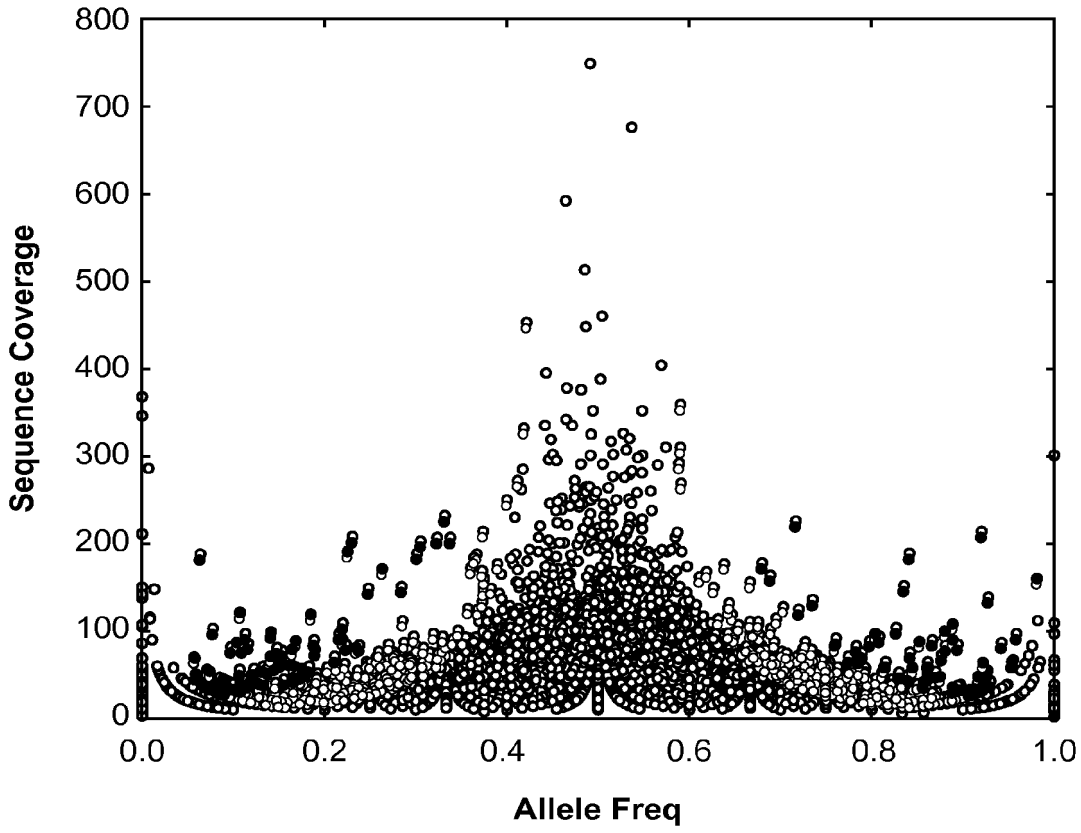


FIG. 17

PARSE EX.1007.083

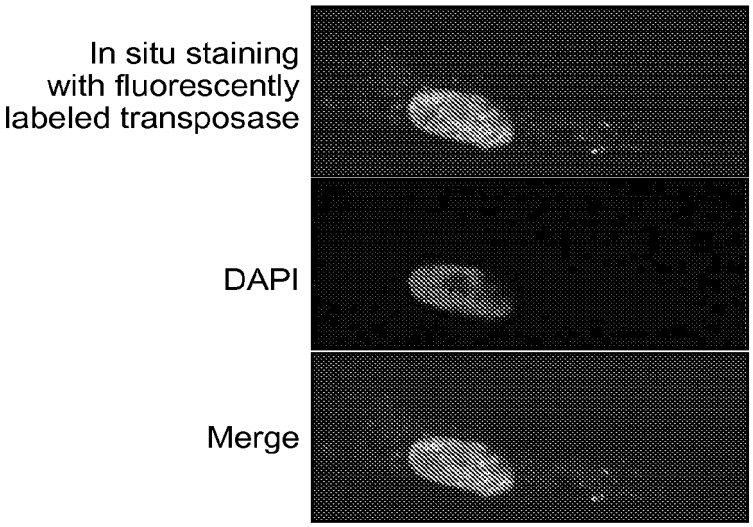


FIG. 18

PARSE EX.1007.084

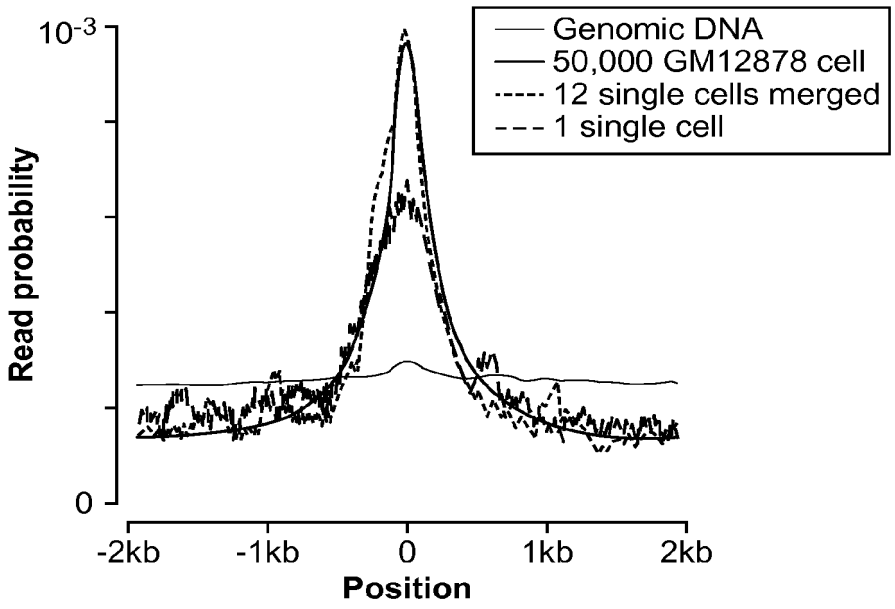


FIG. 19

PARSE EX.1007.085

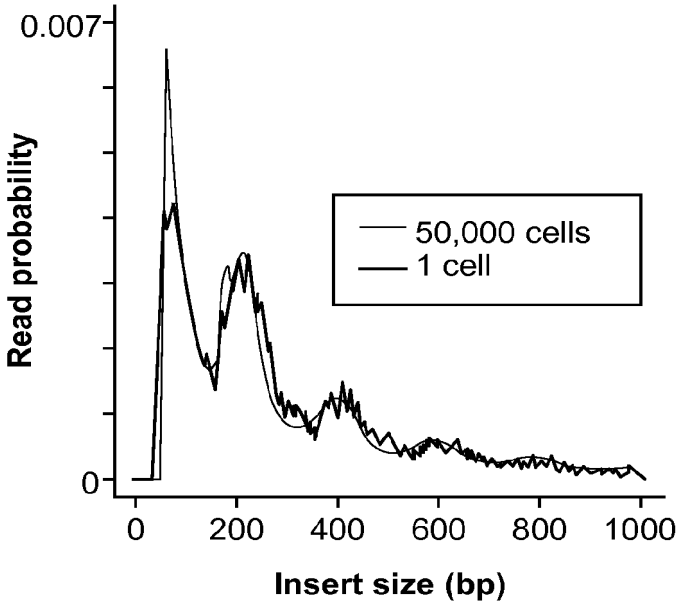


FIG. 20

PARSE EX.1007.086

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82A (07-13)
Approved for use through 01/31/2018. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5, unless the application number and filing date are identified in the Power of Attorney by Applicant form. If neither form PTO/AIA/82A nor form PTO/AIA82B identifies the application to which the Power of Attorney is directed, the Power of Attorney will not be recognized in the application.

Application Number	
Filing Date	
First Named Inventor	Paul GIRESI
Title	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS
Art Unit	
Examiner Name	Not Yet Assigned
Attorney Docket Number	STAN-1111CON

SIGNATURE of Applicant or Patent Practitioner

Signature	/James S. Keddie/	Date (Optional)	July 24, 2018
Name	James S. Keddie	Registration Number	48,920
Title (if Applicant is a juristic entity)			
Applicant Name (if Applicant is a juristic entity)			

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ *Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Doc Code: PA..
Document Description: Power of Attorney

PTO/AIA/82B (07-13)
Approved for use through 11/30/2014. OMB 0651-0051
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number	Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)

☑ I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above: 77974

OR

☐ I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:

☑ The address associated with the above-mentioned Customer Number

OR

☐ The address associated with Customer Number:

OR

Firm or Individual Name					
Address					
City		State		Zip	
Country					
Telephone		Email			

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

The Board of Trustees of the Leland Stanford Junior University

☐ Inventor or Joint Inventor (title not required below)

☐ Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)

☑ Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)

☐ Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature	Katharine Ku	Date (Optional)	August 3, 2015
Name	Katharine Ku		
Title	Director, Office of Technology Licensing		

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.

☑ Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

PARSE EX.1007.088

Original (for SUBMISSION)

VIII-4-1	**Declaration: Inventorship (only for the purposes of the designation of the United States of America)** Declaration of inventorship (Rules 4.17(iv) and 51bis.1(a)(iv)) for the purposes of the designation of the United States of America:	I hereby declare that I believe I am the original inventor or an original joint inventor of a claimed invention in the application. This declaration is directed to international application No. PCT/US2014/038825. I hereby declare that the above-identified international application was made or authorized to be made by me. I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.
VIII-4-1-1-1	Name (LAST, First)	GIRESI, Paul
VIII-4-1-1-2	Residence: (city and either US state, if applicable, or country)	Palo Alto, California
VIII-4-1-1-3	Mailing Address:	264 Durand Way Palo Alto, CA 94304 United States of America
VIII-4-1-1-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-1-5	Date: 6/19/14	
VIII-4-1-2-1	Name (LAST, First)	BUENROSTRO, Jason D.
VIII-4-1-2-2	Residence: (city and either US state, if applicable, or country)	Redwood City, California
VIII-4-1-2-3	Mailing Address:	3523 Altamont Way Redwood City, CA 94062 United States of America
VIII-4-1-2-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-2-5	Date:	
VIII-4-1-3-1	Name (LAST, First)	CHANG, Howard Y.
VIII-4-1-3-2	Residence: (city and either US state, if applicable, or country)	Stanford, California
VIII-4-1-3-3	Mailing Address:	946 Casanueva Place Stanford, CA 94305 United States of America
VIII-4-1-3-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-3-5	Date:	

PARSE EX.1007.089

PCT

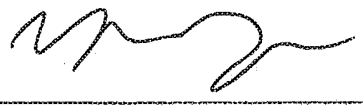
Original (for SUBMISSION)

VIII-4-1	**Declaration: Inventorship (only for the purposes of the designation of the United States of America)** Declaration of Inventorship (Rules 4.17(iv) and 51bis.1(a)(iv)) for the purposes of the designation of the United States of America:	I hereby declare that I believe I am the original inventor or an original joint inventor of a claimed invention in the application. This declaration is directed to international application No. PCT/US2014/038825. I hereby declare that the above-identified international application was made or authorized to be made by me. I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.
VIII-4-1-1-1	Name (LAST, First)	GIRESI, Paul
VIII-4-1-1-2	Residence: (city and either US state, if applicable, or country)	Palo Alto, California
VIII-4-1-1-3	Mailing Address:	264 Durand Way Palo Alto, CA 94304 United States of America
VIII-4-1-1-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-1-5	Date:	
VIII-4-1-2-1	Name (LAST, First)	BUENROSTRO, Jason D.
VIII-4-1-2-2	Residence: (city and either US state, if applicable, or country)	Redwood City, California
VIII-4-1-2-3	Mailing Address:	3523 Altamont Way Redwood City, CA 94062 United States of America
VIII-4-1-2-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	Jason Buenrostro
VIII-4-1-2-5	Date:	9/18/2014
VIII-4-1-3-1	Name (LAST, First)	CHANG, Howard Y.
VIII-4-1-3-2	Residence: (city and either US state, if applicable, or country)	Stanford, California
VIII-4-1-3-3	Mailing Address:	946 Casanueva Place Stanford, CA 94305 United States of America
VIII-4-1-3-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-3-5	Date:	

PARSE EX.1007.090

PCT

Original (for SUBMISSION)

VIII-4-1	**Declaration: Inventorship (only for the purposes of the designation of the United States of America)** Declaration of Inventorship (Rules 4.17(iv) and 51bis.1(a)(iv)) for the purposes of the designation of the United States of America:	I hereby declare that I believe I am the original inventor or an original joint inventor of a claimed invention in the application. This declaration is directed to international application No. PCT/US2014/038825. I hereby declare that the above-identified international application was made or authorized to be made by me. I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.
VIII-4-1-1-1	Name (LAST, First)	GIRESI, Paul
VIII-4-1-1-2	Residence: (city and either US state, if applicable, or country)	Palo Alto, California
VIII-4-1-1-3	Mailing Address:	264 Durand Way Palo Alto, CA 94304 United States of America
VIII-4-1-1-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-1-5	Date:	
VIII-4-1-2-1	Name (LAST, First)	BUENROSTRO, Jason D.
VIII-4-1-2-2	Residence: (city and either US state, if applicable, or country)	Redwood City, California
VIII-4-1-2-3	Mailing Address:	3523 Altamont Way Redwood City, CA 94062 United States of America
VIII-4-1-2-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-2-5	Date:	
VIII-4-1-3-1	Name (LAST, First)	CHANG, Howard Y.
VIII-4-1-3-2	Residence: (city and either US state, if applicable, or country)	Stanford, California
VIII-4-1-3-3	Mailing Address:	946 Casanueva Place Stanford, CA 94305 United States of America
VIII-4-1-3-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-3-5	Date: 7/11/14	

PARSE EX.1007.091

PCT

Original (for SUBMISSION)

VIII-4-1-4-1	Name (LAST, First)	GREENLEAF, William J.
VIII-4-1-4-2	Residence: (city and either US state, if applicable, or country)	Menlo Park, California
VIII-4-1-4-3	Mailing Address:	2372 Branner Drive Menlo Park, CA 94025 United States of America
VIII-4-1-4-4	Inventor's Signature: (The signature must be that of the inventor, not that of the agent)	
VIII-4-1-4-5	Date:	18 Aug 2015

PARSE EX.1007.092

Electronically Filed

INFORMATION DISCLOSURE STATEMENT Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450		
	Attorney Docket	STAN-1111CON
	First Named Inventor	Paul GIRESI
	Application Number	TBD
	Confirmation No.	TBD
	Filing Date	HEREWITH
	Group Art Unit	TBD
	Examiner Name	TBD
	Title: *"Transposition of Native Chromatin for Personal Epigenomics"*	

Sir:

Applicant submits herewith documents which may be material to the examination of this application and in respect of which there may be a duty to disclose in accordance with 37 C.F.R. § 1.56. This submission is not intended to constitute an admission that any document referred to therein is "prior art" for this invention unless specifically designated as such. A listing of the documents is shown on enclosed Form PTO/SB/08A and copies of the foreign patents and non-patent literature are also enclosed.

Applicant would appreciate the Examiner initialing and returning the initialed copy of form PTO/SB/08A, indicating the documents cited therein have been considered and made of record herein.

All of the references identified herein were disclosed in parent application serial number 14/784,250, and as such, copies thereof are not included pursuant to the provisions of 37 CFR § 1.98(d).

Statements

☒ **No statement**

Fees

☒ No fee is believed to be due.

☐ The appropriate fee set forth in 37 C.F.R. §1.17(p) accompanies this information disclosure statement.

Respectfully submitted,
BOZICEVIC, FIELD & FRANCIS LLP

Date: July 24, 2018

By: /James S. Keddie/
James S. Keddie
Reg. No. 48,920

BOZICEVIC, FIELD & FRANCIS LLP
201 Redwood Shores Parkway, Suite 200
Redwood City, CA 94065
Telephone: (650) 327-3400
Facsimile: (650) 327-3231

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	
First Named Inventor	Paul Giresi
Art Unit	
Examiner Name	
Attorney Docket Number	STAN-1111CON

U.S.PATENTS						
Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5137829		1992-08-11	Nag et al.	
	2	5965443		1999-10-12	Reznikoff et al.	
	3	6159717		2000-12-12	Savakis et al.	
	4	6207384		2001-03-27	Mekalanos et al.	
	5	6258571		2001-07-10	Chumakov et al.	
	6	6291243		2001-09-18	Fogarty et al.	
	7	6294385		2001-09-25	Goryshin et al.	
	8	6593113		2003-07-15	Tenkanen et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	9	6613752		2003-09-02	Kay et al.	
	10	7608451		2009-10-27	Cooper et al.	
	11	7745218		2010-06-29	Kim et al.	
	12	7608434		2009-10-27	Reznikoff et al.	
	13	7138267		2006-11-21	Jendrisak et al.	
	14	8124404		2012-02-28	Alphey	
	15	8420386		2013-04-16	Ivics et al.	
	16	8598328		2013-12-03	Koga et al.	
	17	8829171		2014-09-09	Steemers et al.	
	18	8927218		2015-01-06	Forsyth	
	19	9005935		2015-04-14	Belyaev	

PARSE EX.1007.096

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	20	9074251		2015-07-07	Steemers et al.	
	21	9080211		2015-07-14	Grunenwald et al.	
	22	9150916		2015-10-06	Christen et al.	
	23	9175295		2015-11-03	Kaminaka et al.	
	24	9238671		2016-01-19	Goryshin et al.	
	25	9328382		2016-05-03	Drmanac et al.	
	26	9574226		2017-02-21	Gormley et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20030032141		2003-02-13	Nguyen	
	2	20040101880		2004-05-27	Rozwadowski et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	3	20100120098		2010-05-13	GRUNENWALD, et al.	
	4	20120297493		2012-11-22	Cooper et al.	
	5	20130203605		2013-08-08	Shendure et al.	
	6	20150111788		2015-04-23	Fernandez et al.	
	7	20150337298		2015-11-26	Xi et al.	
	8	20150368638		2015-12-24	Steemers et al.	
	9	20150376608		2015-12-31	Kaper et al.	
	10	20160060691		2016-03-03	Giresi et al.	
	11	20160115474		2016-04-28	Jelinek et al.	
	12	20160122753		2016-05-05	Mikkelsen et al.	
	13	20160160235		2016-06-09	Solodushko et al.	

PARSE EX.1007.098

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	14	20160177359		2016-06-23	Ukanis et al.	
	15	20160208323		2016-07-21	Bernstein et al.	
	16	20160376663		2016-12-29	Brown	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	WO 2007/093819	WO		2007-08-23	ISIS INNOVATION LIMITED		☐
	2	WO 2009/147386	WO		2009-12-10	OXFORD BIODYNAMICS LIMITED		☐
	3	WO 2010/048605	WO		2010-04-29	EPICENTRE TECHNOLOGIES CORPORATION		☐
	4	WO 2014/108810	WO		2014-07-17	ILLUMINA CAMBRIDGE LTD		☐
	5	WO 2016/191618	WO		2016-12-01	ZHENG JIANBIAO et al.		☐
	6	WO 2016/207647	WO		2016-12-29	OXFORD BIODYNAMICS LIMITED		☐

PARSE EX.1007.099

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	7	WO 2016/207653	WO		2016-12-29	OXFORD BIODYNAMICS LIMITED		☐
	8	WO 2016/207661	WO		2016-12-29	OXFORD BIODYNAMICS LIMITED		☐
	9	WO 2017/015075	WO		2017-01-26	HARVARD COLLEGE		☐
	10	WO 2017/025594	WO		2017-02-16	CHEM FORSCHUNGSZENTRUM FÜR MOLEKULARE MEDIZIN GMBM		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	ADEY, et al., "Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing", Genome Research, 2012, 22 (6): 1139-1143.	☐
	2	ASON et al., "DNA Sequence Bias During Tn5 Transposition", J. Mol. Biol., 2004, 335: 1213-1255.	☐
	3	BJORNSSON et al., "Intra-individiual change over time in DNA methylation with familial clustering, JAMA, June 25, 2008, Vol. 299 No. 24, pages 2877-2883	☐
	4	BOYLE, et al., "High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells", Genome Res. 2011 Mar;21(3):456-64	☐
	5	BUENROSTRO, et al., "Tranposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position", Nature Methods, 2013, 10(12): 1213-1218	☐

PARSE EX.1007.100

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	6	GANGADHARAN et al., "DNA transposon Hermes insert into DNA in nucleosome-free regions in vivo, Proc nat Ad Sci, December 21, 2010, Vol. 107, No. 51, pages 1966-1972.	☐
	7	GREEN et al., "Insertion site preference of Mu, Tn5, and Tn7 transposons", Mobile DNA, 2012, 3:3, 6 pages.	☐
	8	Parks et al., "ChIP-Seq: advantages and challenges of a maturing technology", Nature, vol. 10, pages 669-680, October (Year: 2009)	☐
	9	Haring et al., "Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization", Plant Methods, 3 (11): 1-25 (Year: 2007)	☐
	10	SIMON, et al., "Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA", Nature Protocols, 2012, 7(2): 256-267.	☐
	11	SONG, et al., "DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells", Cold Spring Harbor Laboratory Press, 2010, 2010(2), doi:10.1101/pdb.prot5384	☐
	12	ZENTNER et al., "Surveying the epigenomic landscape, one base at a time", Genome Biology, 2012, 13:250, 8 pages.	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

PARSE EX.1007.101

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☐ See attached certification statement.

☐ The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/James S. Keddie/	Date (YYYY-MM-DD)	2018-07-24
Name/Print	James S. Keddie	Registration Number	48,920

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

PARSE EX.1007.102

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PARSE EX.1007.103

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS
First Named Inventor/Applicant Name:	Paul GIRESI
Filer:	James Samuel Keddie/Tammy Hu
Attorney Docket Number:	STAN-1111CON

Filed as Small Entity

Filing Fees for Track I Prioritized Examination - Nonprovisional Application under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY FILING FEE (ELECTRONIC FILING)	4011	1	75	75
UTILITY SEARCH FEE	2111	1	330	330
UTILITY EXAMINATION FEE	2311	1	380	380
REQUEST FOR PRIORITIZED EXAMINATION	2817	1	2000	2000
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	2202	3	50	150
Miscellaneous-Filing:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
PUBL. FEE- EARLY, VOLUNTARY, OR NORMAL	1504	1	0	0
PROCESSING FEE, EXCEPT PROV. APPLS.	2830	1	70	70

Petition:	
Patent-Appeals-and-Interference:	
Post-Allowance-and-Post-Issuance:	
Extension-of-Time:	
Miscellaneous:	
Total in USD ($)	**3005**

Electronic Acknowledgement Receipt

EFS ID:	33265675
Application Number:	16043874
International Application Number:	
Confirmation Number:	2837
Title of Invention:	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS
First Named Inventor/Applicant Name:	Paul GIRESI
Customer Number:	77974
Filer:	James Samuel Keddie/Tammy Hu
Filer Authorized By:	James Samuel Keddie
Attorney Docket Number:	STAN-1111CON
Receipt Date:	24-JUL-2018
Filing Date:	
Time Stamp:	15:50:16
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$3005
RAM confirmation Number	072518INTEFSW15512600
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	TrackOne Request	STAN-1111CON_Track-I-Request.pdf	136187 da014cdf4eb8693f0d064c19423ae3251b59fabc	no	2
Warnings:					
Information:					
2	Application Data Sheet	STAN-1111CON_ADS_cute.pdf	120074 caabac9b11c5eae2a71e28111bb2659b7b0e949d	no	8
Warnings:					
Information:					
This is not an USPTO supplied ADS fillable form					
3		STAN-1111CON_Prelim_Amendment.pdf	56687 4aa6292ac8ed6686f06e7d454940e2db4deb95ba	yes	6

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Preliminary Amendment	1	1
	Specification	2	2
	Claims	3	5
	Applicant Arguments/Remarks Made in an Amendment	6	6

Warnings:					
Information:					
4		STAN-1111_Application.pdf	1082964 70826ecaba8df38dd90ea0f9af077a77e1361580	yes	80

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**

	Abstract		1		1
	Specification		2		48
	Claims		49		56
	Drawings-other than black and white line drawings		57		80
Warnings:					
Information:					
5	Sequence Listing (Text File)	STAN-1111_Seq_Listing.txt	5239 	no	-
Warnings:					
Information:					
6	Power of Attorney	STAN-1111CON_GPOA.pdf	2291510 04374230ef0068095a55cf93d7ffef94e900ae0d	no	2
Warnings:					
Information:					
7	Oath or Declaration filed	STAN-1111_Declaration.pdf	2631268 33851ffd2e22ea833687b722be9a9e464b6a3655	no	4
Warnings:					
Information:					
8	Transmittal Letter	STAN-1111CON_IDS-Transmittal.pdf	30473 985de7f3cb7d75a13a9b0e3213267bdbb1395876	no	2
Warnings:					
Information:					
9	Information Disclosure Statement (IDS) Form (SB08)	STAN-1111CON_IDS_cute.pdf	83573 21ad71dc4e1c68463917218d36b5e8c6545cb29b	no	9
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					

10	Fee Worksheet (SB06)	fee-info.pdf	41135 59e1d89fba74fade62fe4603c43045cba7de296d	no	2

Warnings:	
Information:	
Total Files Size (in bytes):	6479110

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Doc Code: TRACK1.REQ
Document Description: TrackOne Request

PTO/AIA/424 (04-14)

CERTIFICATION AND REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 CFR 1.102(e) (Page 1 of 1)

First Named Inventor:	GIRESI Paul	Nonprovisional Application Number (if known):	
Title of Invention:	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS		

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS PRIORITIZED EXAMINATION FOR THE ABOVE-IDENTIFIED APPLICATION.

1. The processing fee set forth in 37 CFR 1.17(i)(1) and the prioritized examination fee set forth in 37 CFR 1.17(c) have been filed with the request. The publication fee requirement is met because that fee, set forth in 37 CFR 1.18(d), is currently $0. The basic filing fee, search fee, and examination fee are filed with the request or have been already been paid. I understand that any required excess claims fees or application size fee must be paid for the application.

2. I understand that the application may not contain, or be amended to contain, more than four independent claims, more than thirty total claims, or any multiple dependent claims, and that any request for an extension of time will cause an outstanding Track I request to be dismissed.

3. The applicable box is checked below:

I. ☑ Original Application (Track One) - Prioritized Examination under § 1.102(e)(1)

i. (a) The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a). This certification and request is being filed with the utility application via EFS-Web.

---OR---

(b) The application is an original nonprovisional plant application filed under 35 U.S.C. 111(a). This certification and request is being filed with the plant application in paper.

ii. An executed inventor's oath or declaration under 37 CFR 1.63 or 37 CFR 1.64 for each inventor, **or** the application data sheet meeting the conditions specified in 37 CFR 1.53(f)(3)(i) is filed with the application.

II. ☐ Request for Continued Examination - Prioritized Examination under § 1.102(e)(2)

i. A request for continued examination has been filed with, or prior to, this form.
ii. If the application is a utility application, this certification and request is being filed via EFS-Web.
iii. The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a), or is a national stage entry under 35 U.S.C. 371.
iv. This certification and request is being filed prior to the mailing of a first Office action responsive to the request for continued examination.
v. No prior request for continued examination has been granted prioritized examination status under 37 CFR 1.102(e)(2).

Signature /James S. Keddie/	Date July 24, 2018
Name (Print/Typed) James S. Keddie	Practitioner Registration Number 48920

Note: *This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.**

☑ *Total of 1 forms are submitted.

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	
Title of Invention	Transposition of Native Chromatin for Personal Epigenomics	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Paul		Giresi	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Palo Alto	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	264 Durand Way		
Address 2			
City	Palo Alto	State/Province	CA
Postal Code	94304	Country i	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Jason	D.	Buenrostro	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

City	Redwood City	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	3523 Altamont Way		
Address 2			
City	Redwood City	State/Province	CA
Postal Code	94062	Country i	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Howard	Y.	Chang	

Residence Information (Select One) ◉ US Residency ○ Non US Residency ○ Active US Military Service

PARSE EX.1007.112

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	
Title of Invention	Transposition of Native Chromatin for Personal Epigenomics	

City	Stanford	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	946 Casanueva Place		
Address 2			
City	Stanford	State/Province	CA
Postal Code	94305	Country i	US

Inventor 4 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	William	J.	Greenleaf	

Residence Information (Select One) (•) US Residency ( ) Non US Residency ( ) Active US Military Service

City	Menlo Park	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	765A Live Oak		
Address 2			
City	Menlo Park	State/Province	CA
Postal Code	94025	Country i	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	77974		
Email Address	docket@bozpat.com	Add Email	Remove Email

Application Information:

Title of the Invention	Transposition of Native Chromatin for Personal Epigenomics		
Attorney Docket Number	STAN-1111CON	Small Entity Status Claimed	☒
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	24	Suggested Figure for Publication (if any)	

Filing By Reference :

PARSE EX.1007.113

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	
Title of Invention	Transposition of Native Chromatin for Personal Epigenomics	

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	◉ Customer Number	○ US Patent Practitioner	○ Limited Recognition (37 CFR 11.9)
Customer Number	77974		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, or 365(c) or indicate National Stage entry from a PCT application. Providing this information in the application data sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the application number blank.

Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)
	Continuation of	14784250	2015-10-13
Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)
14784250	a 371 of international	PCT/US2014/038825	2014-05-20
Prior Application Status	Expired		Remove

PARSE EX.1007.114

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	

Title of Invention	Transposition of Native Chromatin for Personal Epigenomics

Application Number	Continuity Type	Prior Application Number	Filing Date (YYYY-MM-DD)
PCT/US2014/038825	Claims benefit of provisional	61826728	2013-05-23

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button.

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55(d). When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(h)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Remove

Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button.

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

Authorization to Permit Access:

☒ Authorization to Permit Access to the Instant Application by the Participating Offices

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	
Title of Invention	Transposition of Native Chromatin for Personal Epigenomics	

If checked, the undersigned hereby grants the USPTO authority to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the World Intellectual Property Office (WIPO), and any other intellectual property offices in which a foreign application claiming priority to the instant patent application is filed access to the instant patent application. See 37 CFR 1.14(c) and (h). This box should not be checked if the applicant does not wish the EPO, JPO, KIPO, WIPO, or other intellectual property office in which a foreign application claiming priority to the instant patent application is filed to have access to the instant patent application.

In accordance with 37 CFR 1.14(h)(3), access will be provided to a copy of the instant patent application with respect to: 1) the instant patent application-as-filed; 2) any foreign application to which the instant patent application claims priority under 35 U.S.C. 119(a)-(d) if a copy of the foreign application that satisfies the certified copy requirement of 37 CFR 1.55 has been filed in the instant patent application; and 3) any U.S. application-as-filed from which benefit is sought in the instant patent application.

In accordance with 37 CFR 1.14(c), access may be provided to information concerning the date o f filing this Authorization.

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

○ Assignee	○ Legal Representative under 35 U.S.C. 117	○ Joint Inventor
◉ Person to whom the inventor is obligated to assign.	○ Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor :

If the Applicant is an Organization check here. ☒

Organization Name	The Board of Trustees of the Leland Stanford Junior University

Mailing Address Information For Applicant:

Address 1	Office of the General Counsel, Building 170		
Address 2	3rd Floor, Main Quad, P.O. Box 20386		
City	Stanford	**State/Province**	CA
Country	US	Postal Code	94305-2038
Phone Number		Fax Number	

PARSE EX.1007.116

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	

Title of Invention	Transposition of Native Chromatin for Personal Epigenomics

Email Address	

Additional Applicant Data may be generated within this form by selecting the Add button.

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not subsitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication . An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

If the Assignee or Non-Applicant Assignee is an Organization check here. ☐

Prefix	**Given Name**	Middle Name	**Family Name**	Suffix

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1			
Address 2			
City		**State/Province**	
Country i		Postal Code	
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button.

Signature:

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Signature	/James S. Keddie/			Date (YYYY-MM-DD)	2018-07-24
First Name	James S.	Last Name	Keddie	Registration Number	48920

Additional Signature may be generated within this form by selecting the Add button.

PARSE EX.1007.117

PTO/AIA/14 (12-13)
Approved for use through 01/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	STAN-1111CON
	Application Number	
Title of Invention	Transposition of Native Chromatin for Personal Epigenomics	

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

PARSE EX.1007.118

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent C o o p eration Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PARSE EX.1007.119

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 16043874 Document Date: 07/24/2018

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 16043874 Document Date: 07/24/2018

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via eDAN using the Supplemental Content tab, or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: August 26, 2013

==========

Sequence Listing was accepted.

See attached Validation Report.

If you need help call the Patent Electronic Business Center at (866) 217-9197 (toll free).

Reviewer: Wheat Jr, Scott (ASRC)

Timestamp: [year=2018; month=7; day=27; hr=9; min=24; sec=11; ms=684;]

==========

Validated By CRFValidator v 1.0.5

Application No: 16043874 Version No: 1.0

Input Set:

Output Set:

Started: 2018-07-24 16:09:50.535
Finished: 2018-07-24 16:09:51.366
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 831 ms
Total Warnings: 25
Total Errors: 1
No. of SeqIDs Defined: 25
Actual SeqID Count: 25

Error code		Error Description
E	201	Mandatory field data missing in <140>
W	213	Artificial or Unknown found in <213> in SEQ ID (1)
W	213	Artificial or Unknown found in <213> in SEQ ID (2)
W	213	Artificial or Unknown found in <213> in SEQ ID (3)
W	213	Artificial or Unknown found in <213> in SEQ ID (4)
W	213	Artificial or Unknown found in <213> in SEQ ID (5)
W	213	Artificial or Unknown found in <213> in SEQ ID (6)
W	213	Artificial or Unknown found in <213> in SEQ ID (7)
W	213	Artificial or Unknown found in <213> in SEQ ID (8)
W	213	Artificial or Unknown found in <213> in SEQ ID (9)
W	213	Artificial or Unknown found in <213> in SEQ ID (10)
W	213	Artificial or Unknown found in <213> in SEQ ID (11)
W	213	Artificial or Unknown found in <213> in SEQ ID (12)
W	213	Artificial or Unknown found in <213> in SEQ ID (13)
W	213	Artificial or Unknown found in <213> in SEQ ID (14)
W	213	Artificial or Unknown found in <213> in SEQ ID (15)
W	213	Artificial or Unknown found in <213> in SEQ ID (16)
W	213	Artificial or Unknown found in <213> in SEQ ID (17)
W	213	Artificial or Unknown found in <213> in SEQ ID (18)
W	213	Artificial or Unknown found in <213> in SEQ ID (19)

Input Set:

Output Set:

Started: 2018-07-24 16:09:50.535
Finished: 2018-07-24 16:09:51.366
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 831 ms
Total Warnings: 25
Total Errors: 1
No. of SeqIDs Defined: 25
Actual SeqID Count: 25

Error code	Error Description
W 213	Artificial or Unknown found in <213> in SEQ ID (20) This error has occured more than 20 times, will not be displayed

SEQUENCE LISTING

<110> Buenrostro, Jason
Chang, Howard Y
Greenleaf, William J
Giresi, Paul

<120> TRANSPOSITION INTO NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

<130> STAN-1111WO

<140> US 16/043,874
<141> 2018-07-24

<150> US 61/826,728
<151> 2013-05-23

<160> 25

<170> PatentIn version 3.5

<210> 1
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic polynucleotide

<400> 1
aatgatacgg cgaccaccga gatctacact cgtcggcagc gtcagatgtg 50

<210> 2
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 2
caagcagaag acggcatacg agattcgcct tagtctcgtg ggctcggaga tgt 53

<210> 3
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 3
caagcagaag acggcatacg agatctagta cggtctcgtg ggctcggaga tgt 53

<210> 4
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 4
caagcagaag acggcatacg agatttctgc ctgtctcgtg ggctcggaga tgt 53

<210> 5
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 5
caagcagaag acggcatacg agatgctcag gagtctcgtg ggctcggaga tgt 53

<210> 6
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 6
caagcagaag acggcatacg agataggagt ccgtctcgtg ggctcggaga tgt 53

<210> 7
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 7
caagcagaag acggcatacg agatcatgcc tagtctcgtg ggctcggaga tgt 53

<210> 8
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 8

caagcagaag acggcatacg agatgtagag aggtctcgtg ggctcggaga tgt 53

<210> 9
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 9
caagcagaag acggcatacg agatcctctc tggtctcgtg ggctcggaga tgt 53

<210> 10
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 10
caagcagaag acggcatacg agatagcgta gcgtctcgtg ggctcggaga tgt 53

<210> 11
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 11
caagcagaag acggcatacg agatcagcct cggtctcgtg ggctcggaga tgt 53

<210> 12
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 12
caagcagaag acggcatacg agattgcctc ttgtctcgtg ggctcggaga tgt 53

<210> 13
<211> 53
<212> DNA
<213> Artificial sequence

<220>

<223> synthetic polynucleotide

<400> 13
caagcagaag acggcatacg agattcctct acgtctcgtg ggctcggaga tgt 53

<210> 14
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 14
caagcagaag acggcatacg agatatcacg acgtctcgtg ggctcggaga tgt 53

<210> 15
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 15
caagcagaag acggcatacg agatacagtg gtgtctcgtg ggctcggaga tgt 53

<210> 16
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 16
caagcagaag acggcatacg agatcagatc cagtctcgtg ggctcggaga tgt 53

<210> 17
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 17
caagcagaag acggcatacg agatacaaac gggtctcgtg ggctcggaga tgt 53

<210> 18
<211> 53
<212> DNA

<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 18
caagcagaag acggcatacg agatacccag cagtctcgtg ggctcggaga tgt 53

<210> 19
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 19
caagcagaag acggcatacg agataacccc tcgtctcgtg ggctcggaga tgt 53

<210> 20
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 20
caagcagaag acggcatacg agatcccaac ctgtctcgtg ggctcggaga tgt 53

<210> 21
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 21
caagcagaag acggcatacg agatcaccac acgtctcgtg ggctcggaga tgt 53

<210> 22
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 22
caagcagaag acggcatacg agatgaaacc cagtctcgtg ggctcggaga tgt 53

<210> 23
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 23
caagcagaag acggcatacg agattgtgac cagtctcgtg ggctcggaga tgt 53

<210> 24
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 24
caagcagaag acggcatacg agatagggtc aagtctcgtg ggctcggaga tgt 53

<210> 25
<211> 53
<212> DNA
<213> Artificial sequence

<220>
<223> synthetic polynucleotide

<400> 25
caagcagaag acggcatacg agataggagt gggtctcgtg ggctcggaga tgt 53

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
16/043,874	07/24/2018	2676	935	STAN-1111CON	23	1

CONFIRMATION NO. 2837

77974
Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

FILING RECEIPT



Date Mailed: 09/07/2018

Receipt is acknowledged of this non-provisional patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Paul Giresi, Palo Alto, CA;
Jason D. Buenrostro, Redwood City, CA;
Howard Y. Chang, Stanford, CA;
William J. Greenleaf, Menlo Park, CA;

Applicant(s)

The Board of Trustees of the Leland Stanford Junior University, Stanford, CA

Power of Attorney: The patent practitioners associated with Customer Number 77974

Domestic Priority data as claimed by applicant

This application is a CON of 14/784,250 10/13/2015 PAT 10059989
which is a 371 of PCT/US2014/038825 05/20/2014
which claims benefit of 61/826,728 05/23/2013

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: No

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

PARSE EX.1007.131

If Required, Foreign Filing License Granted: 09/06/2018

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 16/043,874**

Projected Publication Date: 12/13/2018

Non-Publication Request: No

Early Publication Request: No
**** SMALL ENTITY ****
Title

TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

Preliminary Class

358

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

PARSE EX.1007.132

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

PARSE EX.1007.133

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 16/043,874

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY FEE($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	75		N/A	
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A	330		N/A	
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	380		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	23 minus 20 =	* 3	x 50 =	150	OR		
INDEPENDENT CLAIMS (37 CFR 1.16(h))	1 minus 3 =	*	x 230 =	0.00			
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).			0.00			
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))				0.00			
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL	935		TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE($)	SMALL ENTITY ADDITIONAL FEE($)	OR	OTHER THAN SMALL ENTITY RATE($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".

The "Highest Number Previously Paid For" (Total or Independent) is the highest found in the appropriate box in column 1.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON	2837

77974 7590 09/19/2018

Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

EXAMINER
WILDER, CYNTHIA B

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
09/19/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket@bozpat.com

PARSE EX.1007.135

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

Doc Code:
TRACK1.GRANT

Decision Granting Request for Prioritized Examination (Track I or After RCE)	Application No.: 16/043,874

1. THE REQUEST FILED July 24, 2018 IS **GRANTED.**

The above-identified application has met the requirements for prioritized examination
- A. ☒ for an original nonprovisional application (Track I).
- B. ☐ for an application undergoing continued examination (RCE).

2. **The above-identified application will undergo prioritized examination.** The application will be accorded special status throughout its entire course of prosecution until one of the following occurs:

   A. filing a **petition for extension of time** to extend the time period for filing a reply;

   B. filing an **amendment to amend the application to contain more than four independent claims, more than thirty total claims**, or a multiple dependent claim;

   C. filing a **request for continued examination**;

   D. filing a notice of appeal;

   E. filing a request for suspension of action;

   F. mailing of a notice of allowance;

   G. mailing of a final Office action;

   H. completion of examination as defined in 37 CFR 41.102; or

   I. abandonment of the application.

Telephone inquiries with regard to this decision should be directed to Irvin Dingle at (571)272-3210, Office of Petitions.

Irvin Dingle
/Irvin Dingle/
[*Signature*]

Lead Paralegal Specialist
(Title)

PARSE EX.1007.136

Office of Petitions: Routing Sheet



Application No. 16043874

This application is being forwarded to your office for further processing. A decision has been rendered on a petition filed in this application, as indicated below. For details of this decision, please see the document PET.OP.DEC filed on the same date as this document.

- [X] **GRANTED**
- [ ] **DISMISSED**
- [ ] **DENIED**

Office of Petitions: Decision Count Sheet

Mailing Month: 9

Application No.: 16043874

16043874

For US serial numbers: enter number only, no slashes or commas. Ex: 10123456
For PCT: enter "51+single digit of year of filing+last 5 numbers", Ex. for PCT/US05/12345, enter 51512345

Deciding Official: Dingle, Irvin

Count (1) - Palm Credit 16043874

Decision: GRANT

FINANCE WORK NEEDED
☐ Select Check Box for YES

GRANT

Decision Type: 643 - Track One request

643

Notes:

Count (2)

Decision: n/a

FINANCE WORK NEEDED
☐ Select Check Box for YES

Decision Type: NONE

Notes:

Count (3)

Decision: n/a

FINANCE WORK NEEDED
☐ Select Check Box for YES

Decision Type: NONE

Notes:

Initials of Approving Official (if required)

If more than 3 decisions, attach 2nd count sheet & mark this box ☐

Printed on: *9/17/2018*

Office of Petitions Internal Document - Ver. 5.0

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

NOTICE OF ALLOWANCE AND FEE(S) DUE

77974 7590 10/02/2018

Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

EXAMINER
WILDER, CYNTHIA B

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 10/02/2018

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON	2837

TITLE OF INVENTION: TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$500	$0.00	$0.00	$500	01/02/2019

THE APPLICATION IDENTIFIED ABOVE HAS BEEN EXAMINED AND IS ALLOWED FOR ISSUANCE AS A PATENT. PROSECUTION ON THE MERITS IS CLOSED. THIS NOTICE OF ALLOWANCE IS NOT A GRANT OF PATENT RIGHTS. THIS APPLICATION IS SUBJECT TO WITHDRAWAL FROM ISSUE AT THE INITIATIVE OF THE OFFICE OR UPON PETITION BY THE APPLICANT. SEE 37 CFR 1.313 AND MPEP 1308.

THE ISSUE FEE AND PUBLICATION FEE (IF REQUIRED) MUST BE PAID WITHIN THREE MONTHS FROM THE MAILING DATE OF THIS NOTICE OR THIS APPLICATION SHALL BE REGARDED AS ABANDONED. THIS STATUTORY PERIOD CANNOT BE EXTENDED. SEE 35 U.S.C. 151. THE ISSUE FEE DUE INDICATED ABOVE DOES NOT REFLECT A CREDIT FOR ANY PREVIOUSLY PAID ISSUE FEE IN THIS APPLICATION. IF AN ISSUE FEE HAS PREVIOUSLY BEEN PAID IN THIS APPLICATION (AS SHOWN ABOVE), THE RETURN OF PART B OF THIS FORM WILL BE CONSIDERED A REQUEST TO REAPPLY THE PREVIOUSLY PAID ISSUE FEE TOWARD THE ISSUE FEE NOW DUE.

HOW TO REPLY TO THIS NOTICE:

I. Review the ENTITY STATUS shown above. If the ENTITY STATUS is shown as SMALL or MICRO, verify whether entitlement to that entity status still applies.

If the ENTITY STATUS is the same as shown above, pay the TOTAL FEE(S) DUE shown above.

If the ENTITY STATUS is changed from that shown above, on PART B - FEE(S) TRANSMITTAL, complete section number 5 titled "Change in Entity Status (from status indicated above)".

For purposes of this notice, small entity fees are 1/2 the amount of undiscounted fees, and micro entity fees are 1/2 the amount of small entity fees.

II. PART B - FEE(S) TRANSMITTAL, or its equivalent, must be completed and returned to the United States Patent and Trademark Office (USPTO) with your ISSUE FEE and PUBLICATION FEE (if required). If you are charging the fee(s) to your deposit account, section "4b" of Part B - Fee(s) Transmittal should be completed and an extra copy of the form should be submitted. If an equivalent of Part B is filed, a request to reapply a previously paid issue fee must be clearly made, and delays in processing may occur due to the difficulty in recognizing the paper as an equivalent of Part B.

III. All communications regarding this application must give the application number. Please direct all communications prior to issuance to Mail Stop ISSUE FEE unless advised to the contrary.

IMPORTANT REMINDER: Utility patents issuing on applications filed on or after Dec. 12, 1980 may require payment of maintenance fees. It is patentee's responsibility to ensure timely payment of maintenance fees when due.

PARSE EX.1007.139

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail **Mail Stop ISSUE FEE**
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

77974 7590 10/02/2018

Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission

I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON	2837

TITLE OF INVENTION: TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$500	$0.00	$0.00	$500	01/02/2019

EXAMINER	ART UNIT	CLASS-SUBCLASS
WILDER, CYNTHIA B	1637	435-006000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

❑ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

❑ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 _____
2 _____
3 _____

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE (B) RESIDENCE: (CITY and STATE OR COUNTRY)

Please check the appropriate assignee category or categories (will not be printed on the patent) : ❑ Individual ❑ Corporation or other private group entity ❑ Government

4a. The following fee(s) are submitted:

❑ Issue Fee

❑ Publication Fee (No small entity discount permitted)

❑ Advance Order - # of Copies _____

4b. Payment of Fee(s): **(Please first reapply any previously paid issue fee shown above)**

❑ A check is enclosed.

❑ Payment by credit card. Form PTO-2038 is attached.

❑ The director is hereby authorized to charge the required fee(s), any deficiency, or credits any overpayment, to Deposit Account Number _____ (enclose an extra copy of this form).

5. **Change in Entity Status** (from status indicated above)

❑ Applicant certifying micro entity status. See 37 CFR 1.29

❑ Applicant asserting small entity status. See 37 CFR 1.27

❑ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature _____ Date _____

Typed or printed name _____ Registration No. _____

PTOL-85 Part B (10-13) Approved for use through 10/31/2013. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PARSE EX.1007.140

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON	2837

77974 7590 10/02/2018

Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

EXAMINER	
WILDER, CYNTHIA B	

ART UNIT	PAPER NUMBER
1637	

DATE MAILED: 10/02/2018

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(Applications filed on or after May 29, 2000)

The Office has discontinued providing a Patent Term Adjustment (PTA) calculation with the Notice of Allowance.

Section 1(h)(2) of the AIA Technical Corrections Act amended 35 U.S.C. 154(b)(3)(B)(i) to eliminate the requirement that the Office provide a patent term adjustment determination with the notice of allowance. See Revisions to Patent Term Adjustment, 78 Fed. Reg. 19416, 19417 (Apr. 1, 2013). Therefore, the Office is no longer providing an initial patent term adjustment determination with the notice of allowance. The Office will continue to provide a patent term adjustment determination with the Issue Notification Letter that is mailed to applicant approximately three weeks prior to the issue date of the patent, and will include the patent term adjustment on the patent. Any request for reconsideration of the patent term adjustment determination (or reinstatement of patent term adjustment) should follow the process outlined in 37 CFR 1.705.

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Customer Service Center of the Office of Patent Publication at 1-(888)-786-0101 or (571)-272-4200.

PARSE EX.1007.141

OMB Clearance and PRA Burden Statement for PTOL-85 Part B

The Paperwork Reduction Act (PRA) of 1995 requires Federal agencies to obtain Office of Management and Budget approval before requesting most types of information from the public. When OMB approves an agency request to collect information from the public, OMB (i) provides a valid OMB Control Number and expiration date for the agency to display on the instrument that will be used to collect the information and (ii) requires the agency to inform the public about the OMB Control Number's legal significance in accordance with 5 CFR 1320.5(b).

The information collected by PTOL-85 Part B is required by 37 CFR 1.311. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, Virginia 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450. Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Notice of Allowability	Application No. 16/043,874	Applicant(s) Giresi et al.	
	Examiner CYNTHIA B WILDER	Art Unit 1637	AIA Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 8/2/2018.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 78-100 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☑ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☑ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☑ Certified copies of the priority documents have been received in Application No. 14784250 .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file areply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☑ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date 7/24/2018.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☑ Examiner's Amendment/Comment
6. ☑ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/CYNTHIA B WILDER/
Primary Examiner, Art Unit 1637

PARSE EX.1007.143

DETAILED ACTION

Notice of Pre-AIA or AIA Status

1. The present application, filed on or after March 16, 2013, is being examined under the first inventor to file provisions of the AIA. Applicant's preliminary amendment filed 7/24/2018 is acknowledged. Claims 1-77 have been canceled. Claims 78-100 have been added and are pending.

Information Disclosure Statement

2. The information disclosure statement (IDS) submitted on 7/24/2018 is acknowledged. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Priority

3. The instant invention is a continuation of 14/784250 filed 10/13/2015, now US Patent 10/059989 which is a 371 of PCT/US14/38825 filed 05/20/2014 which clams benefit to US provisional application 61/826728 filed May 23, 2013.

Drawings

4. The drawings were received on 10/13/2015. These drawings are found acceptable by the examiner.

Reasons for Allowance

5. The following is an examiner's statement of reasons for allowance: The claims are deemed to be in condition for allowance because no prior art was found teaching or suggesting a method for analyzing a biological sample comprising 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; and 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 derivative thereof. Rather the art supports techniques which involves the use of insertional complex that comprises an antibody specific to a protein that is part of the chromatin, such as via chromatin immunoprecipitation) as a main tool for investigating transposition in epigenetic studies. See for example Park et al., Nature 10, 669-680, 2009; see entire reference especially Figure 1 (citation made of record on IDS filed 7/24/2018) and Jelinek et al (US 20160115474, effective filing date May 2013) both of which requires transpososome-antibody complex associated with chromatin for subsequent analysis.

With regards to analysis under double patenting, the clams 78-100 comprise of embodiments sufficiently distinct in scope such that the analysis under double patenting does not apply. Accordingly, the instant invention is deemed novel and non-obvious over the closest prior art.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

6. Any inquiry concerning this communication or earlier communications from the examiner should be directed to CYNTHIA B WILDER whose telephone number is (571)272-0791. The examiner can normally be reached on Flexible.

Examiner interviews are available via telephone, in-person, and video conferencing using a USPTO supplied web-based collaboration tool. To schedule an interview, applicant is encouraged to use the USPTO Automated Interview Request (AIR) at http://www.uspto.gov/interviewpractice.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, GARY BENZION can be reached on 571-272-0782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/CYNTHIA B WILDER/
Primary Examiner, Art Unit 1637

Notice of References Cited	Application/Control No. 16/043,874	Applicant(s)/Patent Under Reexamination	
	Examiner CYNTHIA B WILDER	Art Unit 1637	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-20160115474-A1	04-2016	Jelinek; Mary Anne	C07H21/00	506/4
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

PARSE EX.1007.147

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	**Art Unit**
	CYNTHIA B WILDER	1637

CPC							
Symbol						**Type**	**Version**
C12Q	/	1	/	6874		F	2013-01-01
C12Q	/	1	/	6869		I	2013-01-01

CPC Combination Sets								
Symbol					**Type**	**Set**	**Ranking**	**Version**
	/		/					

NONE		**Total Claims Allowed:**	
(Assistant Examiner)	(Date)	23	
/CYNTHIA B WILDER/ Primary Examiner, Art Unit 1637	21 September 2018	O.G. Print Claim(s)	O.G. Print Figure
(Primary Examiner)	(Date)	1	NONE

U.S. Patent and Trademark Office

Part of Paper No.: 20180921

PARSE EX.1007.148

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	**Art Unit**
	CYNTHIA B WILDER	1637

INTERNATIONAL CLASSIFICATION					
CLAIMED					
C12Q	/	1	/	68	
NON-CLAIMED					
	/		/		

US ORIGINAL CLASSIFICATION	
CLASS	**SUBCLASS**
435	6.1

CROSS REFERENCES(S)						
CLASS	**SUBCLASS (ONE SUBCLASS PER BLOCK)**					

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 23	
/CYNTHIA B WILDER/ Primary Examiner, Art Unit 1637 (Primary Examiner)	21 September 2018 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure NONE

PARSE EX.1007.149

Issue Classification	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	**Art Unit**
	CYNTHIA B WILDER	1637

☐ Claims renumbered in the same order as presented by applicant ☐ CPA ☐ T.D. ☐ R.1.47

CLAIMS

Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original	Final	Original
-	1	-	14	-	27	-	40	-	53	-	66	2	79	15	92
-	2	-	15	-	28	-	41	-	54	-	67	3	80	16	93
-	3	-	16	-	29	-	42	-	55	-	68	4	81	17	94
-	4	-	17	-	30	-	43	-	56	-	69	5	82	18	95
-	5	-	18	-	31	-	44	-	57	-	70	6	83	19	96
-	6	-	19	-	32	-	45	-	58	-	71	7	84	20	97
-	7	-	20	-	33	-	46	-	59	-	72	8	85	21	98
-	8	-	21	-	34	-	47	-	60	-	73	9	86	22	99
-	9	-	22	-	35	-	48	-	61	-	74	10	87	23	100
-	10	-	23	-	36	-	49	-	62	-	75	11	88		
-	11	-	24	-	37	-	50	-	63	-	76	12	89		
-	12	-	25	-	38	-	51	-	64	-	77	13	90		
-	13	-	26	-	39	-	52	-	65	1	78	14	91		

NONE (Assistant Examiner)	(Date)	**Total Claims Allowed:** 23	
/CYNTHIA B WILDER/ Primary Examiner, Art Unit 1637 (Primary Examiner)	21 September 2018 (Date)	O.G. Print Claim(s) 1	O.G. Print Figure NONE

U.S. Patent and Trademark Office

Part of Paper No.: 20180921

PARSE EX.1007.150

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

CPC - Searched*		
Symbol	Date	Examiner
C12Q1/6874	09/21/2018	CW
C12Q1/6869	09/21/2018	CW

CPC Combination Sets - Searched*		
Symbol	Date	Examiner

US Classification - Searched*			
Class	Subclass	Date	Examiner
435	6.1	09/21/2018	CW

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
Inventor search on PALM, text search on west, review of related application and all documents filed therein, review of IDS filed in instant application and copending appl. Review of training material regarding double patenting	09/21/2018	CW

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner
C12Q1/6874		09/21/2018	CW
C12Q1/6869		09/21/2018	CW

/CYNTHIA B WILDER/ Primary Examiner, Art Unit 1637	

PARSE EX.1007.151

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

✓	Rejected	-	Cancelled	N	Non-Elected	A	Appeal
=	Allowed	÷	Restricted	I	Interference	O	Objected

CLAIMS										
☐ Claims renumbered in the same order as presented by applicant			☐ CPA		☐ T.D.		☐ R.1.47			
CLAIM		DATE								
Final	Original	09/21/2018								
-	1	-								
-	2	-								
-	3	-								
-	4	-								
-	5	-								
-	6	-								
-	7	-								
-	8	-								
-	9	-								
-	10	-								
-	11	-								
-	12	-								
-	13	-								
-	14	-								
-	15	-								
-	16	-								
-	17	-								
-	18	-								
-	19	-								
-	20	-								
-	21	-								
-	22	-								
-	23	-								
-	24	-								
-	25	-								
-	26	-								
-	27	-								
-	28	-								
-	29	-								
-	30	-								
-	31	-								
-	32	-								
-	33	-								
-	34	-								
-	35	-								
-	36	-								
-	37	-								
-	38	-								
-	39	-								
-	40	-								
-	41	-								
-	42	-								

PARSE EX.1007.152

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	Art Unit
	CYNTHIA B WILDER	1637

CLAIM		DATE								
Final	Original	09/21/2018								
-	43	-								
-	44	-								
-	45	-								
-	46	-								
-	47	-								
-	48	-								
-	49	-								
-	50	-								
-	51	-								
-	52	-								
-	53	-								
-	54	-								
-	55	-								
-	56	-								
-	57	-								
-	58	-								
-	59	-								
-	60	-								
-	61	-								
-	62	-								
-	63	-								
-	64	-								
-	65	-								
-	66	-								
-	67	-								
-	68	-								
-	69	-								
-	70	-								
-	71	-								
-	72	-								
-	73	-								
-	74	-								
-	75	-								
-	76	-								
-	77	-								
1	78	=								
2	79	=								
3	80	=								
4	81	=								
5	82	=								
6	83	=								
7	84	=								
8	85	=								
9	86	=								
10	87	=								
11	88	=								
12	89	=								
13	90	=								
14	91	=								
15	92	=								
16	93	=								
17	94	=								

PARSE EX.1007.153

Index of Claims	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/043,874	Giresi et al.
	Examiner	**Art Unit**
	CYNTHIA B WILDER	1637

CLAIM		DATE								
Final	Original	09/21/2018								
18	95	=								
19	96	=								
20	97	=								
21	98	=								
22	99	=								
23	100	=								

PARSE EX.1007.154

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (02-18)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	
First Named Inventor	Paul Giresi
Art Unit	
Examiner Name	
Attorney Docket Number	STAN-1111CON

U.S.PATENTS

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	5137829		1992-08-11	Nag et al.	
	2	5965443		1999-10-12	Reznikoff et al.	
	3	6159717		2000-12-12	Savakis et al.	
	4	6207384		2001-03-27	Mekalanos et al.	
	5	6258571		2001-07-10	Chumakov et al.	
	6	6291243		2001-09-18	Fogarty et al.	
	7	6294385		2001-09-25	Goryshin et al.	
	8	6593113		2003-07-15	Tenkanen et al.	

/CYNTHIA B WILDER/ (09/21/2018) 09/21/2018

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

PARSE EX.1007.155

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	9	6613752		2003-09-02	Kay et al.	
	10	7608451		2009-10-27	Cooper et al.	
	11	7745218		2010-06-29	Kim et al.	
	12	7608434		2009-10-27	Reznikoff et al.	
	13	7138267		2006-11-21	Jendrisak et al.	
	14	8124404		2012-02-28	Alphey	
	15	8420386		2013-04-16	Ivics et al.	
	16	8598328		2013-12-03	Koga et al.	
	17	8829171		2014-09-09	Steemers et al.	
	18	8927218		2015-01-06	Forsyth	
	19	9005935		2015-04-14	Belyaev	

/CYNTHIA B WILDER/ (09/21/2018) 09/21/2018

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	
First Named Inventor	Paul Giresi
Art Unit	
Examiner Name	
Attorney Docket Number	STAN-1111CON

	20	9074251		2015-07-07	Steemers et al.	
	21	9080211		2015-07-14	Grunenwald et al.	
	22	9150916		2015-10-06	Christen et al.	
	23	9175295		2015-11-03	Kaminaka et al.	
	24	9238671		2016-01-19	Goryshin et al.	
	25	9328382		2016-05-03	Drmanac et al.	
	26	9574226		2017-02-21	Gormley et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

U.S.PATENT APPLICATION PUBLICATIONS

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20030032141		2003-02-13	Nguyen	
	2	20040101880		2004-05-27	Rozwadowski et al.	

/CYNTHIA B WILDER/ (09/21/2018) 09/21/2018

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT
(Not for submission under 37 CFR 1.99)

Application Number	
Filing Date	
First Named Inventor	Paul Giresi
Art Unit	
Examiner Name	
Attorney Docket Number	STAN-1111CON

	3	20100120098		2010-05-13	GRUNENWALD, et al.	
	4	20120297493		2012-11-22	Cooper et al.	
	5	20130203605		2013-08-08	Shendure et al.	
	6	20150111788		2015-04-23	Fernandez et al.	
	7	20150337298		2015-11-26	Xi et al.	
	8	20150368638		2015-12-24	Steemers et al.	
	9	20150376608		2015-12-31	Kaper et al.	
	10	20160060691		2016-03-03	Giresi et al.	
	11	20160115474		2016-04-28	Jelinek et al.	
	12	20160122753		2016-05-05	Mikkelsen et al.	
	13	20160160235		2016-06-09	Solodushko et al.	

/CYNTHIA B WILDER/ (09/21/2018) 09/21/2018

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

PARSE EX.1007.158

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	14	20160177359		2016-06-23	Ukanis et al.	
	15	20160208323		2016-07-21	Bernstein et al.	
	16	20160376663		2016-12-29	Brown	

If you wish to add additional U.S. Published Application citation information please click the Add button.

FOREIGN PATENT DOCUMENTS

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	WO 2007/093819	WO		2007-08-23	ISIS INNOVATION LIMITED		☐
	2	WO 2009/147386	WO		2009-12-10	OXFORD BIODYNAMICS LIMITED		☐
	3	WO 2010/048605	WO		2010-04-29	EPICENTRE TECHNOLOGIES CORPORATION		☐
	4	WO 2014/108810	WO		2014-07-17	ILLUMINA CAMBRIDGE LTD		☐
	5	WO 2016/191618	WO		2016-12-01	ZHENG JIANBIAO et al.		☐
	6	WO 2016/207647	WO		2016-12-29	OXFORD BIODYNAMICS LIMITED		☐

/CYNTHIA B WILDER/ (09/21/2018) 09/21/2018

EFS Web 2.1.18

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	7	WO 2016/207653	WO		2016-12-29	OXFORD BIODYNAMICS LIMITED		☐
	8	WO 2016/207661	WO		2016-12-29	OXFORD BIODYNAMICS LIMITED		☐
	9	WO 2017/015075	WO		2017-01-26	HARVARD COLLEGE		☐
	10	WO 2017/025594	WO		2017-02-16	CHEM FORSCHUNGSZENTRUM FÜR MOLEKULARE MEDIZIN GMBM		☐

If you wish to add additional Foreign Patent Document citation information please click the Add button

NON-PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	ADEY, et al., "Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing", Genome Research, 2012, 22 (6): 1139-1143.	☐
	2	ASON et al., "DNA Sequence Bias During Tn5 Transposition", J. Mol. Biol., 2004, 335: 1213-1255.	☐
	3	BJORNSSON et al., "Intra-individiual change over time in DNA methylation with familial clustering, JAMA, June 25, 2008, Vol. 299 No. 24, pages 2877-2883	☐
	4	BOYLE, et al., "High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells", Genome Res. 2011 Mar;21(3):456-64	☐
	5	BUENROSTRO, et al., "Tranposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position", Nature Methods, 2013, 10(12): 1213-1218	☐

/CYNTHIA B WILDER/ (09/21/2018) 09/21/2018

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	
	Filing Date	
	First Named Inventor	Paul Giresi
	Art Unit	
	Examiner Name	
	Attorney Docket Number	STAN-1111CON

	6	GANGADHARAN et al., "DNA transposon Hermes insert into DNA in nucleosome-free regions in vivo, Proc nat Ad Sci, December 21, 2010, Vol. 107, No. 51, pages 1966-1972.	☐
	7	GREEN et al., "Insertion site preference of Mu, Tn5, and Tn7 transposons", Mobile DNA, 2012, 3:3, 6 pages.	☐
	8	Parks et al., "ChIP-Seq: advantages and challenges of a maturing technology", Nature, vol. 10, pages 669-680, October (Year: 2009)	☐
	9	Haring et al., "Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization", Plant Methods, 3 (11): 1-25 (Year: 2007)	☐
	10	SIMON, et al., "Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA", Nature Protocols, 2012, 7(2): 256-267.	☐
	11	SONG, et al., "DNase-seq: A High-Resolution Technique for Mapping Active Gene Regulatory Elements across the Genome from Mammalian Cells", Cold Spring Harbor Laboratory Press, 2010, 2010(2), doi:10.1101/pdb.prot5384	☐
	12	ZENTNER et al., "Surveying the epigenomic landscape, one base at a time", Genome Biology, 2012, 13:250, 8 pages.	☐

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature	/CYNTHIA B WILDER/ (09/21/2018)	Date Considered	09/21/2018

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /C.B.W/

PARSE EX.1007.161

WEST Search History for Application 16043874

Creation Date: 2018092119:13

Prior Art Searches

Query	DB	Hits	Op.	Plur.	Thes.	Date
20160115474 or 20150111788	PGPB, USPT, USOC, EPAB, DWPI	6	ADJ	YES		09-21-2018
chromatin same tagged fragment$	PGPB, USPT, USOC, EPAB, DWPI	11	ADJ	YES		09-21-2018
chromatin same tagged fragment$	PGPB, USPT, USOC, EPAB, DWPI	11	ADJ	YES		09-21-2018
(chromatin same tagged fragment$) and transposase	PGPB, USPT, USOC, EPAB, DWPI	6	ADJ	YES		09-21-2018
20160060691	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		09-21-2018
(chromatin same transposase) or (chromatin same HIC integrase)	PGPB, USPT, USOC, EPAB, DWPI	265	ADJ	YES		09-21-2018
(20160060691) and (fragment$ same tag$)	PGPB, USPT, USOC, EPAB, DWPI	3	ADJ	YES		09-21-2018
		112	ADJ	YES		09-21-2018

PARSE EX.1007.162

((chromatin same transposase) or (chromatin same HIC integrase)) and (epigenetic)	PGPB, USPT, USOC, EPAB, DWPI					
insertional enzyme complex or transposase	PGPB, USPT, USOC, EPAB, DWPI	7117	ADJ	YES		09-21-2018
(insertional enzyme complex or transposase) same (tag$ same (nucleic acid or DNA))	PGPB, USPT, USOC, EPAB, DWPI	406	ADJ	YES		09-21-2018
(insertional enzyme complex or transposase same (tag$ same (nucleic acid or DNA))) same chromatin	PGPB, USPT, USOC, EPAB, DWPI	17	ADJ	YES		09-21-2018
(insertional enzyme complex or transposase same (tag$ same (nucleic acid or DNA)) same chromatin) and (cell or nucleus)	PGPB, USPT, USOC, EPAB, DWPI	17	ADJ	YES		09-21-2018
WO 2014190214	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		09-21-2018
WO 2014190214	PGPB, USPT, USOC, EPAB, DWPI	1	ADJ	YES		09-21-2018
61826481	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		09-21-2018
16043874	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		09-21-2018

PARSE EX.1007.163

16/043874	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		09-21-2018
16/043874	PGPB, USPT, USOC, EPAB, DWPI	0	ADJ	YES		09-21-2018
14/784250	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		09-21-2018
(14/784250) and ((C12Q1/6869 | C12Q2521/301 | C12Q2537/164 | C12Q1/6874).CPC.)	PGPB, USPT, USOC, EPAB, DWPI	2	ADJ	YES		09-21-2018

PARSE EX.1007.164

Bibliographic Data

Application No: 16/043,874

Foreign Priority claimed: (•) Yes ( ) No

35 USC 119 (a-d) conditions met: [✓] Yes [ ] No [ ] Met After Allowance

Verified and Acknowledged: /CYNTHIA B WILDER/ (Examiner's Signature) /CBW/ (Initials)

Title: TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

FILING or 371(c) DATE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
07/24/2018	435	1637	STAN-1111CON
RULE			

APPLICANTS

The Board of Trustees of the Leland Stanford Junior University, Stanford, CA, UNITED STATES

INVENTORS

Paul Giresi Palo Alto, CA, UNITED STATES

Jason D. Buenrostro Redwood City, CA, UNITED STATES

Howard Y. Chang Stanford, CA, UNITED STATES

William J. Greenleaf Menlo Park, CA, UNITED STATES

CONTINUING DATA

This application is a CON of 14784250 10/13/2015 PAT 10059989

14784250 is a 371 of PCT/US2014/038825 05/20/2014

PCT/US2014/038825 has PRO of 61826728 05/23/2013

FOREIGN APPLICATIONS

IF REQUIRED, FOREIGN LICENSE GRANTED**

09/06/2018

**** SMALL ENTITY ****

STATE OR COUNTRY

UNITED STATES

ADDRESS

Stanford University Office of Technology Licensing

Bozicevic, Field & Francis LLP

201 REDWOOD SHORES PARKWAY

SUITE 200

REDWOOD CITY, CA 94065

UNITED STATES

FILING FEE RECEIVED

$3,005

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (03-15)
Approved for use through 07/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16043874
Filing Date	2018-07-24
First Named Inventor	Giresi, Paul
Art Unit	1637
Examiner Name	WILDER, CYNTHIA B.
Attorney Docket Number	STAN-1111CON

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20140274740		2014-09-18	Srinivasan et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	WO2012106546	WO	A2	2012-08-09	SHENDURE et al.		
	2	CN102292455	CN	A	2011-12-21	Bio-Rad Laboratories, Inc.		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

PARSE EX.1007.167

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	16043874
	Filing Date	2018-07-24
	First Named Inventor	Giresi, Paul
	Art Unit	1637
	Examiner Name	WILDER, CYNTHIA B.
	Attorney Docket Number	STAN-1111CON

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	Epicentre., "EZ-Tn5TM Custom Transposome Construction Kits", www.epicentre.com, pages 1-17, 2012.	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16043874
Filing Date	2018-07-24
First Named Inventor	Giresi, Paul
Art Unit	1637
Examiner Name	WILDER, CYNTHIA B.
Attorney Docket Number	STAN-1111CON

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/James S. Keddie/	Date (YYYY-MM-DD)	2018-10-08
Name/Print	James S. Keddie	Registration Number	48920

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

PARSE EX.1007.169

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

PARSE EX.1007.170

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2012 (09.08.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/106546 A2

(51) International Patent Classification:
C40B 30/04 (2006.01)

(21) International Application Number:
PCT/US2012/023679

(22) International Filing Date:
2 February 2012 (02.02.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/438,935 2 February 2011 (02.02.2011) US
61/473,083 7 April 2011 (07.04.2011) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION** [US/US]; Suite 500, 4311 11th Avenue NE, Seattle, Washington 98105-4608 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SHENDURE, Jay, Ashok** [US/US]; 1701 NE 80th Street, Seattle, Washington 98115 (US). **SCHWARTZ, Jerrod, Joseph** [CA/US]; Apt. 5, 3635 Francis Ave. N, Seattle, Washington 98103 (US). **ADEY, Andrew, Colin** [US/US]; Apt. 61, 920 N. 34th Street, Seattle, Washington 98103 (US). **LEE, Cho li** [US/US]; Apt. 508, 1305 NE 43rd Street, Seattle, Washington 98105 (US). **HIATT, Joseph, Brian** [US/US]; Apt. 202, 124 Bellevue Avenue, Seattle, Washington 98102 (US). **KITZMAN, Jacob, Otto** [US/US]; Apt. 205, 3815 Woodland Park Ave. N, Seattle, Washington 98103 (US). **KUMAR, Akash** [US/US]; Apt. D, 1602 Ravenna Blvd. NE, Seattle, Washington 98105 (US).

(74) Agent: **DUEPPEN, Lara, J.**; Perkins Coie LLP, P.O. Box 1208, Seattle, Washington 98111-1208 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2012/106546 A2

(54) **Title**: MASSIVELY PARALLEL CONTINGUITY MAPPING

(57) **Abstract**: Contiguity information is important to achieving high-quality de *novo* assembly of mammalian genomes and the haplotype-resolved resequencing of human genomes. The methods described herein pursue cost-effective, massively parallel capture of contiguity information at different scales.

PARSE EX.1007.171

MASSIVELY PARALLEL CONTIGUITY MAPPING

PRIORITY CLAIM

[0001] This application claims priority to U.S. Provisional Patent Application No. 61/438,935, filed February 2, 2011, and U.S. Provisional Patent Application No. 61/473,083, filed April 7, 2011, the subject matter of both of which is hereby incorporated by reference as if fully set forth herein

STATEMENT OF GOVERNMENT SUPPORT

[0002] This invention was made with government support under Grant Numbers 3U54AI057141-06S1880509 and 1R01HG006283-01, awarded by National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] Over the last several years, massively parallel sequencing platforms have reduced the cost-per-base of DNA sequencing by several orders of magnitude (Shendure & Ji 2008). Of the "next-generation" technologies that are commercially available, nearly all rely on iterative cycles of biochemistry and imaging of dense arrays of sequencing features to generate relatively short reads, i.e. "cyclic-array" methods (Shendure et al. 2005; Margulies et al. 2005; Drmanac et al. 2009; Braslavsky et al. 2003; Bentley et al. 2008). The broad dissemination of these platforms represents the culmination of decades of effort to develop practical alternatives to electrophoretic sequencing (Shendure et al. 2004).

[0004] In the context of this success, many developing technologies have the potential to improve the technical capability of what is already feasible today. Such improvements may be accomplished by further development of cyclic array methods, or through the maturation of other promising strategies such as nanopore sequencing (Branton et al. 2008), real-time observation of DNA synthesis (Eid et al. 2009) and sequencing by electron microscopy. Massively parallel sequencing platforms have also given rise to several types of sequencing applications, including resequencing, de novo assembly, exome sequencing (Ng et al. 2009), RNA-Seq (Mortazavi et al. 2008), ChIP-

PARSE EX.1007.172

Seq (Johnson et al. 2007), and genome-wide chromatin interaction mapping (Lieberman-Aiden et al. 2009; Duan et al. 2010).

[0005] Although DNA sequencing technology platforms have improved at a rapid pace, the cost of DNA sequencing remains prohibitive for some goals. Therefore, it is desired to produce methods related to DNA sequencing technology that not only improve the application of existing and developing technology, but also reduce the cost.

SUMMARY

[0006] Short-read sequencing is limited with respect to resequencing of segmental duplications and structurally complex regions of the genome, the resolution of haplotype information, and the *de novo* assembly of mammalian-sized genomes. Moreover, further reductions in the cost-per-base of sequencing will do little to address these limitations. Even as new approaches to DNA sequencing mature and surpass current technology, technologies may continue to be limited in terms of the contiguity information that they generate. Therefore, low-cost methods for obtaining contiguity information at different scales are provided herein.

[0007] In some embodiments, methods for capturing contiguity information comprising are provided herein. Such methods may include treating a target DNA sequence with a transposase resulting in one or more fragmentation or insertion events; adding or inserting one or more recognition sequences to the target DNA sequence (i) during the transposase treatment of (ii) during a subsequent amplification; sequencing the treated DNA; and capturing contiguity information by identifying target DNA sequences or recognition sequences having a shared property.

[0008] In one embodiment, the one or more fragmentation or insertion events results in generation of a library of target nucleic acid molecules derived from the target DNA. In such methods, the one or more recognition sequences are one or more barcodes that are symmetrically tagged to sequences adjacent to each fragmentation or insertion event and the shared property of the one or more barcodes is an identical or complimentary barcode sequence.

PARSE EX.1007.173

[0009] In another embodiment, the target DNA sequence comprises a set of target DNA fragments. Such an embodiment may further include compartmentalizing the target DNA fragments with emulsions or dilutions, generating two or more compartments of target DNA fragments prior to or after treating with the transposase. In this embodiment, the one or more recognition sequences are one or more compartment-specific barcodes, each of which corresponds to the one or more compartments generated in the compartmentalizing step and the shared property of the one or more primer sequences is an identical compartment-specific barcode.

[0010] In another embodiment, the one or more recognition sequences is one or more adaptor sequences that modify the ends of the target DNA sequence or insert within the target DNA sequence. In such an embodiment, the one or more adaptor sequences may be complementary to one or more surface-bound primers. In some aspects, the transposase is bound to a nucleic acid that is complementary to a second surface-bound primer. Further, such a method may include hybridizing the one or more adaptor sequences to the one or more surface bound primers. In some embodiments, the shared property is a constrained physical location, which may be indicated by an x,y coordinate on a flowcell, and the transposase is bound to a surface-bound recognition sequence to form a surface-bound transposase complex. In some embodiments, treating the target DNA sequence comprises exposing a plurality of surface-bound transposase complexes to the target DNA sequence.

[0011] In some embodiments, methods of bisulfite sequencing are provided. Such methods may include performing *in vitro* transposition into target DNA molecules with transposase complexes, each transposase complex comprising a double stranded DNA transposase recognition sequence and a single stranded DNA adaptor overhang having methylated cytosine (C) residues; subjecting transposed target DNA molecules to bisulfite treatment; performing nucleic acid amplification; and sequencing the resulting nucleic acid library.

[0012] In other embodiments, methods for inferring chromosome conformation are provided. Such methods may include cross-linking DNA within cells; isolating cross-linked DNA from cells; fragmenting the cross-linked DNA; end-modifying fragmented,

PARSE EX.1007.174

cross-linked DNA molecules with an adaptor that is complementary to or that corresponds to a first surface-bound primer; e) hybridizing ends of the fragmented, end-modified target DNA molecules to the first surface-bound primer; f) performing transposition with non-surface-bound transposase complexes, each non-surface-bound transposase complex comprising a DNA transposase and one or more sequences corresponding to a second surface-bound primer; g) performing cluster amplification to produce clusters of clonally derived nucleic acids; h) sequencing clusters of clonally derived nucleic acids; and i) determining physical interactions between chromosomal positions by paring neighboring clusters together.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Figure 1 illustrates high density, random, *in vitro* transposition of discontinuous oligonucleotides enables the high efficiency conversion of genomic DNA into adaptor-flanked, shotgun fragments. Light grey area (1) = transposase; dark grey bars (2) = mosaic ends (ME); yellow & red (3a, 3b) = asymmetrical 5' overhangs; blue (4) = genomic DNA).

[0014] Figure 2 is a histogram of fold-coverage with whole genome sequencing (x-axis = fold-coverage; y-axis = % of genome) of libraries from a male human generated by standard methods ('sonication') versus the transposome method ('transposase'), with autosomes ('Aut.') and sex chromosomes ('Sex') plotted separately.

[0015] Figure 3 is a histogram of fragment sizes (x-axis = base-pairs; y-axis = counts) resulting from high-density, *in vitro* fragmentation with a synthetic, discontinuous transposon. The inset shows a model for transposome occupancy consistent with a steric hindrance model for the sharp drop at ~35 bp.

[0016] Figure 4 shows *in vitro*, high-density insertion of transposomes with degenerate, single-stranded "bubbles" (A/B) to genomic DNA (dark gray, (1)) is followed by whole genome amplification (WGA) to resolve each strand of the degenerate stretch (to A/A or B/B). Nicking (at medium gray sites, (2)) and strand displacing polymerization completes fragmentation, but also leaves junctions symmetrically tagged with the same barcode (A/A (shown) or B/B).

PARSE EX.1007.175

[0017] Figure 5 shows independent reads derived from limited sequencing of transposase-based shotgun libraries show enrichment for mapping at 9 bp intervals. This phenomenon is much more pronounced with ultra-low input (10 pg, arrow) relative to low input (50 ng, no arrow), reflecting greater sampling of a lower number of discrete fragmentation events.

[0018] Figure 6 is a schematic diagram, based on examples observed in real data, showing that read-pairs mapping to adjacent locations with 9 bp overlaps are likely to have derived from adjacent fragmentation events. In complexity-limited data based on a library derived from an 'ultra-low-input' sample, chains of 4 to 6 locally derived read-pairs may be identified that collectively span ~1 Kb to ~2 Kb.

[0019] Figure 7 is a graph showing the expected N10, N50, N90 lengths of the total span (y-axis) of chains of read-pairs that are identified as resulting from a contiguous series of fragmentation events along the same genomic DNA molecule, as a function of the efficiency of identifying individual 'joins' (x-axis, percentage; note transition in scale at 99%).

[0020] Figure 8 is a schematic diagram showing that emulsion PCR of a template consisting of common regions ((1), (2)) that flank a degenerate region (A) generates clonally barcoded beads. The common 3' end of the bead-tethered strand (2) can itself serve as a primer in subsequent emulsion PCR reactions.

[0021] Figure 9 is a schematic diagram showing HMW genomic DNA molecules (blue) that are subjected to *in vitro* fragmentation with transposomes bearing adaptors ((1), (2)) that are linked by hybridization of complementary subsequences (brown). DNA densely interspersed with these linked adaptors is then emulsified via microfluidics and subjected to emulsion PCR with primers bearing droplet-specific barcodes (A). Sequence reads from the same HMW genomic DNA fragment may be associated with the same barcode in the final library.

[0022] Figure 10 is a schematic diagram showing emulsions that can be used to support the clonal, isothermal, multiple displacement amplification of HMW DNA (1). These are fused with droplets containing reagents for both transposome fragmentation

PARSE EX.1007.176

and emulsion PCR with primers containing droplet-specific barcodes (color scheme identical to Figures 8 & 9).

[0023] Figure 11 is a graph showing a comparison of experimentally phased assembly with population-based HapMap predictions by HapMap for the same individual for various LD values. In contrast with HapMap inferences, the experimentally phased haplotypes are derived by a method that is LD independent, such that discrepancies predominantly reflect errors in inference-based haplotypes.

[0024] Figure 12 illustrates the use of *in situ* transposition for facilitating methods related to optical sequencing. (A) Single templates are stretched out on a flowcell and fragmented to generate spatially separated clusters at a physical distance proportional to their genomic distance. (B) Randomly coiled DNA is fragmented at its ends to generate clusters that are spatially confined to the area beneath the coil. Reads from either end can be deconvolved by using two different sequencing primers.

[0025] Figure 13 illustrates representative images of a spatially separated "cluster pair" for raw images of a "cluster pair" over four cycles of sequencing (A); and raw integrated basecalling intensities of the two templates over the four cycles (B).

[0026] Figure 14 shows representative images of (A) 48.5 Kb lambda genomes that were stained with JOJO-1, tethered to a modified Illumina flowcell, and stretched with a 15V/cm electric field and (B) stretched DNA like that in (A) that was treated with transposomes for 5 minutes at 55°C and imaged again. Imaging was performed on an Illumina GA2x. Scale bars = 20 μm.

[0027] Figure 15 is a schematic diagram illustrating pretreatment of the library to insert flowcell compatible adapters, without fragmentation, allowing for multiple read pairs to be generated along the axis of the stretched molecule.

[0028] Figure 16 illustrates high-density insertion of synthetic transposons containing single-stranded bubbles into genomic DNA. Lane 1 = ladder (kb); Lane 2 = unfragmented genomic DNA; Lane 3 = post-insertion, post PCR material.

[0029] Figure 17 illustrates the construction of symmetrically tagged, 5'-to-5' linked transposon reagent.

PARSE EX.1007.177

[0030] Figure 18 shows species matching expected size (194 bp) of symmetrically tagged, 5'-5' adaptor (a) and size distribution of post-transposition, post-PCR fragment amplicons is consistent with ~100-200bp of genomic DNA and ~200bp of total adaptor/barcode (b).

[0031] Figure 19 illustrates transposition and polymerase extension in a single reaction volume with no intervening manipulations. Transposase drives fragmentation. Polymerase drives gap closure via nick translation and limited cycles of primer extension to append a barcode (A) bearing adaptor.

[0032] Figure 20 illustrates transposition and polymerase extension in a single reaction volume with no intervening manipulations yields products that can be recovered by PCR after column-cleanup. The primers used in the PCR correspond to sequences added during the extension step. Lane 1 = 100 bp ladder; Lane 2 = no genomic DNA (gDNA) control; Lane 3 = 50 gDNA input.

[0033] Figure 21 illustrates two methods to generate shotgun HMW genomic DNA fragments with appropriate adaptors and 3' ssDNA tails corresponding to flow-cell sequence.

[0034] Figure 22 shows coverage of E. coli genome with reads derived from in situ transposition method. X-axis = genomic coordinates. Y-axis = number of reads (10Kb bins).

[0035] Figure 23 illustrates a Y-adaptor approach for library preparation according to some embodiments.

[0036] Figure 24 illustrates the production of multiple displacing branching rolling circle amplification and polony (i.e., polymerase colony) formation according to some embodiments.

[0037] Figure 25 illustrates a method for direct sequencing of transposon bubbles containing flowcell primers according to some embodiments.

[0038] Figure 26 illustrates a method of transposon insertion using two of the same adaptors in reverse orientation to maintain the resulting "bubble" structure followed by emulsification and amplification according to some embodiments.

PARSE EX.1007.178

[0039] Figure 27 illustrates a transposon-modified fosmid library pool approach to sequencing by using unique barcodes or insertion sites within repetitive regions according to some embodiments.

[0040] Figure 28 illustrates a method used to generate clusters on flowcell: Any combination of the four arms could hybridize to the flowcell and generate a library. In this case, only two arms do.

[0041] Figure 29 illustrates a method that uses "infinipair" to identify interactions between transcription factor binding sites. Cells may be cross linked with formaldehyde subjected to ChIP to pull down DNA:protein complexes. Modified sequencing adaptors may be ligated onto the complexes and used to generate infinipair clusters. The reads may be clustered using "infinipair" technology and used to match clusters. Identification of new cis and trans interactions may be identified using previously described methods (16).

[0042] Figure 30 illustrates a method using infinipair to model chromosome conformation in small numbers of cells.

[0043] Figure 31 illustrates a sample preparation for *in situ* library construction. a) Size-selected HMW genomic DNA is end repaired and then ligated to hairpin adapters containing uracil nucleotides near the loop region. Blue and red indicate different priming sequences and each template molecule has a 50% chance of ligating to two different primer sequences. Treatment of the ligation products with exonuclease III and VII removes unligated DNA molecules that have exposed 3' or 5' ends. Uracil-specific excision reagent (USER™) treatment excises the uracil bases to open the hairpins and generate a flowcell-ready library with single-stranded 3'-tails. b) The library is loaded on a standard Illumina flowcell and both ends are allowed to hybridize. A hyperactive transposase is used to randomly fragment and insert common flowcell adapters in the HMW hybridized library to generate LMW cluster-ready templates. After cluster generation, reads from either end can be deconvolved by using the two different sequencing primers (shown in red and blue).

[0044] Figure 32 shows nearest neighbor pairs that were within 1.5 um of each other and 4,000 bp mapping distance were identified by comparing a) read 1 against

PARSE EX.1007.179

read 1, b) read 2 against read 2, c) read 1 against read 2, and d) read 2 against read 1. The three colors represent three different sized libraries: blue = 1 kb, green = 2 kb, red = 3kb. The cumulative number of cluster pairs is plotted against the numerically sorted mapping distance for each pair.

[0045] Figure 33 shows nearest neighbor cluster pair data for the 1, 2, and 3 kb libraries for different nearest neighbor searches. The white bars are the total number of cluster pairs with <1.5 μm physical separation and < 4000 bp mapping separation. The grey bars are the number of pairs within the targeted size range for that library size (800-1200, 1500-2300, and 2500-3500 bp, respectively). The colored bars are pairs that are within the targeted size range and have reads on opposite strands in opposite directions.

[0046] Figure 34 is a series of data illustrating cluster separation in read 1 and 2 according to one embodiment. a) Every cluster that had a nearest neighbor within 1.5 um and 4,000 bp mapping distance was identified within read 1 for the three libraries (blue = 1 kb, green = 2 kb, red = 3kb). The mapping distance is plotted against the cluster separation distance and histograms along each axis are shown. Note that the native Illumina image processing software will not demarcate two clusters that are closer than ~0.9 μm. b) The nearest neighbors for every cluster in read 1 was identified in read 2 and plotted as above.

[0047] Figure 35 shows illustrative images of stretched DNA according to one embodiment. (a) 48.5kb lambda genomes were stained with JOJO-1, tethered to a modified Illumina flowcell, and stretched with a 15V/cm electric field. Imaging was performed on an Illumina GA2x. (b) The stretched DNA was then treated with transposomes for 5 minutes at 55°C and imaged again. Scale bars = 20 μm.

[0048] Figure 36 shows the tn5mC-seq method and resulting methylation profiles according to one embodiment. (a) Tagmentation-based DNA-seq library construction. Genomic DNA is attacked by transposase homodimers loaded with synthetic, discontinuous oligos (yellow, purple) that allow for fragmentation and adaptor incorporation in a single step. Subsequent PCR appends outer flowcell-compatible primers (pink, green). (b) tn5mC-seq library construction. Loaded transposase attacks

PARSE EX.1007.180

genomic DNA with a single methylated adaptor (yellow). An oligo-replacement approach anneals a second methylated adaptor (purple) which is then subject to gap-repair. Bisulfite treatment then converts unmethylated cytosine to uracil (orange) followed by PCR to append outer flowcell-compatible primers (pink, green). Methylation is represented as black lollipops. (c) Coverage of cytosine positions genome-wide. >96% of Cs in all three contexts are covered at least once. Slight decrease in CpG coverage is due to reduced read alignment ability at regions with a high density of methylation. (d) Normalized methylated cytosine over total cytosine positions in 10kb windows across chromosome 12 (max set to 1.0), black box indicates centromere. (e) Normalized methylated CpG over total CpG residues at annotated genic loci. Promoter is defined as 2kb region upstream of TSS. (f) Elevated CpG methylation levels in gene body (intron, exon) compared to intergenic regions.

[0049] Figure 37 illustrates distribution of average raw quality score for all unmapping read 1's in the 3 kb library (a) and for all nearest neighbor (NN) pairs consisting of one *E. coli* and one unmapped read, the average raw quality score for the unmapped read is shown in a histogram.

[0050] Figure 38 illustrates the average raw quality score across all bases for read 1 (A) and read 2 (B) in the 3 kb library. Reads are those found in nearest neighbor pairs that mapped to E. coli, separation < 1.5 μm, and mapped between 2500 and 3500 bp.

[0051] Figure 39 shows Plots of G_{surf} for the x,y and z components of the end-to-end vector $\vec{r}$ are shown for DNA tethered to a surface (a, b). (c) shows araphic illustration of what may be happening during cluster formation. When two seed templates are localized in close proximity on the surface, as cluster amplification proceeds there is a local depletion of available surface primers. This forces the clusters to grow away from each other. During basecalling, the cluster centers are called at a x-y positions that do not coincide with the original seeding templates.

[0052] Figure 40 is a schematic illustration of the *in situ* stretching process described herein (a). One end of a HMW molecule was hybridized to a surface prior to the application of an electric field. While the field is applied, molecules with a free end

PARSE EX.1007.181

are stretched in the direction of the current flow. The free end is then able to hybridize and sequencing proceeds as usual. (b) shows angles between clusters determined by selecting the cluster furthest from the positive electrode as the reference (r). The angle to the other cluster (oc) was then calculated.

[0053] Figure 41 is a set of scatterplots showing mapping distance vs. physical separation for the 3 kb *E. coli* library in the absence of an applied external electric field (a). For the points shown in the boxes, histograms of the relative angle (in radians) between pairs are shown on the right. (b) shows the plots as in (a) but under-hybridization was performed in the presence of a 28 V/cm electric field. Cluster pairs that were separated by at least 4.5 pixels appear to be aligned along the axis of the flowcell and parallel to the electric field (bottom right).

DETAILED DESCRIPTION

[0054] Methods of capturing contiguity information are provided herein. The contiguity information and the embodiments for receiving such information may be used with any suitable traditional or second generation DNA sequencing technology to improve the efficacy and accuracy of the technology and related uses and applications; and to increase its cost effectiveness. Suitable DNA sequencing technologies that may be used in accordance with the methods described herein may include, but are not limited to, "cyclic-array" methods (e.g., 454 pyrosequencing, Illumina Genome Analyzer, AB SOLiD, and HeliScope), nanopore sequencing methods, real-time observation of DNA synthesis, sequencing by electron microscopy, dideoxy termination and electrophoresis, microelectrophoretic methods, sequencing by hybridization, and mass spectroscopy methods.

[0055] Many of these sequencing methods include several common procedural concepts to sequence a long strand of DNA (or "target DNA sequence"). First, the target DNA sequence is broken up into numerous small sequence fragments (or "DNA fragments"). This may be accomplished by treating the target DNA with a transposase. In some examples. the numerous DNA fragments may be considered to be a DNA fragment library (or "shotgun library"). Next, the DNA fragments may be amplified or cloned, resulting in the generation of clonal copies or clusters. The clonal copies or

PARSE EX.1007.182

clusters are then sequenced by a sequencing platform, such as those described above. After sequencing, the sequenced DNA fragments may be reassembled to reconstruct the original sequence, or mapped to a reference genome to identify sequence variations.

Capturing contiguity information

[0056] As discussed above, when a target DNA sequence is treated with transposase, the target DNA may be broken up into two or more DNA fragments that, prior to the transposase treatment (i.e., prior to fragmentation), were connected via one or more spatial relationships. In one embodiment, the spatial relationship is an adjacent relationship, wherein the DNA fragments were directly adjacent to one another (i.e., the end of one DNA fragment was connected to the end of a second DNA fragment). In another embodiment, the spatial relationship may be a compartmental relationship, wherein the target DNA comprises two or more sequence segments that are categorized as compartments. In such an embodiment, DNA fragments prior to fragmentation by transposase may have been within the same segment of the target DNA, but not necessarily adjacent to one another. In another embodiment, the spatial relationship is a distance relationship wherein the DNA fragments were non-contiguous and non-adjacent prior to fragmentation, but are related by a particular distance or sequence length between each other. These spatial relationships may be determined by capturing contiguity information using methods described herein.

[0057] Contiguity information refers to a spatial relationship between two or more DNA fragments based on shared information. The shared aspect of the information can be with respect to adjacent, compartmental and distance spatial relationships. Information regarding these relationships in turn facilitates hierarchical assembly or mapping of sequence reads derived from the DNA fragments. This contiguity information improves the efficiency and accuracy of such assembly or mapping because traditional assembly or mapping methods used in association with conventional shotgun sequencing do not take into account the relative genomic origins or coordinates of the individual sequence reads as they relate to the spatial relationship between the two or more DNA fragments from which the individual sequence reads were derived.

PARSE EX.1007.183

Therefore, according to the embodiments described herein, methods of capturing contiguity information may be accomplished by short range contiguity methods to determine adjacent spatial relationships, mid-range contiguity methods to determine compartmental spatial relationships, or long range contiguity methods to determine distance spatial relationships. These methods facilitate the accuracy and quality of DNA sequence assembly or mapping, and may be used with any sequencing method, such as those described above.

[0058] According to the embodiments described herein, the methods for capturing contiguity information may include treating a target DNA sequence with a transposase resulting in one or more fragmentation or inserting events. In some embodiments, this step results in the generation of a library of shotgun nucleic acid molecules derived from the target DNA sequence. In an alternative embodiment, the fragmentation or insertion even may be accomplished by a Y adaptor approach as described below. The one or more transposase molecules may be soluble free transposase or may be associated with a surface-bound recognition sequence.

[0059] The target DNA, after treating with the transposase may comprise two or more DNA fragments or a plurality of DNA fragments (also referred to as "the fragmented target DNA") or may comprise an insertion sequence ("the insertion target DNA").

[0060] In some embodiments, the methods for capturing contiguity information may include a step of amplifying the DNA or shotgun library to generate clonal copies or clusters of reads. The amplification step may include, but is not limited to any suitable amplification method such as polony, emulsion PCR, and bridge PCR.

[0061] In some embodiments, after treatment with transposase or after a subsequent amplification, one or more recognition sequences may be added to or inserted into the fragmented or insertion target DNA. The one or more recognition sequences may include, but are not limited to, a barcode, a primer or an adaptor DNA sequence at the site of the fragmentation or insertion that tags the DNA fragment as unique with respect to the adjacent, compartmental or distance spatial relationship.

PARSE EX.1007.184

[0062] After being tagged, the shotgun nucleic acid molecules may be sequenced using a sequencing platform described above contiguity information is captured by identifying recognition sequences that have a shared property. In some embodiments, the shared property is an identical or complementary barcode sequence. For example, read sequences of adjacent origin may be identified via shared barcode sequences; or reads may be defined by compartments based on shared compartment-specific barcodes derived from the same target DNA segment. In other embodiments, the shared property is a shared or constrained physical location, which may be indicated by one or more x,y coordinates on a flowcell. A "constrained" physical location may refer to a close, identical, or nearly identical physical location or to a set of two or more physical locations whose relative physical coordinates are correlated with the relative sequence coordinates on the target DNA sequence from which the DNA fragments were derived. For example, in methods relating to long-range contiguity, *in situ* transposition into stretched, HMW genomic DNA on the surface of a sequencing flowcell is performed using adaptor sequences to obtain distance spatial relationships by identification of the constrained physical locations (i.e. the relative coordinates at which physically linked sequencing templates are immobilized) of the adaptor sequences, hybridized DNA fragments, or a combination thereof. Additional embodiments and details regarding capturing short-range, mid-range and long-range contiguity are described further below.

[0063] *Short range contiguity*. To capture information on short-range contiguity, a modified scheme for *in vitro* transposition in which degenerate barcodes within synthetic transposons are used in methods to symmetrically and uniquely tag shotgun library molecules originating from each flank of any given fragmentation event is provided, such that one can subsequently assign *in silico* "joins" between independent, adjacent-in-origin read-pairs. After sequencing the shotgun library and corresponding barcodes, adjacent fragmentation events can be identified via shared barcode sequences. Importantly, this strategy allows for the determination of local contiguity in a way that is almost completely independent of the primary sequence content.

[0064] *Mid-range contiguity*. Even with long, high accuracy Sanger reads, the hierarchical approach of sequencing BAC clones was important to achieve a high

PARSE EX.1007.185

quality reference assembly of the human genome, particularly in segmentally duplicated and structurally complex regions (Lander et al. 2001; Waterston et al. 2003; Waterston et al. 2002). Therefore, in some embodiments, methods that enable the grouping of short (or "shotgun") reads derived from the same fosmid/BAC-scale region of the genome (e.g., 20 to 200 Kb), to capture information for mid-range congruity are provided. These methods are discussed in detail below in Example 2.

[0065] As described below and in Kitzman *et al.* (Kitzman et al. 2011), this class of information is sufficient to extensively haplotype-resolve an individual human genome sequence. This mid-range contiguity information may also facilitate *de novo* genome assembly. For example, Gnerre *et al.* (Gnerre et al. 2010) recently described the *de novo* assembly of the human and mouse genomes to reasonably high quality using only short-read sequence data. This result, just as with the haplotype contiguity achieved by Kitzman *et al.* (Kitzman et al. 2011), required the use of fosmid library construction in order to partition the genome into ~40 Kb segments. In these methods, emulsions are used to compartmentalize high molecular weight (HMW) genomic DNA fragments, followed by emulsion PCR with primers bearing droplet-specific barcodes. Upon recovery, amplicons are tagged with barcodes that define groups of shotgun reads, with each group derived from the same 20 – 200 Kb region. In preliminary work relying on shotgun libraries derived from complex pools of fosmid clones, the sufficiency of this class of information to extensively haplotype-resolve an individual human genome with next-generation sequencing is demonstrated below.

[0066] Similar to the recently reported "subassembly" strategy (Hiatt et al. 2010), a long fragment library is converted to a population of nested sub-libraries, and a tag sequence directs the *in silico* grouping of short reads that are derived from the same long fragment, thereby enabling the localized assembly of long fragment sequences, i.e. "subassembled" reads. Subassembly extends the utility of short-read sequencing platforms to applications that normally require or benefit from long reads, e.g. metagenomics and *de novo* genome assembly. However, the methods according to the embodiments described herein enable subassembly over 20 – 200 Kb, rather than ~1 Kb, regions as previously described.

PARSE EX.1007.186

[0067] *Long-range contiguity*. High throughput methods that include massively parallel, short read sequencing technologies are inherently limited with respect to several important goals, including the resequencing of segmental duplications and structurally complex regions of the human genome, the resolution of haplotype information in diploid and polyploidy genomes, and the *de novo* assembly of complex genomes. Further reductions in the cost-per-base of sequencing will do little to advance these goals. Rather, what is required are equivalently parallel methods of obtaining contiguity information at different scales. For example, the fact that the original *de novo* assemblies of the human and mouse genomes achieved a high quality (Lander et al. 2001; MSGC 2002), despite an order-of-magnitude less sequence coverage than lower quality assemblies based on short reads alone, is primarily a consequence of the inclusion of a broad spectrum of complementary sources of contiguity information, including: (a) long primary read lengths, (b) mate-paired reads from plasmids, fosmids, and BACs, (c) hierarchical clone-by-clone sequencing, and (d) genetic maps.

[0068] Although new approaches to DNA sequencing may continue to mature and surpass current technology, the most cost-efficient technologies (in terms of cost-per-base) may continue to be read-length limited. Therefore, contiguity information may be obtained, by supplementing low-cost, short-read sequences with contiguity information obtained by other technologies described below. Examples of methods for obtaining contiguity information in this way may include: 1) Long-range "mate-pair" protocols enable one to obtain read-pairs separated by a controlled distance. However, all current *in vitro* protocols employ a circularization step, such that the method is only efficient at separations of several kilobases. 2) Barcoding and sequencing of clone dilution pools (or their *in vitro* equivalent) can yield haplotype information on a genome-wide scale. However, the resolution of the method is limited to the types of fragments (*e.g.* fosmid) and number of pools that one can efficiently process. 3) Optical mapping using restriction enzymes has been successful in generating long-range contiguity maps for *de novo* genome assembly (Schwartz et al. 1993; Zhou et al. 2007; Zhou et al. 2009). However, this process is limited by false positive and negative cut sites due to star activity and inefficient cleavage, necessitating multiple optical maps from the same region to generate a consensus map. Furthermore, the non-uniform distribution of

PARSE EX.1007.187

restriction enzyme recognition sites can limit the amount of useful information derived from repetitive or low complexity regions. 4) Optical sequencing on stretched single DNA molecules (non-fragmented) has yielded up to 3 bp of contiguous sequence information from multiple locations along the same molecule (Ramanathan et al. 2004). Because reads are generated directly from single molecules, issues of sample quantity and PCR bias are largely avoided.

[0069] As described in Example 3 below, *in situ* library construction and optical sequencing within the flow-cells of next-generation sequencing instruments represent an improved and efficient path towards a single technology that simultaneously captures contiguity information and primary sequence at diverse scales. The basic premise is to exploit the physical properties of DNA (by random coiling or stretching of high-molecular weight (HMW) DNA), *in situ* library construction (via *in vitro* transposition of adaptors to HMW DNA within a flow-cell), and the fully developed aspects of an operationally-realized next-generation sequencing instrument (polony amplification, sequencing-by-synthesis, imaging and data-processing), to generate multiple spatially related reads whose physical separation is either known or can be inferred from the relative coordinates at which the reads originate on the flow-cell. In one approach, the random coil configuration adopted by DNA in solution is exploited to spatially confine the ends and generate two reads within a confined surface area. In a related approach, optical sequencing on stretched DNA molecules within a native flowcell may also be performed.

[0070] These approaches are discussed in detail below and, according to some embodiments, illustrate *in vitro* methods for long-distance mate-pairing that are not dependent on any circularization step. Success in obtaining paired-end reads from unstretched 2.7 Kb molecules is shown in Figure 12b. Briefly, flowcell compatible adaptors (FCA1) were end-ligated to linearized, double-stranded puc19. This template was introduced to a flowcell (Illumina) and single-stranded ends were allowed to hybridize to the primer-coated surface. The templates were then treated *in situ* with transposase pre-loaded with FCA2 adaptors. Next, standard cluster PCR was performed, followed by sequencing-by-synthesis. Based on the primers used and the known sequence of pUC19, the first 4 bp were likely to be either AGCT or CGAG,

PARSE EX.1007.188

depending on which end of the molecule the read was coming from. Figure 13A (top) shows representative images of a spatially separated "cluster pair" for the first 4 cycles, and raw integrated basecalling intensities for both templates is shown in Figure 13B (bottom). The observation of many such closely located pairs in an otherwise sparse field is consistent with a common origin from the ends of the same 2.7 Kb molecules. Further diluting the template still produced cluster pairs, strongly suggesting that these are not derived from two different templates that happened to hybridize nearby. Also, only ~20% of templates showed visible physical cluster separation (as in Figure 13), while the remaining 80% of paired ends were co-localized and gave mixed reads. However, the proposed approach of using two different sequencing primers will allow deconvolved mixed reads from such immediately co-localized cluster pairs into two separate reads.

[0071] In other embodiments, the *in situ* fragmentation of linearly stretched 48.5 Kb DNA molecules is also demonstrated with transposomes. Briefly, flow-cells were cleaned using Piranha solution, treated with 2% 3-aminopropyltriethoxysilane (APTES), and loaded with JOJO-1 stained lambda DNA. The flowcell was then loaded with 6M KCl and an electric field of 15V/cm was applied at the input and output ports for 90 sec. Surfaces were imaged directly on an Illumina GA2 sequencer (Figure 14A) to demonstrate that the ends of single 48.5 Kb molecules can be physically stretched over ~30 pixels. Surfaces were then treated *in situ* with transposome and re-imaged (Figure 14B). Individual molecules were fragmented in multiple locations, demonstrating the enzyme's ability to maintain high activity even on surface-immobilized template. These methods may also be used to incorporate flowing in the "lock-down" bridge prior to fragmentation, so that clusters may be generated at the ends of long templates.

[0072] Based on the methods of short, mid-range and long-range contiguity embodiments described herein, several additional embodiments for capturing contiguity are provided below.

[0073] According to some embodiments, methods for capturing contiguity information are provided. In one embodiment, such methods may include constructing a library of shotgun nucleic acid molecules derived from target DNA wherein sequences

PARSE EX.1007.189

adjacent to each fragmentation or insertion event are symmetrically tagged with barcodes; sequencing the shotgun library molecules and corresponding barcodes; and identifying sequences of adjacent origin via shared barcode sequences.

[0074] In another embodiment, methods for capturing contiguity information may include compartmentalizing target DNA fragments with emulsions or dilution; modifying target DNA fragments with transposase to insert primer sequences, either before or after compartmentalization; performing nucleic acid amplification using primers bearing compartment-specific barcodes; and sequencing the resulting library of shotgun nucleic acid molecules derived from target DNA and corresponding barcodes to define groups of shotgun sequence reads. In one aspect, the groups of reads sharing barcodes are derived from the same high molecular weight genomic DNA fragment.

[0075] In a further embodiment, methods for capturing contiguity information may include end-modifying target DNA molecules with an adaptor corresponding to one surface-bound primer; hybridizing both ends of the end-modified target DNA molecules to the surface-bound primer with or without stretching; performing transposition with non-surface-bound transposase complexes that include DNA transposase and sequences corresponding to a second surface-bound primer; performing cluster amplification to produce clusters of clonally derived nucleic acids; sequencing clusters of clonally derived nucleic acids; and determining whether overlapping or closely located clusters are derived from ends of the same target DNA molecules. In one aspect, such a method includes end-modifying high molecular weight DNA molecules with an adaptor corresponding to one flow cell primer; hybridizing both ends of the end-modified high molecular weight DNA molecules to a flowcell with or without stretching; performing *in situ* transposition with transposase loaded with adaptors corresponding to a second flow cell primer; performing cluster PCR to produce visibly overlapping or closely located clusters; and determining whether overlapping or closely located clusters are derived from ends of the same high molecular weight DNA molecule.

[0076] In another embodiment, methods for capturing contiguity information may include modifying target DNA molecules with transposase to insert nucleic acid sequences corresponding to one or several surface-bound primers; hybridizing the

PARSE EX.1007.190

internally modified target DNA molecules to the surface-bound primers with or without stretching; performing cluster amplification to produce clusters of clonally derived nucleic acids; sequencing clusters of clonally derived nucleic acids; and determining whether overlapping or closely located clusters are derived from the same target DNA molecules. In one aspect, such a method includes modifying high molecular weight genomic DNA with transposase to insert primer sequences corresponding to one or two flow cell primers; hybridizing the internally modified high molecular weight DNA molecules to a flowcell with or without stretching; performing cluster PCR to produce visibly overlapping or closely located clusters; and determining whether overlapping or closely located clusters are derived from the same high molecular weight DNA molecules as in Figure 25.

[0077] In another embodiment, methods for capturing contiguity information include steps of (a) generating a surface to which nucleic acid sequences are bound that include a double-stranded DNA sequence corresponding to the recognition sequence of a DNA transposase; (b) assembling complexes comprising a DNA transposase bound to the surface-bound recognition sequence; (c) exposing complexes to target DNA, with or without stretching of the target DNA, and allowing for internal modification of the target DNA by the surface-bound transposase complex; (d) performing cluster amplification to produce clusters of clonally derived nucleic acids; (e) sequencing clusters of clonally derived nucleic acids; and (f) determining whether overlapping or closely located clusters are derived from the same target DNA molecule. In one aspect, an additional step may be included at any point before step (c) wherein target DNA is modified by exposure to non-surface-bound transposase complexes that include DNA transposase and sequences corresponding to a surface-bound primer. In another aspect, an additional step after step (c) and before step (d) may be included, wherein target DNA is further modified by exposure to non-surface-bound transposase complexes that include DNA transposase and sequences corresponding to a surface-bound primer.

PARSE EX.1007.191

Applications of sequencing technologies

[0078] The methods of capturing contiguity information described herein are useful in the improvement of uses and applications of the sequencing technologies described above. Suitable applications of DNA sequencing technologies that may be used in accordance with the methods described herein may include, but are not limited to bisulfite sequencing for determining DNA methylation, resequencing, *de novo* assembly, exome sequencing, RNA-Seq, ChIP-Seq, inferring chromosome conformation and genome-wide chromatin interaction mapping. In some embodiments, the methods for capturing contiguity information may be used with "cyclic-array" methods, for applications such as resequencing, *de novo* assembly, or both as described in detail in the Examples below.

[0079] *Resequencing*. Resequencing human genomes has become relatively straightforward. For example, Bentley *et al*. (2008) sequenced the genome of a Yoruba male to ~40x coverage to identify ~4 million SNPs on the Illumina GA platform (Branton et al. 2008), i.e. massively parallel sequencing-by-synthesis on a dense array of unordered PCR colonies. Today, the Illumina HiSeq platform is able to generate the same quantity of data (135 gigabases (Gb)) in 8 days across 7 sequencing lanes that each yield ~100 million mappable, paired-end, 100 bp reads (PE100). For an exemplar cost of $3,700 per lane, the estimated cost for ~40x human genome resequencing is just over $25,000.

[0080] Furthermore, although short read lengths and modest raw accuracies are compatible with the highly accurate resequencing of ~94% of the human genome, that these technologies continue to fall short in at least two important ways. First, approximately 6% of the human genome consists of gene-rich segmental duplications or structurally complex regions that are prone to recurrent rearrangement. It is likely impossible to uniquely map short sequencing reads within this space, and extremely challenging to decipher complex structural variation. Second, current technology for genome resequencing is almost completely blind to haplotype, i.e., the phase with which polymorphisms along a single chromosome occur. Haplotype information is extremely useful for studies of gene-disease association, as well as for population

PARSE EX.1007.192

genetic analyses. Neither of these deficiencies can be remedied by more sequencing with the same technology. Rather, these deficiencies reflect fundamental limitations of short-read sequencing.

[0081] *De novo assembly*. In contrast with resequencing, there is still a long way to go with respect to generating high-quality *de novo* assembly of mammalian genomes using the same technologies. Generating 20 Gb, i.e. the ~8x coverage (Sanger) used to assemble the 2.5 Gb mouse genome in 2002 (Waterston et al. 2002), is now possible on a single Illumina HiSeq lane (PE100, $3,700). However, even with ~90x coverage, the best "next-generation" *de novo* assembly of the similarly complex human genome yields an N50 contig length of 7.4 Kb, a N50 scaffold length of 446 Kb, and sequence coverage of just 87% of the genome (Li et al. 2010). Further increases in coverage with short-read data would likely only minimally improve assembly quality (Li et al. 2010). By comparison, the initial assembly of the mouse genome, based on over an order of magnitude of less data, had an N50 contig length of 25.9 Kb, an N50 scaffold length of 18.6 megabases (Mb), and sequence coverage of 95% of the genome (Waterston et al. 2002).

[0082] *Bisulfite sequencing*. Methods for bisulfite sequencing for measurement of DNA methylation are provided herein. DNA methylation is a widespread epigenetic modification that plays a pivotal role in the regulation of the genomes of diverse organisms. The most prevalent and widely studied form of DNA methylation in mammalian genomes occurs at the 5 carbon position of cytosine residues, usually in the context of the CpG dinucleotide. Microarrays, and more recently massively parallel sequencing, have enabled the interrogation of cytosine methylation (5mC) on a genome-wide scale (Zilberman and Henikoff 2007). However, the *in vivo* study of DNA methylation and other epigenetic marks, *e.g.* in specific cell types or anatomical structures, is sharply limited by the relatively high amount of input material required for contemporary protocols.

[0083] Methods for genome-scale interrogation of methylation patterns include several that are preceded by the enrichment of defined subsets of the genome (Meissner et al. 2005; Down et al. 2008; Deng et al. 2009), *e.g.,* reduced representation

PARSE EX.1007.193

bisulfite sequencing (RRBS) (Meissner et al. 2005) and anti-methylcytosine DNA immunoprecipitation followed by sequencing (MeDIP-seq) (Down et al. 2008). An advantage of such methods is that they can be performed with limited quantities of starting DNA (Gu et al. 2011). However, they are constrained in that they are not truly comprehensive. For example, the digestion-based RRBS method interrogates only ~12% of CpGs, primarily in CpG islands (Harris et al. 2010), with poor coverage of methylation in gene bodies (Ball et al. 2009) and elsewhere. Furthermore, RRBS does not target cytosines in the CHG or CHH (H = A,C,T) contexts which have been shown to be methylated at elevated levels in the early stages of mammalian development (Lister et al. 2009).

[0084] The most comprehensive, highest resolution method for detecting 5mC is whole genome bisulfite sequencing (WGBS) (Cokus et al. 2008; Lister et al. 2009; Harris et al. 2010). Treatment of genomic DNA with sodium bisulfite chemically deaminates cytosines much more rapidly than 5mC, preferentially converting them to uracils (Clark et al. 1994). With massively parallel sequencing, these can be detected on a genome-wide scale at single base-pair resolution. This approach has revealed complex and unexpected methylation patterns and variation, particularly in the CHG and CHH contexts. Furthermore, as the costs of massively parallel sequencing continue to fall, whole genome bisulfite sequencing is increasingly affordable. However, WGBS is limited in that current protocols call for 5 micrograms of genomic DNA as input (Cokus et al. 2008; Lister et al. 2009; Li et al. 2010), which is essentially prohibitive for many samples obtained *in vivo*.

[0085] In some embodiments, a transposase-based *in vitro* shotgun library construction ("tagmentation") for whole genome bisulfite sequencing is adapted as described below. This method, referred to herein as tn5mC-seq, enables a >100-fold reduction in starting material relative to conventional protocols, such that highly complex bisulfite sequencing libraries are generated from as little as 10 nanograms of input DNA, and ample useful sequence from 1 nanogram of input DNA. tn5mC-seq is demonstrated by sequencing the methylome of a human lymphoblastoid cell line to approximately 8.6X high quality coverage of each strand.

PARSE EX.1007.194

[0086] Further, methods for methylating discontinuous synthetic transposons are provided that use a double stranded DNA portion of the Tn5 recognition sequence as well as a single stranded DNA overhang containing either adaptor sequence 1 or 2 wherein all cytidine or cytosine residues are methylated. In one embodiment, a nick translation step is performed. After the nick traslation, the resulting transposition generates adaptor flanked DNA fragments where each strand has both adaptors, one of which is methylated. PCR is then performed on the nick translated material with an accepted lower efficiency of the unmethylated strand of the adaptor that was generated from the nick translation.

[0087] In another embodiment, the nick translation step is not performed and the second adaptor is added later as described below. The fragment library is then subjected to bisulfite treatment to convert all unmethylated cytidines to uracil residues. The second adaptor is then added added in one of two ways: (1) by adding an A-tail and then using a primer containing poly-T and an adaptor overhang, or (2) by extending a template containing a 3' blocked N6 (at bisulfite treated nucleotide ratios) with a 5' adaptor overhang that will be extended through from the 3' end of the fragment. After addition of the second adaptor, PCR and sequencing is then performed. One advantage of this method is that the high efficiency of conversion of gDNA to adaptor modified fragments will allow for much less DNA to be used in the construction of libraries to be subjected to bisulfite treatment.

[0088] Briefly, the procedure is as follows. First, transposase with adaptors containing the dsDNA transposase recognition sequences are loaded with an ssDNA adaptor overhang in which all cytosine (C) residues are methylated. Next, transposition into genomic DNA is performed, fragmenting the DNA and appending a methylated C, 5' overhang adaptor. If nick translation is performed, the adaptor is extended to both ends of the molecule, however, the 3' adaptor will not be methylated. The library is then subjected to bisulfite treatment to convert all unmethylated C residues to U residues. If nick translation was not performed in the previous step, a second 3' adaptor may be added by one of two approaches: (i) DNA fragments are A-tailed, and the 3' adaptor is appended to the fragments using a 3' poly-T 5' adaptor primer; or (ii) DNA fragments are allowed to extend on an oligo comprised of a 3' blocked N6 (at complementary

PARSE EX.1007.195

bisulfite treated nucleotide composition) and a 5' adaptor overhang. Finally, PCR is performed, followed by sequencing

[0089] According to other embodiments, the method of bisulfite sequencing may include steps of (a) performing *in vitro* transposition into target DNA molecules with transposase complexes that include double stranded DNA transposase recognition sequences with a single stranded DNA adaptor overhang having methylated cytosine residues; (b) subjecting modified target DNA molecules to bisulfite treatment; (c) performing nucleic acid amplification to produce a nucleic acid library; and (d) sequencing the resulting nucleic acid library. In some aspects, a second adaptor to nucleic acid fragments derived from target DNA after step (a) and before step (b), wherein the second adaptor is designed to facilitate nucleic acid amplification in step (c) may be incorporated. In other aspects, a second adaptor to nucleic acid fragments derived from target DNA, after step (b) and before step (c), wherein the second adaptor is designed to facilitate nucleic acid amplification in step (c).

[0090] In other embodiments, the method of bisulfite may include steps of (a) modifying double stranded DNA (dsDNA) transposase recognition sequences with a single stranded DNA (ssDNA) adaptor overhang having methylated cytosine residues; (b) performing *in vitro* transposition with transposase loaded with adaptors containing the modified dsDNA transposase recognition sequences to generate a library of DNA fragments; (c) subjecting the library of DNA fragments to bisulfite treatment; (d) performing a PCR method to amplify a target; and (c) sequencing the target. In some embodiments, an additional step of nick translation may be performed after step b) and before step (c). In other embodiments, nick translation is not performed. In this case, a second adaptor is added after step (c) and before step (d). The second adaptor may be added by (i) adding an adenosine (A) tail to the DNA fragments and appending a 3' adaptor to the fragments using a 3' poly-T 5' adaptor primer; or (ii) allowing the DNA fragments to extend on an oligonucleotide comprising a 3' blocked N6 and a 5' adaptor overhang.

[0091] *Inferring chromosome conformation*. According to some embodiments, methods for inferring chromosome conformation are provided. These methods may

PARSE EX.1007.196

include cross-linking DNA within cells; isolating chromatin fibers; removing and digesting chromatin fragments; purifying chromatin DNA fragments; ligating adaptors to chromatin DNA fragments, forming chromatin DNA fragment complexes; and generating 3-dimensional models of chromosomal positions by pairing neighboring clusters of chromatin DNA fragment complexes. In one embodiment, the method may include steps of (a) cross-linking DNA within cells; (b) isolating cross-linked DNA from cells; (c) fragmenting the cross-linked DNA; (d) end-modifying fragmented, cross-linked DNA molecules with an adaptor corresponding to one surface-bound primer; (e) hybridizing ends of the fragmented, end-modified target DNA molecules to the surface-bound primer; (f) performing transposition with non-surface-bound transposase complexes that include DNA transposase and sequences corresponding to a second surface-bound primer; (g) performing cluster amplification to produce clusters of clonally derived nucleic acids; (h) sequencing clusters of clonally derived nucleic acids; and (i) determining physical interactions between chromosomal positions by paring neighboring clusters together. In some aspects, an isolated cross-linked DNA may be part of a cross linked DNA-protein complex. In this case, the method for inferring chromosome further conformation may additionally include a step of enriching for one or more specific cross linked DNA-protein complexes by immunoprecipitation after step (c) and before step (d).

[0092] In other embodiments, a method for identifying interactions between transcription factor binding sites is provided. Such a method may include inducing a population of cells with a hormone; immunoprecipitating cells to isolate chromatin fibers; producing chromatin fragments by cross linking cells and breaking chromatin fibers; repairing ends of chromatin fragments and ligating ends to adaptors, producing chromatin complexes; generating clusters corresponding to the chromatin complexes; and determining interactions between chromosomal positions by paring neighboring clusters together.

[0093] The following examples are intended to illustrate various embodiments of the invention. As such, the specific embodiments discussed are not to be construed as limitations on the scope of the invention. It will be apparent to one skilled in the art that various equivalents, changes, and modifications may be made without departing from

PARSE EX.1007.197

the scope of invention, and it is understood that such equivalent embodiments are to be included herein. Further, all references cited in the disclosure are hereby incorporated by reference in their entirety, as if fully set forth herein.

EXAMPLES

[0094] Several properties of *in vitro* transposition may be exploited to develop ultra-low-cost, massively parallel sequencing methods for capturing contiguity information at diverse scales. First, modified Tn5 transposomes attack DNA *in vitro* with high efficiency and at high density, in a reaction that catalyzes the insertion of common sequences, with or without fragmentation depending on whether the synthetic transposon is continuous or discontinuous. Second, the pattern of transposome attack is relatively random with respect to sequence content. Third, degenerate subsequences, in addition to common adaptor sequences, may be readily included within the synthetic transposons. Fourth, *in vitro* transposition is inexpensive as a single volume, aqueous-phase, enzymatic reaction. Examples 1-3 are directed at the development of massively parallel methods that exploit *in vitro* transposition to inform short-range, mid-range, and long-range contiguity, respectively. Example 4 is directed at the development of methods that exploit *in vitro* methylated transposition to capture contiguity information. Example 5 is directed at the development of methods for measuring DNA-DNA and DNA-protein interactions within smaller populations of cells that exploit infinipair technology to directly sequence multiple fragments off of immunoprecipitated DNA that has been cross linked. Example 6 is directed at integrating these methods to demonstrate high quality *de novo* genome assembly and haplotype-resolved genome resequencing.

General Approach

[0095] *Contiguity information is a primary goal*. The methods in the Examples described below address a "blind spot" in the next-generation sequencing field. Specifically, the methods address the lack of ultra-low-cost methods to determine contiguity information at broader scales.

[0096] These methods and their associated costs are dependent on the sequencing technology with which they are integrated, as this is the method by which

PARSE EX.1007.198

the primary sequence coupled to the contiguity information is decoded. The methods below are performed using a commercially available, cyclic-array platform (e.g., Illumina GA2x or HiSeq). However, all of the methods described herein may be integrated with other approaches to DNA sequencing, e.g. nanopore sequencing, other cyclic-array platforms. Broad compatibility will ensure that these methods can be combined with any technology that emerges as the best in terms of cost-per-base.

Materials and Methods

[0097] *In vitro transposition for capturing contiguity information*. Although Examples 1-6 are technically diverse, a common thread is their reliance on high density, random, *in vitro* transposition as a novel means of physically shattering genomic DNA in creative ways that facilitate the recovery of contiguity information at different scales. The initial interest in this technology was based on its potential utility for low-cost, low-input, *in vitro* preparation of shotgun libraries. As shown in Figure 1, a modified Tn5 transposase catalyzes fragmentation and adaptor incorporation in a single, 5 minute step. In conventional *in vitro* transposition, inverted 19 bp mosaic-end (ME) sequences flanking transposon DNA are recognized by the transposase and form a stable homodimer synaptic complex in solution. This "transposome" inserts the transposon into target DNA. When applied for library preparation, the transposome is instead comprised of enzyme and free ME sequences with adaptor overhangs. Insertion of the discontinuous transposon results in fragmentation via symmetrical insertion of the ME sequence with asymmetrical 5' adaptor overhangs. PCR amplification with primers complementary to the adaptors yields a shotgun fragment library.

[0098] To address concerns regarding insertion bias and library complexity, extensive comparisons were performed with traditional methods of *in vitro* shotgun library construction (Adey et al. 2010). The analysis revealed a slightly greater bias with respect to sequence content at fragmentation sites with the transposome-based method. However, this was of negligible impact in terms of the coverage distribution during whole human genome resequencing (Figure 2), and the methods exhibited equivalent G+C bias. Critically, it was noted that the complexities of transposome

PARSE EX.1007.199

libraries made from as little as 400 nanograms were equivalent to or greater than the complexities of standard libraries made from much larger amounts of input DNA.

[0099] The library complexities observed with this method suggests that the mass conversion efficiency of genomic DNA into adaptor-flanked library is high, as fragmentation events may be occurring in close succession along any given stretch of genomic DNA in order to generate sequencing-compatible fragments of several hundred base-pairs. Indeed, in analyzing the distribution of fragment lengths resulting from this method, we observe a sharp decrease at ~35 bp that is likely secondary to steric hindrance from adjacent, attacking transposomes (Figure 3). Even with a PCR-free version of the protocol (to avoid skewing the fragment size distribution), the data suggests that the bulk of adjacent transposome reactions (>95%) are separated by 35 to 600 bp. In principle, this high efficiency of mass conversion should translate into low input requirements. Consistent with that, even with input as low as 100 picograms (30 haploid equivalents of the human genome), obtain complex libraries may be obtained. At 10 picograms (3 haploid equivalents complexity begins to bottleneck, but millions of uniquely mapping read-pairs may be observed nonetheless.

Example 1: Short-Range Contiguity

1.A. Symmetrically and uniquely tagging fragmentation events

[00100] The fragmentation of genomic DNA, whether by mechanical or enzymatic methods, results in a complete loss of information as to the pairing of molecules that derive from either side of any single "break". To preserve this information, a method was devised to associate a unique barcode with both ends of fragments derived from each break introduced by *in vitro* transposition (Figure 4). In brief, transposase may be used to catalyze *in vitro* insertion of synthetic transposons containing a degenerate single-stranded "bubble" flanked by nicking restriction endonuclease site into very low amounts of genomic DNA, i.e., less than 5 haploid human genome equivalents. In contrast with the approach described in Figure 1, the synthetic transposons are continuous, containing the 19 bp ME sequences along with two endonuclease nicking sites flanking a 25 bp degenerate sequence. Since the degenerate region is not complementary between the top and bottom strands, a single-stranded bubble is present, increasing

PARSE EX.1007.200

flexibility to aid in the formation of a synaptic complex with two transposase monomers. After inserting these synthetic transposons to high density (every 35 to 600 bp), a 9 bp lesion, resulting from the transposition mechanism, is repaired via a gap-fill and ligation reaction.

[00101] The construct is then subjected to primase-based whole genome amplification (pWGA), which resolves the bubbles at the degenerate regions while yielding a relatively uniform amplification (Li et al. 2008). This material is then digested to completion by both nicking endonucleases, which introduce nicks on opposite strands flanking the degenerate region. Finally, extension with a strand-displacing polymerase fragments the target DNA, yielding molecules that terminate in an identical barcode sequence, i.e. symmetrical tagging. At this point, standard protocols (A-tailing, adaptor ligation, PCR) can be applied for compatibility with massively parallel sequencing-by-synthesis. Separate reads can be used to access the barcodes and primary sequence at each end of each library molecule.

[00102] The barcodes used herein should be unique to each fragmentation event because they are derived from a 25 bp degenerate stretch and can be used *in silico* to successively link strings of read-pairs derived from adjacent transposome insertions. These "joins" are based on barcodes alone, thus they are entirely independent of the primary sequence content.

[00103] To test the feasibility of this approach, a synthetic transposon containing a single-stranded bubble with fixed, non-complementary sequences corresponding to two primers was designed (as shown in the first step of Figure 4, but with fixed non-complementary sequences for A & B rather than degenerate sequences). These synthetic transposons were loaded to EZ-Tn5 transposase and reacted with genomic DNA under appropriate conditions. After gap-fill and repair of the expected 9 base-pair lesions resulting from transposition events, PCR with primers corresponding to the non-complementary synthetic bubble sequences yielded amplicons with a broad distribution of sizes ranging from ~0.5 to ~3 Kb (Figure 16). This experiment confirms that synthetic, contiguous transposons containing single-stranded bubbles can be inserted with reasonable efficiency. To achieve a denser distribution of insertion sites, this method

PARSE EX.1007.201

should be optimized. In particular, the efficiency with which the transposase is loaded with synthetic transposons may be improved. As the steric hindrance of adjacent, attacking transposase complexes puts an upper bound on insertion density (Figure 3), a large molar excess of properly loaded transposome complexes will likely achieve a denser insertion distribution.

1.B. Evidence that adjacent events are detectable

[00104] To evaluate whether adjacent fragmentation events are potentially detectable by sequencing, ~2 million uniquely mapping read-pairs were mined from the sequencing of a transposome-fragmented shotgun library derived from 10 picograms of human genomic DNA (3 haploid equivalents). Because 9 bp duplication occurs at each end of each fragmentation event, molecules derived from either side of each event should map to the genome with a 9 bp overlap. As a consequence, a clear increase in "read 2" mapping locations was observed 9 bp from the "read 1" start-sites of other read-pairs (Figure 5). This signature was markedly more pronounced in this ultra-low-input library as compared to libraries that were generated from larger amounts of starting material. Using this 9 bp overlap as evidence for fragments originating from the same breakpoints, chains of 4 to 6 read-pairs were identified that were derived from successive, adjacent fragments that collectively span ~1 Kb to ~2 Kb (Figure 6).

1.C. Method development and performance parameters

[00105] The strategy described above (see 1.A) is one of several related methods that have been devised to (1) attain symmetrically and uniquely tagging fragmentation events and (2) successively link strings of sequence read-pairs derived from adjacent transposome insertions by exploiting these tags during analysis. An alternative approach for symmetrical tagging has also been developed, wherein individual transposases are loaded with symmetrically tagged but formally discontiguous oligonucleotides (or "oligos"), such that both tagging and fragmentation can take place in a single step.

[00106] A method based on this alternative approach was devised to construct a symmetrically tagged, 5'-to-5' linked transposon reagent (Figure 17). To generate this reagent, two primers were linked, one of which contains a 5'-5' inverted adenine RNA

PARSE EX.1007.202

moiety and a 3' phosphate blocking group. Single-stranded ligation between the terminal RNA base with the 5' phosphorylated DNA base of the other oligonucleotide is carried out with T4 RNA ligase. The 5'-5' linked primers are then hybridized to an oligonucleotide containing appropriate complementary sequences for both primers, a degenerate stretch to serve as the tag (e.g. 20 randomized nucleotides – shown in black in Figure 17), and the 19 bp mosaic-end (ME) sequence recognized by the transposon. The first primer in the 5'-5' pair is extended while the other end is blocked by the 3' phosphate. Next, T4 polynucleotide kinase (T4 PNK) is used to remove the 3' blocking phosphate and the second primer is extended with a strand-displacing polymerase. Each single molecule of the resulting species includes two oligonucleotides, linked 5'-to-5' by the inverted adenine moiety, that are identical across the degenerate stretch and that each terminate in the 19 bp mosaic-end (ME) sequence. Gel-based purification is used to remove extension byproducts, and then appropriate oligonucleotides are hybridized to double-strand the ME subsequences at each end. The resulting species are both symmetrically tagged at the single molecule level, and readily loadable to the Tn5 transposase.

[00107] The 5'-5' linked, symmetrically tagged transposon species was constructed as described. The full length product (194 bp) of said transposon, prior to gel purification to remove extension byproducts, is shown in Figure 18a, Successful fragmentation of genomic DNA with said transposon is shown in Figure 18b).

[00108] The success of this approach is dependent on at least two parameters: (1) *Maintenance of library complexity*: The chaining of read-pairs terminates when adjacent fragments on either end of a chain are undetectable in sequencing. For example, at the extreme, if 100% of fragments derived from synthetic transposition were successfully sequenced along with corresponding tags, then in principle it would be possible to chain from end-to-end of entire chromosomes. (2) *Uniformity of representation*: The extent of sequencing required to sample tags and primary sequences from both ends of a large fraction of fragmentation events is heavily dependent on library uniformity. Significant skewing of relative representation may require a correspondingly large amount of sequencing to overcome. It is therefore important that such skewing be minimized.

PARSE EX.1007.203

[00109] Through simulation, the N10, N50, and N90 span of chained read-pairs were determined based on the empirical size distribution of transposome fragmentation (Figure 3 above) and as a function of the fraction of transposase-derived fragments that are successfully sequenced (which is in turn a function of sequencing depth and the above-described performance parameters). As shown in Figure 7, contiguity rises sharply at efficiencies above 90%. At 95% efficiency, the N50 is 1.4 Kb and the N10 is 4.7 Kb. At 99% efficiency, the N50 is 8 Kb, and the N10 is 24 Kb. At 99.9% efficiency, the N50 is 71 Kb, and the N10 is 237 Kb.

[00110] An important aspect of this technology is that the *in silico* "joins" between independent read-pairs are almost completely independent of the primary sequence content (as would largely be the case with conventional *de novo* assembly, were it not confounded by the ubiquity of closely related sequences). Rather, joins are based on the shared barcode sequences which result from the synthetic transposons that are used to symmetrically tag fragmentation events. It is noted that 25 bp barcodes (which would only require a 25% increase in the amount of sequencing relative to a PE100 run) are unlikely to be identical by chance, even when sequencing millions of independent tags and allowing for a reasonable edit distance. Furthermore, the expected 9 bp overlaps between primary sequences can serve as a "verification key" for correct joins. Thus, the chance that coincidence or errors resulting in incorrect joins is very low.

[00111] This method may enable the equivalent of "strobe reads" (i.e., multiple sub-reads from a single, long contiguous fragment (Ritz et al. 2010)), while using a short-read technology. The gaps result when any given genomic fragment along the chain is too long to be spanned by single-end or paired-end sequencing. The frequency and length distribution of gaps is a function of the read-length of the short-read technology with which this method is integrated. For example, assuming that: 1) genomic fragments are interrogated by paired-end, 100 bp reads (PE100); 2) a terminal overlap of 20 bp is sufficient to merge read-pairs sequencing the same fragment from either end; 3) the fragmentation size-distribution shown in Figure 3 holds, then simulations show an average of 0.7 gaps per Kb, with gap sizes averaging 53 ± 48 bp (less than 5% of the overall scaffold length).

PARSE EX.1007.204

1.D. LoxP insertion via transposase followed by Cre recombination

[00112] In another embodiment, the bacterial transposase Tn5 may be used to insert a transposon containing the 34bp directional LoxP site flanked by inverted mosaic end (ME) sequences as well as an internal biotinylation and potentially alternate sequencing primers. Target insertion density is roughly one insertion event every 10 kilobases.

[00113] The resulting population of molecules has insertion events in the same order or in an inverted manner. Recombination with Cre recombinase will excise a 10kb circular stretch of DNA where two LoxP sites were inserted in tandem in the same orientation. Where tandem LoxP sites are inverted, the 10kb region will be inverted, yet the DNA will remain linear. Finally, inter-strand LoxP sites will recombine and swap strands which will also result in linear DNA.

[00114] Linear molecules may then be digested using a plasmid safe exonuclease, leaving behind the circularized DNA resulting from recombination between two tandem LoxP sites in the same orientation.

[00115] Circularized DNA may then be used for library preparation by any method and the DNA flanking the LoxP transposons can be enriched for by streptavidin bead pulldown. PCR followed by sequencing from either within the LoxP sites, or from the terminal ends of the molecules will yield ~10kb mate paired reads.

1.E. Y-Transposons

[00116] In another embodiment, a Y-adaptor approach (Figure 23) may be used in place of traditional transposase catalyzed adaptor insertion as a method for a library preparation where the resulting species are either A-B (50%), A-A (25%), or B-B (25%) where A and B are the two different adaptors and only 50% of the molecules are viable for sequencing.

[00117] In this case, Tn5 may be loaded with oligonucleotides complementary for the 19bp mosaic end recognition sequence along with an extension of complementarity to provide a higher melting temperature followed by non-complementary single stranded DNA (ssDNA) adaptor overhangs of A and B'. Transposition will result in one of the

PARSE EX.1007.205

adaptors (ME at the 3' end) being directly linked with the other remaining bound via hybridization.

[00118] Non-displacing polymerization and nick-repair may result in molecules where each insertion event can result in a viable sequencing amplicon.

[00119] An alternative embodiment involves a hairpin transposon containing a U (or other linker or targetable site for degradation or polymerase discontinuity) that links the ends of Y-adaptors to prevent loss of the other strand due to melting as in Figure 23.

1.F. Double-Bubble Barcode Transposons

[00120] In another embodiment, a synthetic transposon containing two degenerate barcodes (on each strand) and two sets of primer sites as well as several restriction sites, can be inserted by high frequency into genomic DNA as shown in the example below:

ES – SbfI/AsiSI – N1/N2 – Barcode – X/Y –NotI – X/Y – Barcode – N1/N2 – SbfI/AsiSI – ES

[00121] The resulting transposition and gap-repair followed by whole genome amplification (WGA) resolves the degenerate regions. Digestion using the outermost restriction sites (SbfI, AsiSI in the example above) followed by PCR using N1/N2 and overhanging flowcell primers will allow for a sequencing run to associate the two degenerate barcodes within each inserted transposon.

[00122] The other digestion in the middle of the transposon (NotI in the example above) and amplification and sequencing from the X/Y as well as N1/N2 gives the outer barcode sequences and intervening genomic DNA.

1.G. Subassembly with transposase inserted barcodes

[00123] In another embodiment, a discontinuous transposon may be inserted where each loaded DNA sequence is comprised of an outer flowcell primer, a degenerate barcode, an inner sequencing primer, and the double stranded transposase recognition sequence. The target insertion density may be every 1 to 2 kb.

PARSE EX.1007.206

[00124] After transposition, a degenerate sequence primer with a sequencing or flowcell primer overhang can be used to anneal to different positions along the molecule and extended back to the terminating transposase added sequence under dilute template or, more likely, emulsion conditions.

[00125] Sequencing will allow barcode association with every read that comes from the degenerate primer extension that occurred throughout the long molecule.

1.H. Mate-pair (ssDNA circularization) Transposase based library prep

[00126] In another embodiment, a standard, barcoded transposase-based library prep with a fragment size of approximately 1 to 2 kb, wherein size selection may be required, may be used to form a mate pair library.

[00127] The large fragment barcoded transposase based library prep will be subject to PCR using 5' phosphorylated flowcell (outermost) primers, in which one also has an internal biotin as well as a uracil near the 5' end.

[00128] The resulting PCR product will be circularized, followed by mechanical shearing. The fragmented DNA is then denatured and circularized in a single-stranded manner. The fragments containing the ends of the initial circularization are selected for using a streptavidin coated bead. The circles are then made linear by digestion at the uracil which will flip out the molecule. Sequencing allows for mate pair reads from the ends of the original library, also preserving the barcode.

1.I. Transposon modified fosmid or plasmid library pool sequencing

[00129] In another embodiment, continuous, synthetic transposons may be inserted into genomic DNA (gDNA), followed by gap repair. DNA is then sheared to 40kb (or roughly 5kb) and used to make a complex fosmid library (or plasmid) library respectively. This will allow for repetitive regions of the genome to be broken up by transposons that either have unique barcodes, or are identified by their unique insertion site into the repetitive region (Figure 27).

[00130] Briefly, synthetic, continuous transposons are inserted into genomic or high molecular weight DNA using *in vitro* transposition methods to a density between 100 and 1000 base pairs (bp). Transposons are either all the same, or contain unique

PARSE EX.1007.207

barcodes. Lesions 9 bp in length that result from the transposition mechanism are then repaired. Next, DNA is sheared to approximately ~40kb (or ~5kb) and a size selection is performed, followed by end-repair. Next, a complex fosmid (or plasmid) library is generated using the modified, sheared, and repaired DNA. Finally, fosmid (or plasmid) library pools are sequenced to provide phasing information as well as information regarding transposon insertions that will allow for differentiation between similar regions of the genome, using either unique barcodes and/or unique transposon insertion sites.

Example 2: Mid-Range Contiguity

2.A. Emulsion PCR with droplet-specific barcodes

[00131] Emulsion PCR is well established, but the methods below require droplets containing reagents including primers with droplet-specific barcodes. These reagents can be generated through emulsion PCR of common sequences flanking a degenerate subsequence, with recovery of products to micron scale beads (Figure 8) (Dressman et al. 2003). Specifically, large numbers of clonally amplified beads (each bearing a presumably unique barcode) may be generated by emulsion PCR with limiting dilution, followed by enrichment of "amplified" beads by hybridization (Shendure et al. 2005). These beads can be emulsified again for use in the below methods. Inclusion of a single clonally amplified bead per droplet, along with appropriate design of common sequences and emulsion PCR primers will result in the capture of barcoded amplicons to the beads themselves for convenient recovery.

2.B. Barcoding of "pre-transposed" HMW genomic DNA

[00132] In one embodiment (Figure 9), transposomes are loaded with adaptors containing the transposase recognition sequence with 5' ssDNA extensions of two different subsequences with complementary termini. This results in HMW genomic DNA densely interspersed with linked adaptor sequences. These "pre-transposed" molecules are then compartmentalized to emulsions with limiting dilution, using microfluidics to minimize shear and control size while maintaining a high throughput of droplet production (Zeng et al. 2010). Emulsion PCR, with primers bearing droplet-specific barcodes (2.A above), will amplify many fragments derived from the same HMW molecule within each droplet. Sequence reads derived from the same droplet will be

PARSE EX.1007.208

associated with the same barcode in the final library, thereby facilitating the *in silico* grouping and localized assembly of each progenitor 20 – 200 Kb molecule.

2.C. Barcoding of "pre-amplified" HMW genomic DNA

[00133] In another embodiment (Figure 10), HMW DNA is directly compartmentalized to emulsions, again using microfluidics to minimize shear, with reagents that support clonal, isothermal multiple displacement amplification (MDA) within droplets (Mazutis et al. 2009). These droplets will then be fused (with a relatively straightforward and cost-effective microfluidics device) with droplets containing standard transposomes as well as reagents for emulsion PCR, using primers bearing droplet-specific barcodes (2.A above). As with the previous embodiment described above, recovery and sequencing of the resulting library can interrogate both shotgun primary sequence and the barcode sequence on each molecule, with the expectation that reads sharing the same barcode derive from the same progenitor 20 – 200 Kb molecule.

[00134] This method may be used in transposome fragmentation followed by PCR within a single emulsion. When a "single-step" method is used to generate sequencing libraries from bacterial colonies, transposition may be performed followed by PCR with no cleanup step. In some aspects, the transposome reaction is diluted by addition of PCR reagents (Adey et al. 2010). Notably, in this method it is the PCR polymerase that facilitates the repair of the 9 bp lesion resulting from transposition by nick translation. At a minimum, these data illustrate that the MDA droplets can be fused with droplets supporting the transposome reaction, and these could subsequently be fused with larger droplets containing PCR reagents and barcoded primers.

[00135] The same effective end-results may be achieved exclusively with *in vitro* methods. Each of the methods described herein (2.B, 2.C) are dependent on capturing transposome fragmentation products within each emulsion droplet to a uniquely barcoded bead. In order for one of these approaches to be successful (for example, the approach described in 2.C.; "barcoding" of "pre-amplified" HMW genomic DNA"), both transposase-based fragmentation and polymerase-based extension must take place within the same emulsion compartment, i.e. within the same buffer. Initial experiments have been focused on this specific step, and are being conducted in non-emulsion

PARSE EX.1007.209

reaction volumes to facilitate optimization (schematic in Figure 19). Recently, this compatibility in selected buffers was demonstrated. In brief, a reaction volume was prepared containing 50 ng of genomic DNA in Nextera HMW buffer, dNTPs, adaptors 1 & 2, loaded transposase, and PCR polymerase. Adaptors 1 & 2 were designed to include both sequences complementary to the synthetic transposons, as well as unique sequences at their 5' ends (P1 & P2). The transposase+extension reaction was carried out at 55C for 5:00 to facilitate transposition, followed immediately by a single round of thermocycling to facilitate the nick translation and to append adaptors 1 & 2 (72C for 10:00, 95C for 0:30, 62C for 0:30, 72C for 10:00). Reactions were subjected to column-based cleanup and then used as template in a PCR using only outer primers P1/P2. The resulting distribution of amplicon sizes (Figure 20) is consistent with transposase-based fragmentation and polymerase extension taking place in the same buffer, albeit with limited insertion density. This reaction may be demonstrated it in the context of a water-in-oil emulsion, capture of extension products or beads loaded with barcoded oligonucleotides.

[00136] Mid-range contiguity information is likely sufficient to extensively support haplotype resolution in the resequencing of an individual human genome. To test this, a straightforward "short-cut" scheme was implemented by barcoding and sequencing complex pools of large-insert (fosmid) clones. Specifically, randomly sheared human genomic DNA (~35 Kb) was cloned from a single individual to yield a complex fosmid library (>2x10^6 clones). This library was then transformed to cultured *E. coli*. The resulting transformed E. *Coli* cultures were split into 115 fractions, and selected for transformants. The initial transformation was titrated to yield ~5,000 clones per pool. Given the uniform insert size of ~35 Kb, this corresponds to ~3% physical coverage of a diploid human genome per pool. Transposome fragmentation was then performed to generate a barcoded library from each of the 115 pools. This library was sequenced across 18 lanes on the Illumina GA2x for a total of 120 Gb of sequence (PE76 or PE101 + barcode). A shotgun library from this same individual was also sequenced across 7 lanes on the Illumina HiSeq for a total of 86 Gb of sequence (PE50), or 28-fold coverage of the haploid genome. The latter data alone yielded 3.6 million SNP and indel

PARSE EX.1007.210

calls, but as with all individual human genome sequences to date, these calls are blind to haplotype.

[00137] After deconvolving barcodes and mapping reads, the approximate boundaries of individual clones within each pool were easily identified by read-depth. A total of 538,009 clones (4678 ± 1229 per pool) for ~3x physical coverage of the diploid genome were identified. 98.6% of the genome was covered by 1+ clones, and 93.6% by 3+ clones. Long outgrowths of clone pools were avoided to minimize the impact of growth effects on representation. This was successful, as on average 82% of clones per pool had read depth within one order of magnitude. Because each pool only sparsely samples the genome as a whole, the rate of overlap, or 'clone-collision' within any given pool is low. Therefore, short reads derived from each pool overwhelmingly represent alleles from only one of the two homologous chromosomes at any given location. Haploid genotype calls from clones were assembled across all pools using a parsimony-maximization approach (Bansal & Bafna 2008). The resulting haplotype assembly covered 93% of ascertained heterozygous SNPs, with an N50 of 386 Kb. Of all RefSeq genes, 63% were entirely encompassed by a single phased haplotype block, while 75% were at least half encompassed by a single block.

[00138] This phased assembly was compared to HapMap predictions for this same individual (Figure 11). Within regions of exceptionally high LD (D'>0.90), a nearly perfect concordance with HapMap predictions was observed (>99.5% agreement). Because the sample chosen was not part of a trio, HapMap predictions rely upon LD between alleles to predict phase from genotype calls. Correspondingly, concordance was reduced to ~71% in more highly recombinogenic regions (D'<0.10), which includes the majority (66%) of pairwise SNP combinations. The haplotype-resolved resequencing of this genome is direct and experimental, and therefore completely orthogonal to population-based measures such as LD and allele frequency. Consequently, this trend likely reflects errors on the part of HapMap phasing (Lecy et al. 2007).

[00139] In contrast with population-based inferential methods, direct haplotyping allows for phasing of rare alleles and structural variants, including at complex, duplicated loci (Kidd et al. 2008). For instance, in these data, clones containing a

PARSE EX.1007.211

common inversion polymorphism on chromosome 7q11 were observed as well as clones containing a rare deletion polymorphism on chromosome 1p36. Similar approaches may be used to leverage the unambiguous assignment of short sequence reads to 20 – 200 Kb regions by the methods described herein. Whether relying on clones, or entirely *in vitro,* mid-range contiguity information facilitates the long-range haplotype resolution of individual human genomes. Further, mid-range contiguity information may also facilitate the *de novo* assembly of large, complex genomes.

2.D. Emulsion transposition with bead-immobilized transposomes

[00140] In another emboidmet, beads coated in a primer flanked, degenerate, monoclonal barcode oligonucleotide (or "oligo") terminating in the double stranded DNA (dsDNA) transposase recognition sequence and beginning with a flowcell primer may be emulsified with high molecular weight genomic DNA and free transposase. The bead-immobilized oligos and attack genomic DNA may be loaded within the emulsion the transposase. Resulting fragments are PCR ready and able to be sequenced along with their barcode. Barcode association can then be used to group reads that came from the same progenitor molecule.

[00141] This approach encompasses several variants. For exmple, many clonal copies of a barcoded oligo ending in the mosaic end sequence (ME) are immobilized at their 5' ends on each bead. These beads may be generated by emulsion PCR with 5'-biotinylated primers and a degenerate region, or alternatively a smaller set of barcoded oligos may be synthesized and immobilized to the beads. A short oligo comprising the reverse complement of ME (ME') is present in the emulsion mix to support transposase loading. Alternatively, the ME' may be annealed and loaded onto transposases prior to emulsification. Bead-bound oligos may be designed with an enzymatically cleavable moiety to allow the loaded transposomes to diffuse within the droplet.

2.E. Emulsion transposition and bead capture

[00142] In another embodiment, beads are coated by oligos with an internal, inverted base, thereby having two 3' ends. On the bead-distal 3' side of the inverted base is a primer site flanked, degenerate, monoclonal barcode, and a fixed adaptor sequence ("N1 prime"). These are emulsified with substrate (e.g., HMW gDNA) and

PARSE EX.1007.212

transposase pre-loaded with oligos 5'-N1-ME. Transposition then proceeds within each droplet, generating fragments covalently linked to the 5'-N1-ME sequence. The mixture is then heated, inactivating the transposase enzyme and denaturing the fragmented substrate. After slowly cooling, 5'-N1 flanked fragments generated by transposition anneal to the free ends of bead-bound oligos. Bead bound oligos are then extended using a thermostabile polymerase either present in each droplet, or after breaking the emulsion. Barcode association is then used to group reads originating from the same progenitor molecule.

[00143] In an altenrative approach, beads are coated in a primer flanked, degenerate, monoclonal barcode oligo. Then, a pool of random hexamers (DNA or LNA) having a 3'-blocking moiety is attached to the 3' end of each bead-immobilized oligos by ssDNA or RNA ligase. Beads, substrate (e.g., HMW gDNA) and pre-loaded transposomes are then emulsified. Transposition results in fragments with covalently attached 5' linkers. These fragments are denatured and allowed to anneal to the random 3' portion of the bead-bound oligos. The hybridized fragments are then extended into the barcode either by polymerase present in each droplet or by breaking the emulsion and adding polymerase. Barcode association is then used to group reads originating from the same progenitor molecule.

2.F. End capture of long molecules using transposase and emulsification

[00144] In another embodiment, long genomic DNA molecules with an adaptor B' ligated to the ends may be subject to transposon insertion of a bubble transposon in which inverted adaptor A sequences make up the bubble which is flanked by transposon recognition sequences. The molecules may then be emulsified where a portion of microreactions contain a large molecule, a bead coated in a monoclonal degenerate barcode terminating in adaptor B, and adaptor A. Performing PCR is then performed, which results in amplification of the outer most ends with the ligated B adaptors on the bead, appending the unique barcode.

[00145] After performing a subsequent PCR using the washed beads, the library may be sequenced and barcodes may be used to associate the two end sequences

PARSE EX.1007.213

from each of the ends of the original long molecule, effectively creating a jumping library of whatever size the original long molecules were.

2.G. T7 Promoter insertion via transposase

[00146] In another embodiment, transposomes are loaded with a bubble structure, flanked by a T7 terminator to one side and a T7 promoter to the other. This structure is integrated into a substrate (e.g., HMW gDNA) by bulk transposition at a density of at least one integration per kilobase. The resulting material is then emulsified with T7 RNA polymerase and with beads containing monoclonal degenerate barcodes flanked by priming sites and ending in sequence (X) complementary to the portion of the integrated bubble structure preceding the T7 terminator. *In vitro* transcription is carried out within each droplet, and the resulting RNA molecules, ending in X', hybridize to their bead-bound complementary sequences. Reverse transcription is then carried out to extend the bead-bound oligos, either within each droplet or after breaking the emulsion. Barcode association is then used to group reads originating from the same progenitor molecule.

2.H. Extension of emulsion PCR on adaptor bubble inserted high molecular weight fragments to allow for subassembly

[00147] In another embodiment, a transposon that forms a "bubble" may be inserted, wherein the bubble within the transposon includes two of the same adaptors in reverse orientation. Using the same adaptors in reverse orientation maintains the "bubble" structure. Bubbles may be inserted at a frequency of approximately 1,000 bp.

[00148] As shown in Figure 26, large fragments will then be emulsified where a portion of the microreactions will contain a single, long DNA fragment, a single bead coated in a primer bound via biotin at its 5' end and consisting of an outer primer, a degenerate barcode (monoclonal for each bead), and the complement to the adaptor inserted via transposition. Multiple displacing PCR (MDPCR) then generates many copies extending out from each adaptor insertion site.

[00149] Emulsions are then broken and beads are pulled out. Transposition with a second adaptor on a discontinuous transposon will occur at random distances away

PARSE EX.1007.214

from the bead for each copy of the amplified fragment. Removal of non-bead bound products and amplification will produce a library in which all amplicons from a large fragment can be associated with one another. The library also retains the ability to use the sequence acquired by sequencing genomic DNA from the original barcode adaptor as an anchor to associate reads where the paired read for all locally associated reads can be used for subassembly, as they arise from different secondary (post emulsion PCR) transposition events.

2.I. Clonally barcode-tailed, randomly primed amplification in nanoliter reactors

[00150] In another embodiment, beads are coated in primer sequences, or templates thereof, having a degenerate barcode monoclonal for each bead as well as a non-clonal, fully degenerate short kmer (k=5 to 9). These primers are released by excision of the immobilized DNA oligo from the bead, or alternatively by *in vitro* transcription of the immobilized DNA oligo into RNA primers. Oligos immobilized on the beads are designed such that the resulting DNA or RNA primers are structured as follows:

5'—[common1]—[clonal barcode]—[common2]—[random_k-mer]—3'OH

[00151] Bead-immobilized oligos may be prepared either by directly immobilizing these full sequences (e.g., in sets of 96 different barcodes) to beads, or by emulsion PCR.

[00152] Beads thus constructed are emulsified with: (i) substrate DNA (e.g., high molecular weight genomic DNA) at a target concentration of one substrate fragment per droplet, (ii) reagents for primer release/synthesis including, but not limited to, T7 RNAP and NTPs, any suitable restriction enzyme, or uracil N-glycoslyase and DNA glycosylase-lyase, and (iii) reagents to support DNA polymerization from the cleaved/synthesized primers (e.g., phi29 or Bst DNA polymerase, dNTPs).

[00153] Following their release or synthesis, primers anneal by random priming to sites throughout the substrate molecule. The included DNA polymerase extends the annealed primers along the template, resulting in multiple, randomly spaced dsDNA

PARSE EX.1007.215

fragments containing at the 5' end a tag clonal to a given droplet, and the 3' end sequence derived from various points along the substrate. In one aspect, the DNA poymerase has strong strand displacement activity (e.g., phi29 DNAP).

[00154] Following this DNA polymerization, the emulsion is broken. If RNA primers were used, the barcode encoded in each primer is reverse transcribed into DNA by methods known in the art. Finally, the resulting fragments are subjected to a standard library construction technique (transposase-based or otherwise) and amplified using a left primer consisting of common1 and a right primer corresponding to the adaptor added by transposition or ligation. Barcode association can then be used to group reads that came from the same progenitor molecule.

Example 3: Long-Range Contiguity

[00155] Methods for long-range contiguity, (e.g., 100 Kb – 10 Mb) by *in situ* transposition into very HMW genomic DNA on the surface of a sequencing flowcell are developed using spatial information (i.e. the relative coordinates at which physically linked sequencing templates are immobilized), as opposed to capturing contiguity information to barcodes as described above.

[00156] Such methods are accomplished by (1) exploiting *in situ* transposition to obtain paired-end reads from arbitrarily large single DNA molecules, and (2) developing related methods whereby multiple reads along the full length of arbitrarily large single DNA molecules are obtained.

[00157] In one approach (Figure 12b), HMW DNA molecules may be end-modified with an adaptor (FCA1), but are then hybridized to the flowcell without stretching. Long DNA molecules typically adopt a random coil configuration in solution. One end hybridizes, while the other end is spatially confined to an area proportional to the square root of the contour length of the template. This increases the probability that it will also hybridize at a close or nearly identical physical location (or a "constrained" physical location). The immobilized templates are then subjected to *in situ* transposition with transposomes loaded with hybrid adaptors corresponding to the second flow cell primer (FCA2), but also containing sequence corresponding to one of two possible sequencing primers (p1 or p2). After cluster PCR, approximately ~50% of the original templates will

PARSE EX.1007.216

likely produce two visibly overlapping or closely located clusters that each contain shotgun sequence derived from one or the other end of the very HMW molecule adjacent to either p1 or p2. Reads originating from p1/p2 with the same or closely located physical coordinates are highly likely to have been derived from the ends of the same very HMW molecule.

[00158] For this approach (Figure 12b), the molecules should have 3' single-stranded tails that are complementary to one of the flow-cell sequences (i.e. the cluster PCR primers). To achieve this, flow-cell adaptor A (or B) may be appended to the ends of HMW DNA molecules in solution, and then inserting adaptor B (or A) via *in situ* transposition. In practice, two different species of the adaptor are needed for one of these steps, i.e. A1 and A2 (or B1 and B2). This is because clusters derived from fragments at either end of any given HMW DNA molecule will be located in close proximity, with the potential to interfere with the sequencing of the other unless different sequencing primers are used. This can be achieved by using two different adaptors (i.e. A1 and A2 (or B1 and B2)) that both contain the flow-cell adaptor sequence (A (or B)) but also contain unique sequence at their 3' end to facilitate the design of distinct, non-cross-hybridizing sequencing primers. In one embodiment, A1 and A2 were added to the ends of HMW DNA in solution, and B via transposition. This scheme enables the sequencing of the HMW molecule ends (adjacent to A1 and A2), rather than the transposition junction (adjacent to B). The transposition junction necessarily includes the 19 bp mosaic end (ME) sequence, which complicates the design of two sequencing primers with distinct specificities. However, the alternative (sequencing through the 19 bp ME), would be unnecessarily wasteful.

[00159] It is noted that optical mapping is routinely used to analyze molecules as long as 1 Mb. The system described herein may be applied to molecules of similar lengths.

[00160] In another approach, optical sequencing on stretched single DNA molecules has been shown to be capable of yielding up to 3 bp of contiguous sequence information from multiple locations along the same molecule (Ramanathan et al. 2004). Since reads are generated directly from single molecules, issues of sample quantity and

PARSE EX.1007.217

PCR bias are largely avoided. However, in order for this approach to be practical in facilitating *de novo* genome assembly, read-lengths must be significantly improved.

[00161] Here, *in situ* transposition may be used to facilitate methods related to optical sequencing but with existing next-generation sequencing hardware, software, and reagents. In one approach (Figure 12a), a library of very HMW DNA molecules (10^5-10^7 bp) are end-modified with an adaptor (FCA1), hybridized to the surface of a primer-coated flowcell, and physically stretched using an electric field. While the field is still applied, a second adaptor is flushed into the flowcell and allowed to hybridize (similar to (Geiss et al. 2008)). This locks down the free end of every template and holds it in a stretched position. Transposomes pre-loaded with a second flowcell compatible adaptor (FCA2) can then be introduced to randomly fragment the stretched molecules while simultaneously inserting these adaptors. The majority of fragments will receive two FCA2 adaptors, except for the ends, which have both FCA1 and FCA2. Cluster PCR via these adaptors will only produce clusters at the ends of the stretched molecule. In this way, we obtain spatially co-linear clusters that are known to be derived from the same parent molecule and are related by the physical distance between the clusters.

3.A. Optical sequencing & *in situ* library construction

[00162] Optical mapping using restriction enzymes has been successful in generating long-range contiguity maps for genome assembly (Zhou et al. 2009; Zhou et al. 2007; Lin et al. 1999; Lim et al. 2001; Lai et al. 1999; Schwartz et al. 1993). However, this process is limited by false positive and negative cut sites due to star activity and inefficient cleavage, necessitating multiple optical maps from the same region to generate a consensus map. The non-uniform distribution of restriction enzyme recognition sites can also limit the amount of useful information derived from repetitive or low complexity regions.

[00163] As discussed above, the relatively short read lengths associated with the most cost-effective DNA sequencing technologies have limited the quality and completeness of *de novo* genome assembly as well as of human genome sequencing. There are currently few or no robust methods that capture mid-range and long-range

PARSE EX.1007.218

contiguity information at a throughput commensurate with the current scale of massively parallel sequencing. To address this limitation, an *in situ* library was constructed and optical sequencing was performed on the flow-cells of currently available next-generation sequencing platforms. This produced an efficient method of capturing both contiguity information and primary sequence with a single technology by generating >30,000 *E. coli* paired-end reads separated by 1, 2, or 3 kb using *in situ* library construction on standard Illumina flow-cells.

[00164] Surface-mediated bridge PCR performs poorly for inserts >= 1 kb which limits the Illumina platform's ability to generate native long paired end reads from high molecular weight (HMW) DNA. To circumvent this, HMW DNA molecules constrained to a specific size-range are end-modified with two flowcell-compatible adaptor sequences (FCA1 and FCA2), each of which contains one of two possible priming sequencing primers (p1 or p2). The templates are then hybridized to the flowcell surface under stationary flow, during which they typically adopt a random coil configuration. When one template end hybridizes, it spatially confines the other template end thereby increasing the probability that it will also hybridize in close physical proximity. The immobilized templates are subsequently subjected to *in situ* transposition with transposomes loaded with hybrid adaptors corresponding to the second flow cell adaptor (FCB1). Without a transposition event, each template molecule contains only one of the two required flowcell adaptors required to generate a cluster. For templates that are transposed, this process generates two low molecular weight (LMW) templates that are both capable of cluster formation and hybridized to the surface in close proximity. After bridge PCR amplification, 50% of the templates should produce two overlapping or closely located clusters that each contain shotgun sequence derived from one or the other end of the HMW molecule. p1 is then serially used to sequence one end and p2 to sequence the other end of the template, and reads originating from closely located physical coordinates are likely to have been derived from the ends of the same HMW parent molecule. In this way, the information provided by the spatial coordinates at which clusters are generated to infer long-range contiguity. In a similar way, HMW DNA molecules that are tethered at one end and stretched using flow or an electric field could

PARSE EX.1007.219

be transposed *in situ* with appropriate adapters to generate multiple co-linear clusters derived from the same parent molecule.

Materials and Methods

[00165] *Library synthesis*. Genomic DNA from *Eschericia coli* type B cells were obtained from USB (Part # 14380) and physically sheared for 30 sec on a Bioruptor (Diagenode). The DNA was then size selected on a 1% agarose gel run at 100V for 2 hours, purified (Qiagen QIAquick Gel Extraction Kit), and end-repaired (End-It, Epicentre). Hairpin adapters were self annealed and then blunt-ligated using Fast-Link Ligase (Epicentre) overnight. Unligated genomic DNA and adapters were removed with treatment by Exonuclease III (NEB) and VII (Epicentre). The molecules were then treated with Uracil-specific excision reagent (USER™) (NEB) to generate single-stranded flow cell complementary 3' tails.

[00166] *Transposome loading*. Synthetic DNA oligonucleotides containing transposase mosiac, primer sites, and flowcell adapter sequence were obtained from IDT. Adapters were annealed and loaded on the transposase (Tn5, Epicentre) by mixing and incubating at room temperature for 20 minutes.

[00167] *In situ flowcell library construction and sequencing*. A custom cluster generation protocol was written to accommodate template and transposome loading on a standard Illumina Cluster Station. The flowcell was first primed with hybridization buffer and then heated to 96°C at rate of 1°C/s. At 96°C a standard Illumina sequencing library was loaded into a separate lane as a control while the other seven lanes received hybridization buffer. After a 2 min. incubation, the temperature was lowered to 65 °C at 0.05 °C/s to hybridize the control library. At this point, the tubing on the manifold for the control lane was removed on both the input and output sides of the flowcell. The *E. coli* libraries were added to each lane at 15 µL/min for 2.5 minutes, followed by slowly cooling the flowcell to 40 °C at 0.02 °C/sec. After a 5 min. incubation, the flowcell was heated to 55°C at 1 °C/s. Loaded transposomes were then added to the lanes containing *E. coli* at 15 µL/min. The flowcell was incubated at 55°C for 5 minutes to allow transposition to take place and then cooled to 40°C. A new manifold was then installed on the cluster station and Illumina wash/amplification buffer was injected

PARSE EX.1007.220

across the entire flowcell. First strand synthesis was performed at 65°C for 5 minutes and 74°C for 5 minutes using library-specific DNA polymerases. Standard human control libraries were than hybridized to each *E. coli* lane as per the manufacturer's protocol. Clusters were generated with 35 cycles of bridge amplification. Two separate single end 36 bp (SE36) reads were obtained on an Illumina Genome Analyzer Iix with RTA 1.8 and SBS v5 as per the manufacturer's protocol.

[00168] *Data collection and analysis*. The X-Y coordinates of every cluster from read 1 and read 2 were extracted from the fastq files using a custom Perl script. This data was used to calculate the image offsets using the normxcorr2 function in MATLAB and the X-Y coordinates for read 2 were corrected accordingly. Reads were then mapped separately to the *E. coli* genome using the Burrows-Wheeler Aligner (BWA) and the identities of neighboring clusters between read 1 and read 2 were determined using a custom Perl script.

Results

[00169] *In vitro and in situ library construction and sequencing*. An efficient approach for generating HMW DNA libraries containing single-stranded flowcell compatible 3'-tails is provided herein. Briefly, genomic DNA from *Eschericia coli* was physically sheared, size selected for 1, 2 or 3 kb size molecules, purified, and end-repaired. Hairpin adapters containing three uracil bases near the loop of the hairpin (Figure 31a) were self annealed and then blunt ligated to the size-selected libraries. Unligated genomic DNA and adapters were removed with treatment by Exonuclease III and VII to yield an enriched population of molecules with hairpin adapters on both ends. The molecules were then treated with USER™ to open the hairpin loop and release single-stranded flow cell complementary 3' tails. Both ends of the molecules were then hybridized to standard Illumina flowcell surfaces using a slightly modified thermal cycling protocol. Tn5 transposase loaded with flowcell–compatible adapters was added to the flowcell to randomly fragment and add adapters to the HMW molecules thereby generating LMW sequencing-ready templates (Figure 31b). Each *E. coli* library was pooled with a human control library, loaded onto a separate lane, and two separate single-end 36 bp reads (SE36) were obtained on an Illumina GAIIx.

PARSE EX.1007.221

[00170] *Reconstructing contiguity information.* Table 1 below illustrates the distribution of mapping reads for the 1, 2 and 3 kb libraries constructed as described above.

Table 1

	1 kb no filter	1 kb >= Q30	2 kb no filter	2 kb >=Q30	3 kb no filter	3 kb >=Q30
E. coli	4,532,112	3,428,616 (76%)	3,668,061	2,667,329 (73%)	2,340,128	1,523,035 (65%)
human	155,966	97,328 (62%)	794,123	504,299 (64%)	5,370,959	4,883,197 (91%)
adaptor/mosiac	44,189	23,563 (53%)	34,801	16,581 (47%)	9,337	5,037 (54%)
unmapping	6,269,729	780,191 (12%)	5,930,170	565,338 (10%)	2,755,611	215,931 (8%)
total	11,001,996	4,329,698 (39%)	10,427,155	3,753,547 (36%)	10,476,035	6,627,200 (63%)

[00171] An average of 3.5M reads mapping to *E. coli* were obtained in each of the three lanes (Table 1). The X-Y coordinates of the clusters in every tile were used to calculate the spatial offset between read 1 and read 2 X-Y coordinates. For each cluster mapped to *E. coli* in read 1 and read 2, its nearest physical *E. coli* mapping neighbor within 1.5 µm was identified within the same read and the mapping distances of all pairs were numerically ordered (Figure 32a and b). Table 2 below shows the nearest neighbor cluster pair data.

Table 2. Nearest neighbor cluster pair data for the 1, 2 and 3 kb libraries when E. coli reads are compared against E. coli reads. The expected size ranges were set at 800-1200, 1500-2300 and 2500-3500 bp, respectively.

	Ref. read	Pairing read	NN pairs <1.5 µm <4000 bp	+ within expected mapping distance	+ reads have opposite orientation
1 kb	1	1	4,952	1,206	1,060
2 kb			5,820	3,402	3,236
3 kb			3,464	2,424	2,334
1 kb	2	2	5,426	766	602
2 kb			2,276	870	794
3 kb			2,704	1,710	1,612
1 kb	1	2	33,393	25,708	25,502
2 kb			36,656	32,653	32,457
3 kb			39,743	37,916	37,769
1 kb	2	1	33,256	25,305	25,117
2 kb			35,686	31,643	31,466
3 kb			39,204	37,351	37,196

PARSE EX.1007.222

1 kb	1	1+2	38,256	26,894	26,544
2 kb			42,242	35,885	35,525
3 kb			43,097	40,272	40,036
1 kb	2	2+1	38,597	26,894	25,708
2 kb			37,841	32,438	32,192
3 kb			41,761	38,950	38,701
1 kb	1+2	2+1	29,676	23,028	22,863
2 kb	mutually exclusive		33,064	29,505	29,350
3 kb			35,701	34,082	33,946

[00172] Between 766-3,402 cluster pairs with the expected mapping distance were observed for each library (Figure 32a and b). A low number of cluster pairs were seen within a single read because clusters often physically overlap on the surface and Illumina's image analysis software is unable to distinguish them. Plotting mapping distance as a function of physical separation (Figure 34a) revealed the default lower limit of resolution between two clusters in a single read to be ~0.94 μm.

[00173] The nearest-neighbor search was repeated by looking for cluster pairs <1.5 μm between reads 1 and 2. Up to 37,916 distinct cluster pairs were identified within the expected mapping separation range (Table 2; Figure 33 and Figure 32c and d). Of these, over 99% were cluster pairs that gave reads on opposite template strands going in the opposite direction, which is the is expected orientation based on the design of the *in situ* library preparation. With this approach of serially obtaining paired reads, cluster pairs closer than 0.94 μm were clearly demarcated, including some that were almost completely overlapping (Figure 34b). The mean mapping separation for the cluster pair libraries was 946 bp, 1,770 bp, and 2,995 bp for the 1, 2, and 3 kb libraries, respectively (Figure 34b, top histogram). The 2kb library was likely a little low due to a wider size selection.

[00174] Separation distances were calculated based on a freely-jointed chain model of DNA tethered to a surface. Using a freely-jointed chain model, the free-space distribution function for the end-to-end vector of a DNA molecule is Gaussian. It is described by the equation:

$$G(\vec{r}, \vec{r_0}) = C_0 e^{\left(-\frac{3(\vec{r}-\vec{r_0})^2}{2bL}\right)} \quad (1)$$

PARSE EX.1007.223

where L is the contour length, b is the Kuhn length (twice the persistence length), and C_0 is a normalization constant. In the case where we have a surface at $z = 0$ and the molecule starting at [0, 0, z_0], the distribution function becomes

$$G_{surf}(\vec{r}, [0,0,z_0]) = C_0' \left(G(\vec{r}, [0,0,z_0]) - G(\vec{r}, [0,0,-z_0]) \right) \quad (2)$$

where C_0' is a new normalization constant and the second term repesents the entropic repulsion from the surface. As z_0 -> 0, the difference becomes a derivative which gives

$$G_{surf}(\vec{r}, [0,0,0]) = C_0'' r_z G(\vec{r}, [0,0,0]) \quad (3)$$

where C_0'' is a normalization constant. Plots of G_{surf} for the x,y and z components of $\vec{r}$ are shown in Figure 39a and 39b.

[00175] The mode physical cluster separation for the 1 kb pairs was 0.44 μm and for the 2 kb and 3 kb pairs it was 0.67 μm, with the tail of the distribution showing some cluster pairs separated by > 1.0 μm. These observed physical separation distances between Infinipair reads are was significantly larger (3-4 fold) than expected (Figure 39a). For example, the mean physical distance between 3kb cluster pairs was ~1000 nm which roughly corresponds to the contour length of the molecule. At least two possible explanations were contemplated for this discrepancy: 1) the image offsets are slightly off thereby giving rise to overestimates of the true physical distances, or 2) the large separation distances arise as an artifact during cluster formation. To verify the offsets, histograms showing the distribution of angles between every cluster pair were generated and the cumulative direction vector for all pairs was calculated. One would expect a random distribution of angles between pairs if the images are properly aligned and a net zero vector sum; a bias towards a subset of angles within a given tile or a non-zero vector sum suggests the offsets are slightly off. This did not appear to be a significant source of error. Therefore, these large separation distances arise as an artifact during cluster formation. Therefore, this observed discrepancy arises due to the way in which the clusters were generated on the flowcell (Figure 39c). If two cluster-capable molecules are hybridized within 50-100 nm of each other, there will be a highly localized depletion of available adapters between the templates during the initial cycles

PARSE EX.1007.224

of bridge PCR. This effectively forces the clusters to grow away from each other. As a result, the X-Y coordinate of each cluster will not accurately reflect the X-Y coordinate of the initial seed templates.

[00176] Using read 1 as a reference, the closest nearest neighbor was screened for from either read 1 and read 2 (Figure 33). For pairs within the expected mapping distance and in the correct orientation, fewer than 1% had a different nearest neighbor in the combined dataset, and this observation remained true when using read 2 as the reference. Applying a more restrictive filter that requires mutual exclusivity (i.e., the nearest neighbor of cluster A is B and that of B is A) reduces the number of candidate pairs by up to 10% but does not yield any significant gain in sensitivity. It is also noted that as the library size increases, a greater fraction of the total cluster pairs give rise to pairs within the target size range with reads in the correct orientation. This may be due to steric effects whereby larger molecules occupy larger volumes, and thereby prevent other molecules from hybridizing to the surface nearby.

[00177] Although the number of related cluster pairs represented only accounts for approximately 1% of the total reads mapping to *E. coli*, it demonstrates that *in situ* transposition and library preparation is technically feasible. At least two factors may contribute to the low efficiency: 1) a low probability for DNA to adopt the appropriate conformation to favor both ends annealing to the surface; and 2) transposon insertion in close proximity to the ligated adaptor sequence. The 3D probability distribution for the end-to-end vector of a DNA molecule with one end tethered to a surface indicates that the free end has a much higher probability of being far away from the surface than close to it. This problem is exacerbated with increasing DNA length. When only one end hybridizes and the molecule undergoes transposition, it generates a singleton read and not have a related nearest neighbor. Further, factor (2) is evidenced bythe finding of 9,294 nearest-neighbor cluster pairs where read 1 mapped to *E.coli* and read 1 mapped to transposase mosaic and/or flowcell adaptor sequence. Fine tuning of the transposase concentration and incubation time may help improve this but it may be difficult to completely eliminate it.

PARSE EX.1007.225

[00178] Surprisingly, the majority of reads for all three libraries did not map to human, *E. coli*, or adapter sequences (Table 1). The average quality score for the unmapping reads was typically low: only 8% had average raw quality scores > 30 and 69% had the lowest possible average raw quality score of 2 (Figure 37a). When all reads were considered, and not just those mapping to *E. coli* for nearest-neighbor proximity, 15.7% of the nearest neighbor pairs had one read mapping to *E. coli* and one unmapping read, and 6.8% had both mapped to *E. coli*. For the pairs that had one unmapping read, only 6% of them had an unmapping read with an average raw quality score > 30 and 78% had the lowest possible raw quality score (Figure 37b). Although the source of these unmapping reads is not clear, they can largely be filtered out based solely on quality score alone.

[00179] *In situ stretching and tagging of HMW molecules*. In an effort to improve the hybridization efficiency and explore further applications of this sequencing paradigm, *in situ* stretching and fragmentation of HMW molecules was successfully performed within Illumina flow cells. Briefly, flowcells were cleaned using Piranha solution, treated with 2% 3-aminopropyltriethoxysilane (APTES), and loaded with JOJO-1 stained lambda DNA. The flowcell was then loaded with 6M KCl and an electric field of 15V/cm was applied at the input and output ports for 90 sec. Surfaces were imaged directly on an Illumina GA2 sequencer (Figure 35a) to demonstrate that the ends of single 48.5 Kb molecules can be physically stretched over ~40 pixels of imaging space. Surfaces were then treated *in situ* with transposome and re-imaged (Figure 35b). Individual molecules were clearly fragmented in multiple locations, demonstrating the enzyme's ability to maintain high activity levels even on a surface-immobilized template. It should be straightforward to build on these methods to incorporate flowing in the "lock-down" bridge prior to fragmentation on native flowcells, so that clusters may be generated at the ends of long templates.

[00180] Using the 3 kb *E. coli* library described above, *in situ* stretching and sequencing of the ends of stretched molecules was also successfully performed within Illumina flow cells (Figure 40a). Template libraries were loaded into a flow cell at 75°C and the chamber was slowly cooled at 0.1°C/s to 55°C. Next, hybridization buffer containing 5X SSC and 200 mM KCl was flowed into the chamber and a 28 V/cm

PARSE EX.1007.226

electric field was applied for either 0 or 2 seconds. Wash buffer was then flushed through the chamber prior to *in situ* transposition and sequencing. In the absence of an applied electric field, the angles between clusters in a pair were randomly distributed and not correlated with the distance between the clusters. For cluster pairs that were separated by at least 45 tenths of pixels (~1.6 μm), 46% had angles (Figure 40b) between $-\pi/4$ and $\pi/4$ with respect to the axis of current flow in the chamber (Figure 41a). However, in the presence of the electric field, 78% of cluster pairs separated by > 45 tenths of pixels had angles within this range (Figure 41b). This is strongly indicative that these pairs had at least one end of the molecule hybridized at the time the field was applied, at which point the other end was stretched by the electric field before it hybridized to the surface. These results demonstrate that *in situ* stretching and sequencing of HMW DNA can be accomplished within native flowcells.

Discussion

[00181] Diverse technologies currently exist for determining contiguity information on a variety of length scales, including optical mapping (Schwartz et al. 1993), stretching single molecules in nanochannels (Riehn et al. 20057), single chromosome sorting (Fan et al. 2011), long-read single molecule sequencing (Eid et al. 2009), large insert cloning (Kitzman et al. 2011), and transmission electron microscopy. However, all of these technologies remain prohibitive for widescale use due to capital equipment costs or the expertise required for implementation. In the experiments described above, it was successfully demonstrated that *in situ* library preparation of HMW DNA molecules enables the capture of long-range sequence information up to 3 kb apart on an existing sequencing platform. The method described herein may overcome these limitations by taking advantage of existing sequencing hardware and single-step enzyme-based *in situ* library preparation. Further, the methods described have shown that paired-end sequencing can be accomplished without circularization.

[00182] There are at least four factors that affect the generation of related nearest neighbor clusters: 1) the production of a HMW library with uniform single-stranded flowcell compatible 3' adapters, 2) the hybridization of both ends to the flowcell surface, 3) the uniform and nondestructive *in situ* transposition of bridged molecules, and 4) the

PARSE EX.1007.227

generation of clusters that are largely overlapping. Control experiments suggest that the aforementioned approach using hairpin adapters followed by Exo III/VII treatment is highly effective at eliminating any library molecules that do not have two hairpin adapters. Additional control experiments have shown that USER™ treatment is also very efficient at uracil excision for making adapters single stranded, suggesting that the initial library construction is robust. Although it may be argued that having both ends of a molecule be situated near the surface is unfavored due to entropic arguments, it is more favorable than the circularization of a same-length single molecule due to the fact that each end can hybridize to any one of thousands of flowcell adapters. There may also be ways to force both ends to be closer to the surface to improve the hybridization efficiency, such as with the use of tethered magnetic beads or an electric field. For the *in situ* transposition step, a range of transposase concentrations and incubation times were used to identify the optimal balance between too little activity and too much activity, both of which result in a failure to generate clusters.

[00183] The effect on sequence quality of intentionally generating clusters that are largely overlapping is harder to interrogate. For example, it may be that when there are two cluster-ready templates hybridized on the surface in close proximity that one of them will out-compete the other during bridge amplification, as often happens during conventional bulk solution PCR. This can be due to differences in sequence composition, melting temperature, length, and the stochasticity of polymerase binding events. In the method described herein, control of the final length of the related nearby templates has been limited after transposition (one could be 200 bp and the other could be 800 bp). While it could a concern that the clusters are too close together, this does not appear to be the case here. In such a case, the quality scores would decrease with nearest neighbor cluster distance due to having fewer numbers of molecules within either cluster and/or the potential for mixed reads. Here, however, quality scores for read 2 were generally better than for read 1 and there does not appear to be a correlation between average quality score and nearest neighbor cluster distance (**Figure 38**).

[00184] Although the high background currently makes the approach impractical to use for *de novo* assembly, further improvements to the library preparation and *in situ*

PARSE EX.1007.228

transposition methods mal lead to a concomitant improvement in signal to noise. Ultimately *in situ* library preparation methods may enable the generation of reads whose physical relationship to one another on a flow cell is correlated with genomic distance, enabling the routine optical sequencing of multiple, ordered reads from many single HMW molecules as described below.

3.B. Achieving multiple *in situ* reads per single DNA molecule

[00185] Stretching prior to *in situ* transposition offers a number of advantages over the random-coil method. First, with stretching, the physical distance between co-linear clusters is expected to be directly proportional to the distance between reads, rather than proportional to the square root, thereby providing greater resolution. Second, stretching is more readily amenable to the second goal of this aim, which is to generate large numbers of independent reads along the full length of arbitrarily long single DNA molecules. In one scheme, diagrammed in Figure 15, stretched single molecules are subjected to *in vitro* transposition with synthetic, continuous transposons, containing the 19 bp ME sequences that are connected by sequence that includes a single-stranded bubble. This is similar to the first steps of the strategy in Figure 4, except that rather than degenerate sequences, each arm of the bubble is corresponds to the forward or reverse sequences of the primers that coat the flowcell. Once these synthetic, continuous transposons are inserted to high density (every 35 to 600 bp), the highly interspersed single molecules are stretched on the flowcell with current (without removal of the transposome complexes, such that there is no need for repair of the 9 bp lesions). Assuming a modest efficiency of hybridization and cluster PCR initiation from each bubble, this will likely yield multiple sequencing reads along the length of each stretched molecule.

[00186] The full area of each lane of the flow-cells that are used according to the long-range contiguity method described above is 245,760 pixels in the dimension of flow (2.5 cM) and 3,776 pixels in the orthogonal dimension. As lambda DNA (48.5 Kb) was stretched to ~30 pixels, as many as 400 x 1 Mb molecules may be stretched end-to-end along the full lane. At ~1/20th density, a single lane would be sufficient to support 14x physical coverage of a diploid human genome. For data analysis, published algorithms

PARSE EX.1007.229

for optical mapping (Zhou et al. 2009; Zhou et al. 2007; Lin et al. 1999; Lim et al. 2001; Lai et al. 1999; Schwartz et al. 1993) may be used as well as previous experience in image analysis for sequencing applications (Shendure et al. 2005; Mitra et al. 2003). Such analysis may be performed directly from images, or alternatively from platform-generated sequence reads annotated with position-of-origin information. The positional information can be correlated with sequence data generated from co-linear or overlapping clusters.

[00187] The resulting data should be similar to that generated by optical maps, but has the following advantages:

1) Sequence reads represent data points that have much greater information content than restriction enzyme sites for both *de novo* assembly and haplotype resolution;

2) Issues that impact optical mapping such as restriction enzyme star activity and incomplete digestion will not occur with this approach; and

3) The positions of data points along the length of stretched DNA molecules will be random, rather than dependent on the restriction enzyme cut-site distribution.

[00188] The effect of extra templates hybridizing near to the stretched or coiled templates (which can confound the interpretation of the physical coordinates) can be mitigated by size-restricting the single DNA molecule populations and/or tuning template concentrations. However, these methods can be implemented without major sacrifices to cluster density, as the sequencing should be as dense as would normally be the case on the same platform.

[00189] To generate HMW DNA with single-stranded tails corresponding to A1 and A2 appended to each end, two strategies are pursued. In the first (Figure 21, left), genomic DNA is physically sheared (e.g. with HydroShear), and then end-repaired, A-tailed, and ligated to adaptor sequences corresponding to A1 and A2. The library is then PCR amplified using primers corresponding to A1 and A2 in which all thymine bases are replaced by uracil. Post-PCR treatment with USER™ is expected to yield the desired 3'

PARSE EX.1007.230

single-stranded, flow-cell compatible tails flanking the double-stranded HMW DNA molecule. One advantage of this approach is that self-complementarity of end-sequences is expected to limit accumulation of A1-A1 and A2-A2 products, whereas a disadvantage is that it may not be practical for HMW fragment sizes that are largely incompatible with PCR, i.e. >10 Kb. As an alternative, libraries were generated in which A1 and A2 adaptors containing single-stranded, flow-cell compatible tails are directly ligated to blunt-end or restriction digested HMW genomic DNA (Figure 21, right). An advantage of this method is that it is independent of the length of the HMW molecules. However, only 50% of products will be A1-A2 flanked (with the remainder either A1-A1 or A2-A2).

[00190] Both library preparation methods shown in Figure 21 were applied to generate adaptor A (A1/A2) flanked shotgun HMW molecules from *E. coli* genomic DNA. Then, transposase loaded with synthetic transposons that include adaptor B for *in situ* fragmentation on the flow-cell surface (i.e. the method shown in Figure 12b) was used.

[00191] The results showed that *in situ* transposition may be successfully performed to introduce adaptor B into shotgun, A1/A2 adaptor-flanked HMW genomic DNA molecules that are pre-hybridized to the flow-cell. This is an improvement over the experiment described above and in Figure 13, which involved only a single puc19 fragment. Mapping density across the *E. coli* genome of a representative experiment is shown in Figure 22. The distribution is largely uniform, indicating that introduction of a cluster PCR compatible adaptor via *in situ* transposition does not result in overt biases in genomic representation.

[00192] Further, the conversion of library molecules into useful sequencing templates is currently quite inefficient. The data shown in Figure 22 was from a single Illumina GA2x lane loaded with 10x the amount of usual template, but generated 100-fold fewer clusters than expected. Several reasons may explain this inefficiency, including: (a) *Inefficient generation of properly tailed molecules:* these approaches (Figure 21) may be significantly less than 100% efficient in their conversion of target material to appropriately adapted molecules, and require further optimization; (b)

PARSE EX.1007.231

Inefficient capture of 3' tailed dsDNA molecules to flow-cell primers: It is possible that additional manipulations of the flow-cell prior to cluster PCR (e.g. the transposase reaction; a wash including SDS to remove transposase, etc.) remove a substantial fraction of library molecules. (c) *Failure of first strand synthesis on the flow-cell:* Phusion DNA polymerase, which is normally used for first strand synthesis on the Illumina platform, has a very low strand displacement activity. Strand displacement during this first cycle is required for the method but not for conventional sequencing on this platform. Alternatives have shown, for example, that Bst polymerase can be substituted for Phusion for first strand synthesis on the flow-cell. (d) *Transposase loading and/or in situ transposition is inefficient:* Even if the molecules are hybridizing properly and first strand synthesis is successful, it won't form a cluster pair unless it receives at least one transposase insertion relatively proximal to an end.

[00193] Finally, although the data represented in Figure 22 represents over 200,000 reads from a single lane, only a negligible fraction of these reads came from clusters that had a "paired read" from a neighboring cluster (as in Figure 13). This problem may be related to the lower than expected densities of cluster formation (i.e. inefficient generation of molecules with proper tails at both ends, inefficient *in situ* transposition). Alternatively, this may be consequent to the relative rigidity of double-stranded DNA limiting both ends of a molecule with single-stranded tails from hybridizing to the surface.

3.C. Transposase followed by ssDNA circularization and multiple displacing, branching RCA

[00194] In another embodiment, as shown in Figure 24, transposase may be loaded with discontinuous oligos terminating in 5' phosphorylated A', followed by B and the dsDNA transposase recognition sequence. Transposition followed by denaturation will result in ssDNA terminating in a 5' phosphate, A', B, ME and then genomic DNA. Next, ssDNA circularization may be performed and then rolling-circle amplification (RCA) using flowcell bound A and B primers will result in multiple displacing branching rolling circle amplification and polony (i.e., polymerase colony) formation.

PARSE EX.1007.232

[00195] In an alternative embodiment, fosmids may be used in place of ssDNA which may result in long-range amplification and may allow production of 40kb mate paired libraries.

3.D. Ordered transposition using long ssDNA backbones for DNA nanoball formation or barcode association

[00196] In another embodiment, a circularized ssDNA template may be prepared using four primers (A, B, C, and D) interspersed with ~100bp of filler DNA sequence and circularized (dsDNA sticky-end circularization, followed by selective digestion of one strand). Rolling-circle amplification (RCA) then results in long ssDNA molecules of repeating A, B, C, and D primer sites with intervening filler DNA sequences.

[00197] Next, a set of four differently loaded transposase complexes may be pooled where the first has complementarity to the A sequence and includes a mid-way cut site, and the other three are complementary to the B, C, and D sequences. Transpositon into genomic DNA likely allows for partial or complete insertions occurring in the A, B, C, and D order. After gap repair, the A restriction sites may be digested and the molecules circularized which results in circularized molecules of A, gDNA, B, gDNA, C, gDNA, D, gDNA. These molecules may then be used as templates in RCA that will generate DNA nanoballs containing 4 adaptor sites.

[00198] In an alternative embodiment, the original backbone template may be comprised of an adaptor flanked degenerate barcode with ~100bp of filler sequence which is circularized, denatured, and subjected to RCA. The resulting backbone includes many repeats of the original template in succession. The transposase complexes are loaded with oligos that terminate in sequence complementary to the adaptors that flank the barcode where the transposase adaptors will anneal in a padlock formation. Gap repair of the degenerate region allows each transposome bound to any given backbone to have the same barcode. Transposition results in adjacent transposition events, likely occurring from transposomes of the same barcode, thereby allowing association of numerous reads with one original large progenitor molecule.

PARSE EX.1007.233

3.E. Direct sequencing of transposon bubbles containing flowcell primers

[00199] In another embodiment, a transposon that forms a "bubble" such as those described above may be inserted, wherein the bubble within the transposon includes primers complementary to flowcell bound primers as the bubble adaptors.

[00200] After insertion and subsequent gap repair, these long molecules can be directly hybridized to the flowcell either with or without stretching. Some portion of tandem transposons will be A and B' or A' and B which will be able to form clusters via standard bridge PCR methods. This will allow clusters originating proximal to one another will likely have arisen from the same high molecular weight progenitor molecule.

[00201] Alternatively, long molecules of known (to allow for an expected distance) or unknown length can have adaptors ligated to both ends containing a 5' overhang complementary to one of the flowcell primers. Transposition of a bubble transposon containing the other flowcell primer followed by gap repair will result in a molecule terminating in the complement to one flowcell primer and interspersed with the other. Hybridization to the flowcell with or without stretching will allow for the ends of the molecule to anneal. An initial displacing extension will copy through the transposon inserted second adaptor and produce the reverse complement. Subsequently, standard bridge PCR can be performed and after sequencing a proportion of proximal clusters will have arisen from the terminal ends of the original long molecule.

3.F. Transposomes assembled on the flowcell

[00202] In another embodimet, a flowcell is reprogrammed to include (1) oligos terminating in the transposase recognition sequence, or (2) bridge oligos that are hybridized that terminate in the transposase recognition sequence. The transposase is then added to the loading buffer and allowed to load the flowcell bound oligos.

[00203] Genomic DNA is then added to the transposase reaction buffer and wherever a molecule comes in contact with the flowcell, the immobilized transposase will attack at multiple positions along its length. After an initial extension, bridge PCR may be performed on the resulting fragments. Sequencing results in a portion of proximal clusters having arisen from the same large progenitor molecule.

PARSE EX.1007.234

[00204] In an alternative embodiment, long molecules may be added in which adaptors have been ligated containing one of the primer sequences that is not part of the flowcell-immobilized transposomes. Post transposition, one strand may be denatured and removed and the other is able to form clusters. A portion of resulting proximal clusters originate from the ends of the same large progenitor molecule.

Example 4: Low-input transposase library preparation for bisulfite sequencing

[00205] As described above, a transposase-based *in vitro* shotgun library construction method ("tagmentation") that allows for construction of sequencing libraries from greatly reduced amounts of DNA (Figure 36a) (Adey et al. 2010). Briefly, the method utilizes a hyperactive derivative of the Tn5 transposase loaded with discontinuous synthetic oligonucleotides to simultaneously fragment and append adaptors to genomic DNA. The resulting products are subjected to PCR amplification followed by high-throughput sequencing. The increased efficiency of genomic DNA conversion to viable amplicons and the greatly reduced number of steps allows the construction of low-bias, highly complex libraries from less than 50 nanograms of genomic DNA.

[00206] An approach, referred to herein as tn5mC-seq, that retains the advantages of transposase-based library preparation in the context of whole-genome bisulfite sequencing is described herein. Because the target of the transposition reaction is double-stranded DNA, whereas bisulfite treatment yields single stranded DNA, the method was extensively modified such that the tagmentation reaction could take place prior to bisulfite treatment (Figure 36b). First, the adaptors to be incorporated were methylated at all cytosine residues to maintain cytosine identity during bisulfite treatment, with the exception of the 19 base-pair transposase recognition sequence (in order to minimize differential binding during transposome assembly). Second, an oligonucleotide replacement scheme (Grunenwald et al. 2011) was utilized to ensure that each strand would have adaptors covalently attached to both ends of the molecule. Specifically, this entails initial transposition with a single adaptor in which the double-stranded transposase recognition sequence is truncated to 16 base-pairs (Tm = 36°C), thereby facilitating its post-incorporation removal by

PARSE EX.1007.235

denaturation. A second adaptor is then annealed and the gap repaired, resulting in each strand being covalently flanked by both a 3' and 5' adaptor. The fragmented, adapted, double-stranded genomic DNA is then subjected to standard bisulfite treatment for the conversion of unmethylated cytosine to uracil. This yields single-stranded, converted DNA that is PCR-amplified and sequenced.

Materials and methods

[00207] *tn5mC-seq library construction and sequencing.* Transposome complexes were generated by incubating 2.5µl of 10µM tn5mC-A1 (tn5mC-A1top: 5'- GAT [5mC] TA [5mC] A[5mC] G [5mC] [5mC] T [5mC] [5mC] [5mC] T [5mC] G [5mC] G [5mC] [5mC] AT [5mC] AGA GAT GTG TAT AAG AGA CAG -3', IDT (SEQ ID NO:1), annealed to tn5mC-A1bot: 5'-[Phos]- CTG TCT CTT ATA CAC A -3', IDT (SEQ ID NO:2), by incubating 10µl of each oligo at 100µM and 80µl of EB (QIAGen) at 95°C for 2 minutes then cooling to RT at 0.1°C/s) with 2.5µl 100% glycerol and 5µl Ez-Tn5 transposase (Epicentre – Illumina) for 20 minutes at RT.

[00208] Genomic DNA prepared from NA20847 cell lines was used at respective input quantities with 4µl Nextera® HMW Buffer (Epicentre – Illumina), nuclease-free water (Ambion) to 17.5µl and 2.5µl prepared tn5mC transposomes (regardless of the quantity of DNA used). Reactions were incubated at 55°C for 8 minutes in a thermocycler followed by SPRI bead cleanup (AMPure) using 36µl of beads and the recommended protocol with elution in 14µl nuclease-free water (Ambion). Adaptor 2 annealing was then carried out by adding 2µl of 10X Ampligase Reaction Buffer (Epicentre –Illumina), 2µl 10X dNTPs (2.5mM each, Invitrogen), and 2µl 10µM tn5mC-A2top (IDT) to each reaction and incubating at 50°C for 2 minutes followed by 45°C for 10 minutes and cooling at 0.1°C/s to 37°C and subsequent incubation for 10 minutes. Gap repair was then performed by adding 3µl of Ampligase at 5U/µl (Epicentre – Illumina) and 1µl of either T4 DNA Polymerase (tn5mC libraries A-G, NEB) or Sulfolobus DNA Polymerase IV (tn5mC libraries H-J, NEB) and additional incubation at 37°C for 30 minutes. Reactions were then cleaned up using SPRI beads (AMPure) according to recommended protocol using 36µl beads and elution in 50µl nuclease-free water (Ambion).

PARSE EX.1007.236

[00209] Bisulfite treatment was performed using an EZ DNA Methylation™ Kit (Zymo) according to recommended protocols with a 14 hour 50°C incubation and 10µl elution. Eluate was then used as the template for PCR using 12.5µl Kapa 2G Robust HotStart ReadyMix (Kapa Biosystems), 1µl 10µM tn5mC-P1 (5'-[Phos]- CTG TCT CTT ATA CAC ATC TCT GAG [5mC] GGG [5mC] TGG [5mC] AAG G [5mC] AGA [5mC] [5mC] GAT [5mC] -3', IDT) (SEQ ID NO:3), 1µl 10µM Barcoded P2 (From Adey *et. al.* (2010)), 0.15µl 100X SYBR Green (Invitrogen), and 0.35µl nuclease-free water (Ambion). Thermocycling was carried out on a BioRad Opticon Mini real-time machine with the following parameters: 5:00@95°C; (0:15@95°C; 0:15@62°C; 0:40@72°C; Plate Read; 0:10@72°C) X 99. Reactions were monitored and removed from thermocycler as soon as plateau was reached (12-15 cycles).

[00210] Sequencing was carried out using either a full or partial lane on an Illumina HiSeq2000 using custom sequencing primers: Read 1: tn5mC-R1 (5'- GCC TCC CTC GCG CCA TCA GAG ATG TGT ATA AGA GAT AG -3', IDT) (SEQ ID NO:4), Index Read: tn5mC-Ix (5'- TTG TTT TTT ATA TAT ATT TCT GAG CGG GCT GGC AAG GC -3', IDT) (SEQ ID NO:5), Read 2: tn5mC-R2 (5'- GCC TTG CCA GCC CGC TCA GAA ATA TAT ATA AAA AAC AA -3', IDT) (SEQ ID NO:6). Read lengths were either single-read at 36bp with a 9bp index (SE36, libraries A and B, not included in table) or 101bp paired-end with a 9bp index (PE101, libraries C-J). Libraries were only sequenced on runs that did not have lanes containing Nextera® libraries as a precaution due to the similarity between sequencing primers.

[00211] *Read alignment.* The hg19 reference genome was first bisulfite-converted *in silico* for both the top (C changed to T, C2T) and bottom (G changed to A, G2A) strands. Prior to alignment reads were first filtered based on the run metrics, as several libraries were run on lanes in which instrument valve failures resulted in poor quality or reads consisting primarily of "N" bases. Next, reads were filtered to contain no more than 3 "N"s in the first 75 bases and subsequently aligned to both the C2T and G2A strands using *BWA* with default parameters. Reads that aligned to both strands were removed. Read pairs in which neither aligned to either strand were then pulled and trimmed to 76bp (except for SE36 runs) and again aligned to both C2T and G2A strands. For library F, an initial trimming of 25bp from the start of read 2 was performed

PARSE EX.1007.237

due to instrument valve failure during those cycles. Duplicate reads (pairs sharing the same start positions for both reads 1 and 2) were removed and complexity determined. Reads with an alignment score < 10 were then filtered out prior to secondary analysis. Total fold coverage was calculated using the total bases aligned from unique reads over the total alignable bases of the genome (slightly below 3Gb per strand).

[00212] *5mC Calling*. Methylated cytosines were called using a binomial distribution as in Lister *et. al*. (2009) whereby a probability mass function is calculated for each methylation context (CpG, CHG, CHH) using the number of reads covering the position as the number of trials and reads maintaining cytosine status as successes with a probability of success based on the total error rates which were determined by the combined non-conversion rate and sequencing error rate. The total error rate was initially determined by unmethylated lambda DNA spike-ins, however we found that the error rate estimation from the gap-repair portion of reads (as described in the main text) gave a more comprehensive estimate which was slightly higher than that of the lambda estimate, therefore to be conservative, we used the highest determined error rate at 0.009. If the probability was below the value of M, where M * (num. total unmethylated CpG) < 0.01 * (num. total methylated CpG), the position was called as being methylated, thus enforcing that no more than 1% of positions would be due to the error rate.

Results

[00213] *Ultra-low-input transposase-based WGBS library performance*. tn5mC-seq was performed to sequence the methylome of a lymphoblastoid cell line (NA20847) using libraries constructed from 1 nanogram to 200 nanograms of input genomic DNA. Each library was barcoded during PCR amplification and subjected to either a spike-in (5%) or majority (80-90%) of a lane of sequencing on an Illumina HiSeq2000 (PE100; v2 chemistry). These data are summarized in Table 3, below.

PARSE EX.1007.238

TABLE 3 Summary of tn5mC-seq libraries and sequencing

Name	Input DNA (ng)	Percent Aligning	Percent Unique	Unique Aligned Reads	Mean Insert Size (bp)
tn5mC-C	200	68	93	127,098,152	198
tn5mC-D	50	75	90	133,383,834	254
tn5mC-E*	1	12	76	11,181,960	134
tn5mC-F*	10	65	95	118,170,302	168
tn5mC-G*	50	61	97	87,294,793	180
tn5mC-H	1	11	78	12,393,357	126
tn5mC-I**	10	62	n/a	29,546,077	n/a
tn5mC-J	50	71	95	132,144,644	196
TOTAL				**651,213,119**	

*Valve failures in Read 2 resulted in extensive read trimming (50-70bp)
**Complete valve failure on Read 2.

[00214] Raw reads were initially filtered for instrument valve failures at specific locations of reads and then removal of reads containing over three Ns or extremely low quality bases (phred score <= 2) in the first 50 bases. Alignment was then performed using *BWA*(Li and Durbin 2009) to *in silico* converted top and bottom strand references of hg19 (GRC37) followed by trimming and re-alignment. Duplicate reads were identified and removed according to their start position and insert size. The percentage of post-filtering reads that align for each library is shown, as is the percentage of these that are non-duplicates.

[00215] Reads were aligned to an *in silico* converted hg19 (GRC37) to both the top (C=>T) and bottom (G=>A) strands using *BWA* (Li and Durbin 2009) followed by read trimming of unmapped reads and secondary alignment using the same parameters. Because unmethylated nucleotides are incorporated during the gap-repair step (first 9 base-pairs of the second read and last 9 base-pairs before the adaptor as determined by insert size on the first read), the gap-repair regions must be excluded from methylation analysis. However, these bases also serve as an internal control for the conversion rate of the bisulfite treatment. This was found to be >99% for all libraries, and this was independently confirmed using unmethylated lambda DNA spike-ins to two libraries.

PARSE EX.1007.239

[00216] For each library constructed using ≥10 nanograms of genomic DNA, over 100 million aligned reads were obtained (60-75% of total filtered reads; see Methods) of high complexity (90-97% non-duplicates). Despite the significantly reduced performance of libraries prepared from 1 nanogram, approximately12 million reads were still aligned and the library was of reasonable complexity (78% non-duplicates). Post-alignment reads were merged and quality filtered for a total of 51.7 gigabases of aligned, unique sequence. The average read depth was 8.6X per strand with >96% of CpG and >98% of non-CpG cytosines covered genome-wide (Figure 36c).

[00217] *Lymphoblastoid cell line methylation.* Approximately 46 million 5mC positions (1% FDR; see Methods) were detected, accounting for 4.2% of total cytosines with coverage. The majority of methylation observed was in the CpG context (97.1%), and the global CpG methylation level was 69.1%. This level is similar to that of the fetal fibroblast cell line IMR90 sequenced by Lister and colleagues (Lister et al. 2009) (67.7%), and consistent with the observation that CpG methylation levels are reduced in differentiated cell types. Additionally, CHG and CHH methylation levels were substantially lower than in ES cells, at 0.36% and 0.37% respectively, again consistent with the differentiated cell type. On the chromosome scale, the methylation was greater in sub-telomeric regions (Figure 36d), as expected by the miRNA-mediated pathways that act to control telomere length (Benetti et al. 2008). An analysis of functionally annotated genic regions revealed a sharp decrease in CpG methylation through the promoter region followed by a minor increase in the 5'UTR and then elevated levels of methylation throughout the gene body, particularly at introns (Figure 36e,f), consistent with previously described CpG methylation profiles (Lister et al. 2009).

Discussion

[00218] tn5mC-seq was developed as a novel method for rapidly preparing complex, shotgun bisulfite sequencing libraries for WGBS. In brief, the method utilizes a hyperactive Tn5 transposase derivative to fragment genomic DNA and append adaptors in a single step, as previously characterized for the construction of DNA-seq libraries (Adey et al. 2010). In order for library molecules to withstand bisulfite treatment, the adaptors are methylated at all cytosine residues and an oligonucleotide

PARSE EX.1007.240

replacement strategy is employed to make each single-strand covalently flanked by adaptors. The high efficiency of the transposase and overall reduction in loss-associated steps permits construction of high quality bisulfite sequencing libraries from as little as 10ng as well as useful sequence from 1ng of input DNA.

[00219] These results illustrate how derivatives of the transposase-based method for DNA-Seq library preparation enable important applications of next-generation sequencing where its advantages are perhaps even more relevant. The ability to generate such libraries from very low amounts of input genomic DNA substantially improves the practicality of whole methylome sequencing, and removes an important advantage of less comprehensive methods such as RRBS (Meissner et al. 2005; Harris et al. 2010). Specifically, low-input WGBS with tn5mC-seq may make possible the comprehensive interrogation of methylation in many contexts where DNA quantity is a bottleneck, *e.g.* developing anatomical structures, microdissected tissues, or pathologies such as cancer, where the epigenetic landscape is of interest but tissue quantity limits resolution.

Example 5: Identifying distant regulatory sites and measuring chromosome conformation

[00220] Recent studies have shown the importance and complexity of physical interactions between genetic elements within a genome. Measuring these interactions can help to explain how distant cis and trans regulatory DNA plays a role in gene regulation (including which genes are affected by which enhancers, how chromosomes are arrayed within a cell, how certain transcription factors like AR and ER bind and influence gene expression). It can also provide clues to the formation of the rearrangements and inversions involved in cancer and other genetic diseases.

[00221] Current methods of assaying physical interactions using high throughput sequencing include chromatin interaction analysis using paired end tag sequencing (ChIA-PET; interactions of transcription factor-binding sites) and Hi-C (method of producing maps of genome). Both approaches have limitations that can be attributed to the low efficiency and specificity of intramolecular ligation, which the methods use to pair distant regions of DNA together for sequencing. Such a ligation step requires large

PARSE EX.1007.241

amounts of input DNA (100+ ug) and can result in technical artifacts in which fragments of DNA are ligated to each other even when they are not typically associated with each other.

[00222] These problems are important when trying to understand the genomic architecture of a small population of cells, like embryonic stem cells and cancer cells. It also means that any inferences of DNA interactions resulting from transcription factor binding that are measured are on average of extremely large cell populations (e.g., 10^8 cells or ~500 micrograms of DNA). Given that different cells can have different genomic architectures or patterns of transction factor binding, an approach that requires less starting DNA may be useful.

[00223] Therefore, methods for measuring DNA-DNA and DNA-protein interactions within smaller populations of cells are provided below. Such methods use the "infinipair" technology (described in example 3A above) to directly sequence multiple fragments off of immunoprecipitated DNA that has been crosslinked (Figure 28). These methods differ from the CHIA-PET and Hi-C methods because they directly assay crosslinked fragments of DNA without the extra step of intramolecular ligation.

5.A. Identification of distant regulatory binding sites

[00224] In one embodiment, modifications of the technology described in example 3A (referred to herein as the "inifinpair" technology) may be used to identify interactions between transcription factor (TF) binding sites, such as, for example, those found on the Estrogen receptor. As shown in Figure 29, approximately 10^4 cells are induced with hormone (~10 ng), followed by chromatin immunoprecipitation of the cells. Next, the chromatin fibers are broken by cross-linking the cells with 1% formaldehyde followed by sonication. An ER/AR/receptor specific antibody is then used to enrich binding chromatin fragments.

[00225] Next, end repair is performed using T4 polymerase to ligate to A+B adaptors. No phosphorylation, ligation of half linkers or dilution of ligation is necessary in this method. The infinipair technology is then used to generate clusters corresponding to immunoprecipitated complexes. Neighboring clusters are paired together to create a list of interactions between chromosomal positions.

PARSE EX.1007.242

[00226] To narrow down the list of putative interactions, the data collected is then intersected with CHIP-Seq information, which provides information on known binding sites of transcription factors. Information from multiple libraries is overlapped to increase confidence in called interactions. The structure of the chromatin interactions is characterized using this data, and is also used to link regulatory regions to DNA (i.e link genes to enhancers).

[00227] Some of the benefits of using this method may include, but are not limited to, (1) Higher accuracy in pairing interactions – (no inter-fragment ligation), (2) Lower input DNA required, resulting in more applications may be used (i.e., interactions within smaller sets of samples (ES cells/cancer cells/smaller groups of healthy cells) may be identified, (3) Easier workflow – no dilute ligations, no PCR, no MmeI digestions etc., and (4) Less sequencing required.

5.B. Inferring chromosome conformation

[00228] In another embodiment, infinipair technology is used to model chromosome conformation in small numbers of cells. Previous methods such as the Hi-C method required a larger number of cells (~10^7 cells; ~ 50 ug DNA). As shown in Figure 30, genomic DNA is cross linked. The cells are lysed using a homogenizer and the chromatin is then spun down. Chromatin proteins are removed by incubating in 1% SDS followed by Triton X-100. Chromatin is then digested by incubating in HindIII overnight.

[00229] Chromatin is purified using columns with beads directed against an anti-chromatin antibody. A and B adaptors are then ligated to DNA fragments without the need to biotinylate. Next, the infinipair technology is used to generate inifinipair clusters corresponding to complexes. Neighboring clusters are paired together to create a list of interactions between chromosomal positions. The information generated is then used to generate 3 dimensional models and to better understand the conformation of specific cell types.

PARSE EX.1007.243

Example 6: Integration of Short-Range, Mid-Range and Long-Range Contiguity For a More Cost-Effective Sequencing Method

[00230] The focus of this Example includes 1) integrating methods developed in Examples 1-3 for the high-quality *de novo* assembly of the mouse genome; 2) integrating these same methods for the haplotype-resolved resequencing of a human genome; and 3) extending compatibility to other next-generation sequencing paradigms.

6.A. Cost analysis and the path to the $1,000 mark

[00231] An important aspect of the methods described herein is that the costs are almost entirely dependent on the costs of the sequencing platform with which they are integrated. If "X" is the cost of genome resequencing, then the cost the methods described herein can be abstracted as "a + bX", where "a" is the fixed cost per sample of capturing contiguity information (e.g. the cost of an *in situ* transposition reaction), and "b" is the proportion of sequencing required to recover that information relative to genome resequencing. Estimates for "a" are low, i.e. less than $30 per method. This is because reactions such as *in situ* transposition and PCR manipulate genomic DNA *en masse* within single reagent volumes. Furthermore, reagents such as degenerate oligonucleotides and microfluidic devices are relatively inexpensive, and their costs can be amortized over many uses. The value of "b" is more difficult to predict, and is dependent on the extent of success in implementing and optimizing each method. However, it is noted that the barcodes themselves are short as compared to the primary reads with which they are *in cis* (e.g. SE25 barcode versus PE76 primary).

[00232] Sequencing costs associated with each of the following application of the methods described herein should be roughly the *same* as the cost of 40x resequencing of a mammalian genome with the same platform, i.e. "b≈1". As demonstrated by the original assemblies of the mouse and human genomes, it is possible to achieve a high quality *de novo* assembly of a mammalian genome with substantially less sampling than is currently used for genome resequencing, provided that sufficient contiguity information is also obtained.

PARSE EX.1007.244

6.B. *De novo* assembly of the mouse genome

[00233] Using the contiguity information obtained from the methods described above, a high-quality *de novo* assembly of a mammalian genome may be obtained *de novo* To accomplish this, existing tools for either conventional or 'next-generation' *de novo* assembly (Schatz et al. 2010) will be repurposed and applied to these data, and additional software will be developed as necessary. To minimize costs without significantly compromising quality, the optimal mix of contiguity mapping methods (i.e. at different scales) will be determined. This may require, for example, oversampling the genome with each contiguity mapping method, and then downsampling to include different proportions of data from each method and evaluating the impact on the quality of *de novo* assembly. Focusing on the *de novo* assembly of the mouse genome as a test case, the contiguity of the original assembly (i.e. contig N50 of 24.8 Kb; supercontig N50 of 16.9 Mb) will be exceeded with the same amount of data as is required for 40x resequencing (2.5 Gb x 40 = ~100 Gb), i.e. "b≈1". Initially, sequencing costs will predominate, i.e. "bX >> a", but even as this changes the total costs of preparatory reactions ("a") should be kept to <$100, even if all scales of contiguity mapping methods are used (i.e. Short-Range, Mid-Range and Long-Range Contiguity).

6.C. Haplotype resolved resequencing of a human genome

[00234] Preliminary data (2.D) shows that a modest amount of contiguity information may provide extensive haplotype resolving power. For this data, the software required for haplotype-resolved genome resequencing has been developed or will be developed. Additionally, algorithms will be developed to discover SNPs resolve haplotypes using the same data, as accurately calling haploid genotypes requires less than half as much sequencing as calling diploid genotypes. Contiguity mapping methods will be integrated to resequence and simultaneously haplotype resolve a human genome, with a target of >95% coverage in haplotype-resolved blocks with an N50 of at least 1 Mb while maintaining >99.5% concordance with HapMap data at D'>0.90. As with *de novo* assembly of the mouse genome, this may be achieved with the same amount of sequencing as would be required for 40x haplotype-blind resequencing of the human genome (3 Gb x 40 = ~120 Gb), i.e. "b≈1".

PARSE EX.1007.245

6.D. Extending the compatibility of contiguity mapping methods to other sequencing paradigms

[00235] Although the methods for capturing contiguity information as described herein are directed to being used with the sequencing technologies with which they are integrated, such methods may be developed for other sequencing technologies and with other sequencing platforms. These include other cyclic-array platforms (e.g. Polonator, SOLiD), as well as emerging paradigms such as nanopore sequencing.

PARSE EX.1007.246

REFERENCES

The references, patents and published patent applications listed below, and all references cited in the specification above are hereby incorporated by reference in their entirety, as if fully set forth herein.

Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X et al. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11(12): R119.

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol 27(4): 361-368.

Bansal V, Bafna V. HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. Bioinformatics. 2008;24(16):i153-9.

Benetti R, Gonzalo S, Jaco I, Munoz P, Gonzalez S, Schoeftner S, Murchison E, Andl T, Chen T, Klatt P et al. 2008. A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. Nat Struct Mol Biol 15(3): 268-279.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53-9. PMCID: PMC2581791.

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol. 2008;26(10):1146-53. PMCID: PMC2683588.

Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A. 2003;100(7):3960-4. PMCID: PMC153030.

Clark SJ, Harrison J, Paul CL, Frommer M. 1994. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22(15): 2990-2997.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452(7184): 215-219.

Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J et al. 2009. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 27(4): 353-360.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM et al. 2008. A Bayesian deconvolution strategy for

PARSE EX.1007.247

immunoprecipitation-based DNA methylome analysis. Nat Biotechnol 26(7): 779-785.

Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A. 2003;100(15):8817-22. PMCID: PMC166396.

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science.2009; 327(5961):78-81.

Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, et al. A three-dimensional model of the yeast genome. Nature 2010; 465(7296):363-7. PMCID: PMC2874121.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133-8.

Fan HC, Wang J, Potanina A, & Quake SR (2011) Whole-genome molecular haplotyping of single cells. Nat Biotech 29(1):51-57.

Fullwood, M.J. et al. An oestrogen-receptor-α-bound human chromatin interactome. Nature 462, 58–64 (2009).

Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol. 2008;26(3):317-25.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2010 Dec 27. [Epub ahead of print] PubMed PMID: 21187386.

Grunenwald H, Baas B, Goryshin I, Zhang B, Adey A, Hu S, Shendure J, Caruccio N, Maffitt M. 2011. Nextera PCR-Free DNA Library Preparation for Next-Generation Sequencing. . (Poster Presentation, AGBT).

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 6(4): 468-481.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y et al. 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol 28(10): 1097-1105.

Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J. Parallel, tag-directed assembly of locally derived short sequence reads. Nat Methods. 2010; 7(2):119-22. PMCID: 2848820.

http://www.epibio.com/nextera/nmeth_f_269.pdf.

PARSE EX.1007.248

Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316(5830):1497-502.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. Nature. 2008;453(7191):56-64. PMCID: PMC2424287.

Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, Sudmant PH, Ng SB, Alkan C, Qiu R, Eichler EE, Shendure J. Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotechnol. 2011 Jan;29(1):59-63. Epub 2010 Dec 19. PubMed PMID: 21170042.

Kitzman JO, et al. (2011) Haplotype-resolved genome sequencing of a Gujarati Indian individual. Nat Biotech 29(1):59-63.

Lai Z, Jing J, Aston C, Clarke V, Apodaca J, Dimalanta ET, et al. A shotgun optical map of the entire Plasmodium falciparum genome. Nat Genet. 1999;23(3):309-13.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol. 2007;5(10):e254. PMCID: PMC1964779.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010; 20(2):265-72. PMCID: PMC2813482.

Li Y, Kim HJ, Zheng C, Chow WH, Lim J, Keenan B, et al. Primase-based whole genome amplification. Nucleic Acids Res. 2008;36(13):e79. PMCID: PMC2490742.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14): 1754-1760.

Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J et al. 2010. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol 8(11): e1000533.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289-93. PMCID: PMC2858594.

Lim A, Dimalanta ET, Potamousis KD, Yen G, Apodoca J, Tao C, et al. Shotgun optical maps of the whole Escherichia coli O157:H7 genome. Genome Res. 2001;11(9):1584-93. PMCID: PMC311123.

Lin J, Qi R, Aston C, Jing J, Anantharaman TS, Mishra B, et al. Whole-genome shotgun optical mapping of Deinococcus radiodurans. Science. 1999;285(5433):1558-62.

PARSE EX.1007.249

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462(7271): 315-322.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376-80. PMCID: PMC1464427.

Mazutis L, Araghi AF, Miller OJ, Baret JC, Frenz L, Janoshazi A, et al. Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis. Anal Chem. 2009;81(12):4813-21.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33(18): 5868-5877.

Mitra RD, Shendure J, Olejnik J, Edyta Krzymanska O, Church GM. Fluorescent in situ sequencing on polymerase colonies. Anal Biochem. 2003;320(1):55-65.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621-8.

MGSC (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520-562.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-6. PMCID: PMC2844771.

Ramanathan A, Huff EJ, Lamers CC, Potamousis KD, Forrest DK, Schwartz DC. An integrative approach for the optical sequencing of single DNA molecules. Anal Biochem. 2004;330(2):227-41.

Riehn R, et al. (2005) Restriction mapping in nanofluidic devices. Proceedings of the National Academy of Sciences of the United States of America 102(29):10012-10016.Schatz MC, Delcher AL, Salzberg SL. Assembly of large genomes using second-generation sequencing. Genome Res. 2010; 20(9):1165-73. PMCID: PMC2928494.

Ritz A, Bashir A, Raphael BJ. Structural variation analysis with strobe reads. Bioinformatics.2010; 26(10):1291-8.

Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK. Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science. 1993;262(5130):110-4.

Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135-45.

Shendure J, Mitra RD, Varma C, Church GM. Advanced sequencing technologies: methods and goals. Nat Rev Genet. 2004;5(5):335-44.

PARSE EX.1007.250

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005;309(5741):1728-32.

Steensel B and Dekker J. Genomics tools for unraveling chromosome architecture. Nature Bitoechnology 13 October 2010.

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., Dekker, J., Lander, E. S., Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. http://www.jove.com/details.stp?id=1869 doi: 10.3791/1869. J Vis Exp. 39 (2010).

Waterston RH, Lander ES, Sulston JE. More on the sequencing of the human genome. Proc Natl Acad Sci U S A. 2003;100(6):3022-4; author reply 5-6. PMCID: PMC152236.

Waterston RH, Lander ES, Sulston JE. On the sequencing of the human genome. Proc Natl Acad Sci U S A. 2002;99(6):3712-6. PMCID: PMC122589.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420(6915):520-62.

Zeng Y, Novak R, Shuga J, Smith MT, Mathies RA. High-performance single cell genetic analysis using microfluidic emulsion generator arrays. Anal Chem. 2010; 82(8):3183-90. PMCID: PMC2859697.

Zhou S, et al. (2007) Validation of rice genome sequence by optical mapping. BMC Genomics 8(1):278.

Zhou S, et al. (2009) A Single Molecule Scaffold for the Maize Genome. PLoS Genet 5(11):e1000711.

Zilberman D, Henikoff S. 2007. Genome-wide analysis of DNA methylation patterns. Development 134(22): 3959-3965.

PARSE EX.1007.251

CLAIMS

What is claimed is:

1. A method for capturing contiguity information comprising:

treating a target DNA sequence with a transposase resulting in one or more fragmentation or insertion events;

adding or inserting one or more recognition sequences to the target DNA sequence (i) during the transposase treatment or (ii) during a subsequent amplification;

sequencing the treated DNA; and

capturing contiguity information by identifying target DNA sequences or recognition sequences having a shared property.

2. The method of claim 1, wherein the one or more fragmentation or insertion events results in generation of a library of target nucleic acid molecules derived from the target DNA.

3. The method of claim 2, wherein the one or more recognition sequences are one or more barcodes that are symmetrically tagged to sequences adjacent to each fragmentation or insertion event.

4. The method of claim 3, wherein the shared property of the one or more barcodes is an identical or complementary barcode sequence.

5. The method of claim 1, wherein the target DNA sequence comprises a set of target DNA fragments.

6. The method of claim 5, further comprising compartmentalizing the target DNA fragments with emulsions or dilutions, generating two or more compartments of target DNA fragments prior to or after treating with the transposase.

7. The method of claim 6, wherein the one or more recognition sequences are one or more compartment-specific barcodes, each of which corresponds to the one or more compartments generated in the compartmentalizing step.

8. The method of claim 7, wherein the shared property of the one or more primer sequences is an identical compartment-specific barcode.

PARSE EX.1007.252

9. The method of claim 7, further comprising performing nucleic acid amplification using primers bearing compartment-specific barcodes.

10. The method of claim 1, wherein the one or more recognition sequences is one or more adaptor sequences that modify the ends of the target DNA sequence, or insert within the target DNA sequence.

11. The method of claim 10, wherein the one or more adaptor sequences are complementary to one or more surface-bound primers.

12. The method of claim 11, wherein the transposase is bound to a nucleic acid that is complementary to a second surface-bound primer.

13. The method of claim 11, further comprising hybridizing the one or more adaptor sequences to the one or more surface bound primers.

14. The method of claim 13, wherein the shared property is a constrained physical location.

15. The method of claim 1, wherein the transposase is bound to a surface-bound recognition sequence to form a surface-bound transposase complex.

16. The method of claim 15, wherein the shared property is a constrained physical location.

17. The method of claim 15, wherein treating the target DNA sequence comprises exposing a plurality of surface-bound transposase complexes to the target DNA sequence.

18. A method of bisulfite sequencing comprising:

a) performing *in vitro* transposition into target DNA molecules with transposase complexes, each transposase complex comprising a double stranded DNA transposase recognition sequence and a single stranded DNA adaptor overhang having methylated cytosine (C) residues

b) subjecting transposed target DNA molecules to bisulfite treatment;

c) performing nucleic acid amplification

PARSE EX.1007.253

d) sequencing the resulting nucleic acid library

19. The method of claim 18, further comprising incorporating a second adaptor to nucleic acid fragments derived from target DNA, after step (a) and before step (b), wherein the second adaptor is designed to facilitate nucleic acid amplification in step (c).

20. The method of claim 18, further comprising incorporating a second adaptor to nucleic acid fragments derived from target DNA, after step (b) and before step (c), wherein the second adaptor is designed to facilitate nucleic acid amplification in step (c).

21. A method for inferring chromosome conformation comprising:

a) cross-linking DNA within cells;

b) isolating cross-linked DNA from cells;

c) fragmenting the cross-linked DNA;

d) end-modifying fragmented, cross-linked DNA molecules with an adaptor that is complementary to or that corresponds to a first surface-bound primer;

e) hybridizing ends of the fragmented, end-modified target DNA molecules to the first surface-bound primer;

f) performing transposition with non-surface-bound transposase complexes, each non-surface-bound transposase complex comprising a DNA transposase and one or more sequences corresponding to a second surface-bound primer;

g) performing cluster amplification to produce clusters of clonally derived nucleic acids;

h) sequencing clusters of clonally derived nucleic acids; and

i) determining physical interactions between chromosomal positions by paring neighboring clusters together.

22. The method of claim 21, wherein the isolated cross-linked DNA are part of cross-linked DNA-protein complexes.

PARSE EX.1007.254

23. The method of claim 22, further comprising enriching for one or more specific cross-linked DNA-protein complexes by immunoprecipitation after step (c) and before step (d).

PARSE EX.1007.255

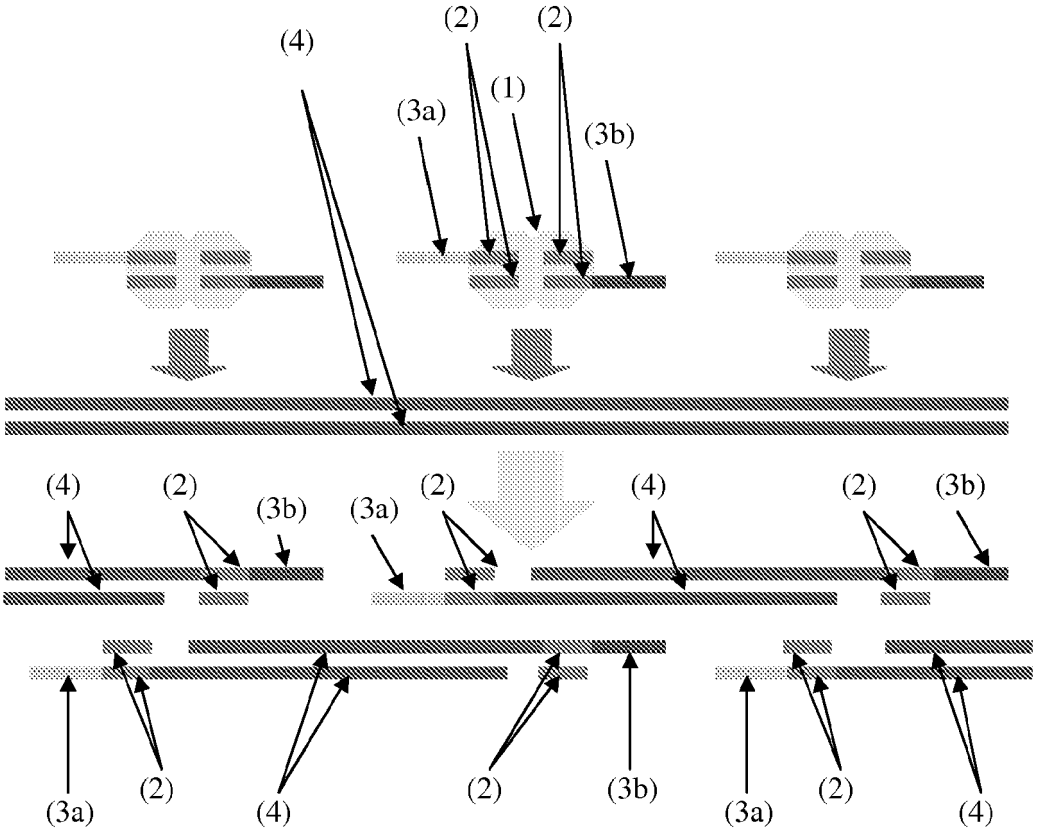


Figure 1

PARSE EX.1007.256

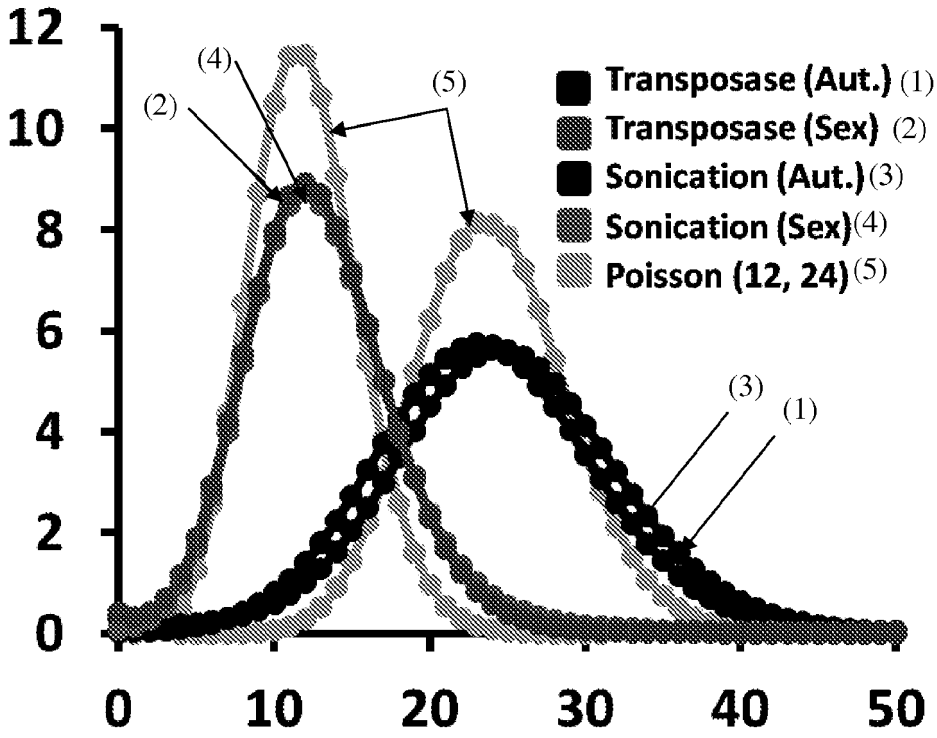


Figure 2

PARSE EX.1007.257

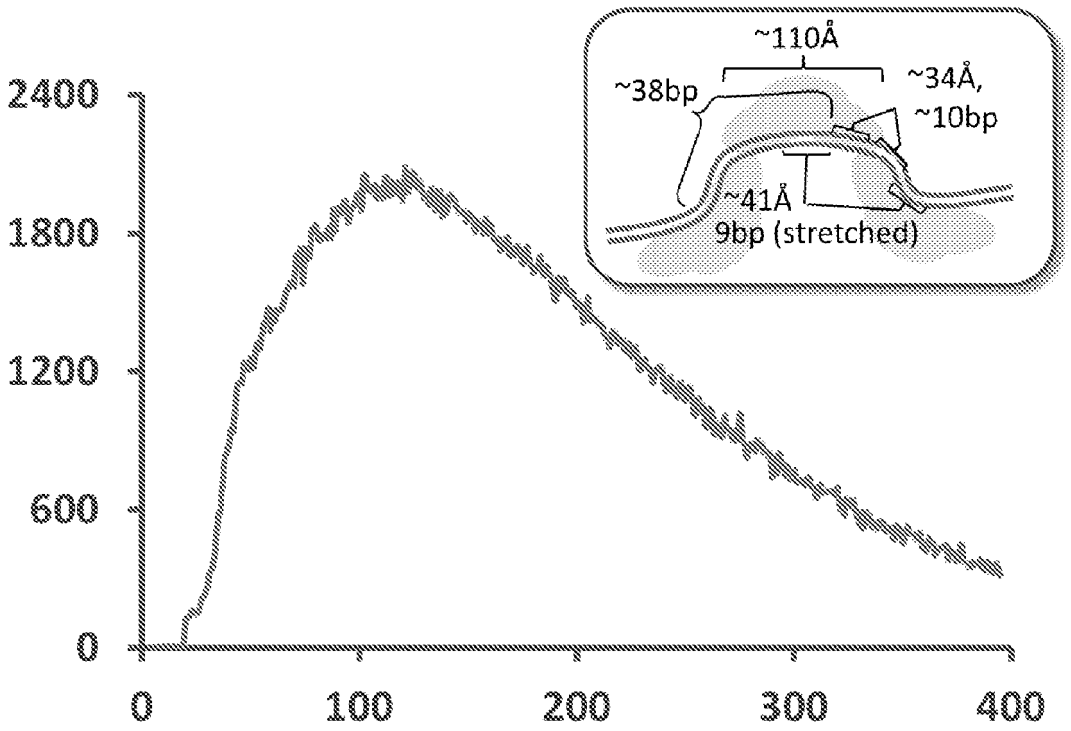


Figure 3

PARSE EX.1007.258

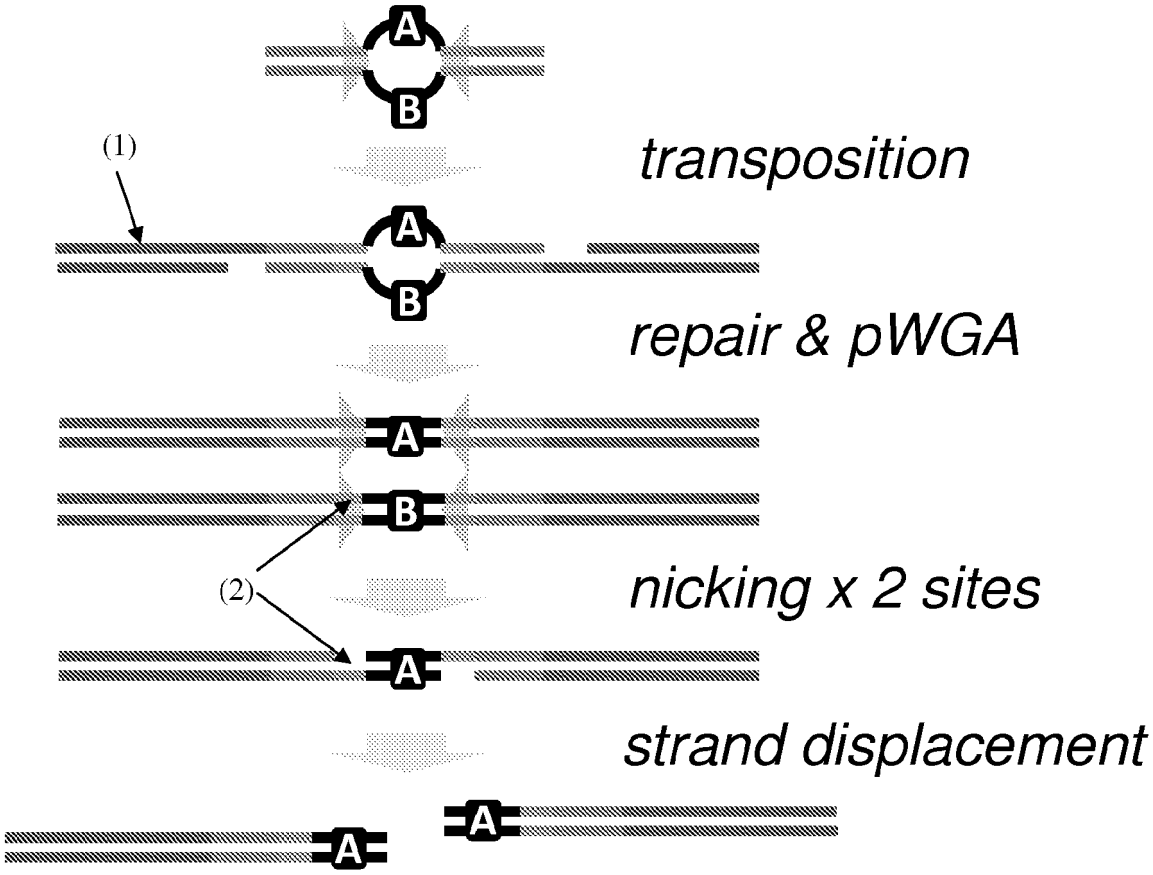


Figure 4

PARSE EX.1007.259

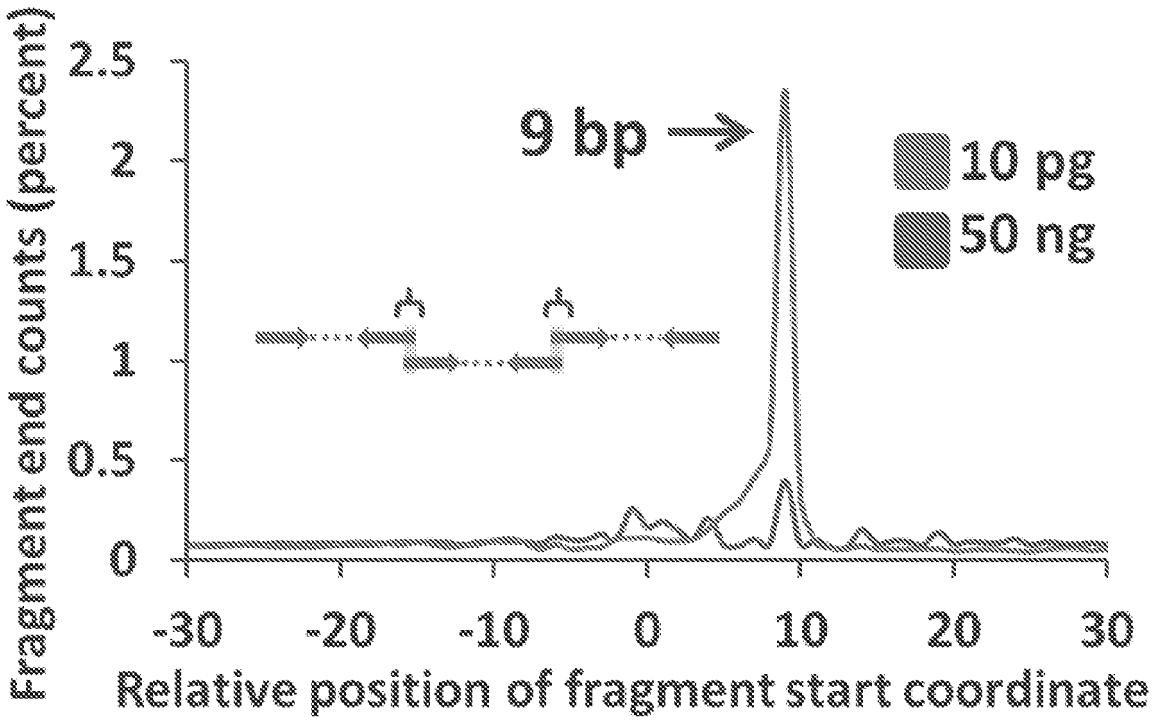


Figure 5

PARSE EX.1007.260

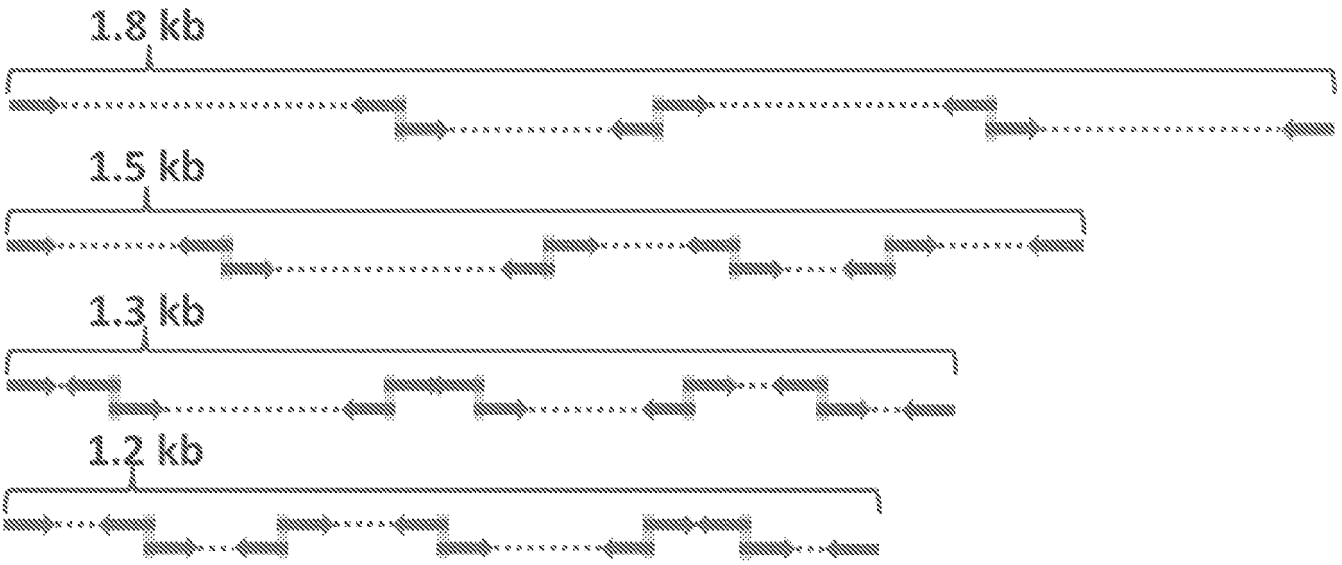


Figure 6

PARSE EX.1007.261

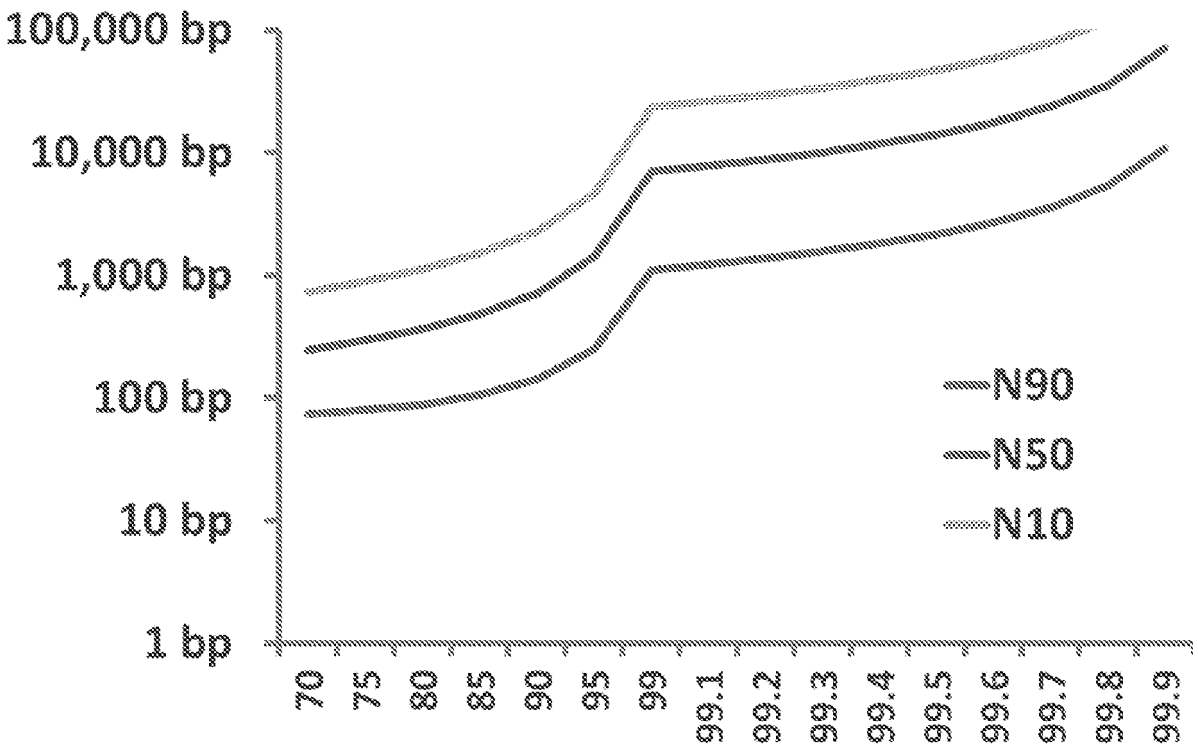


Figure 7

PARSE EX.1007.262

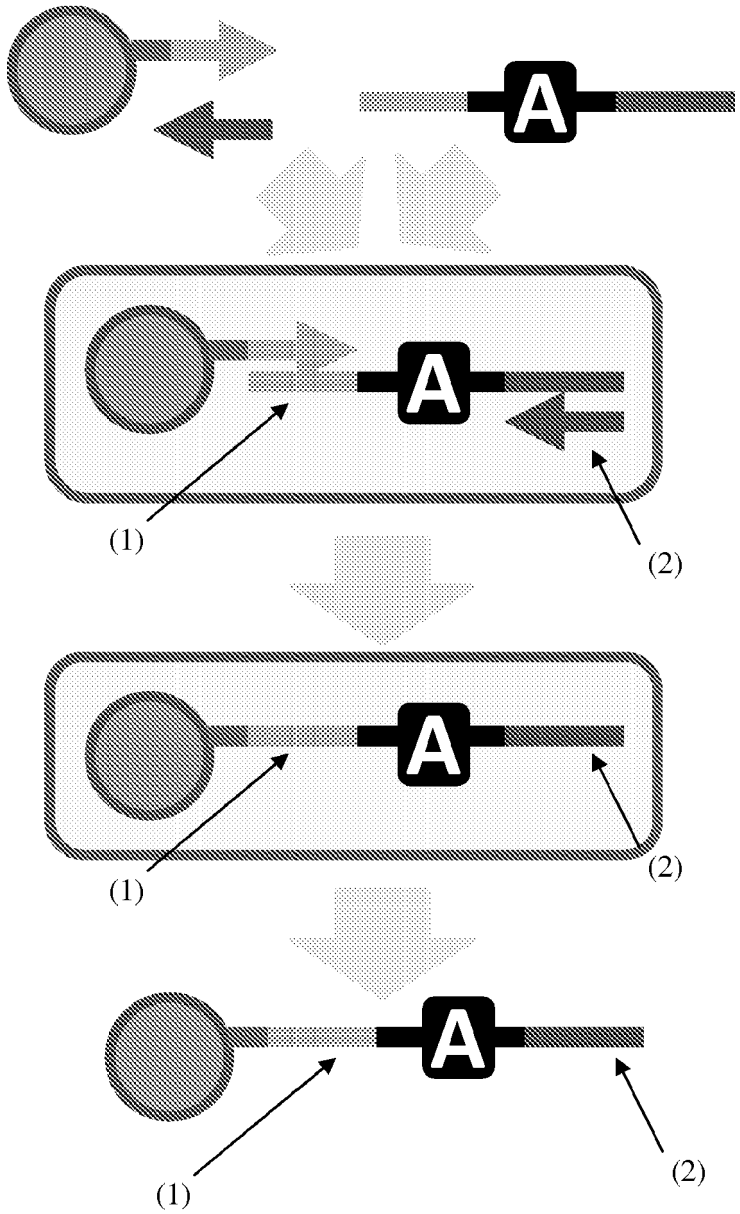


Figure 8

PARSE EX.1007.263

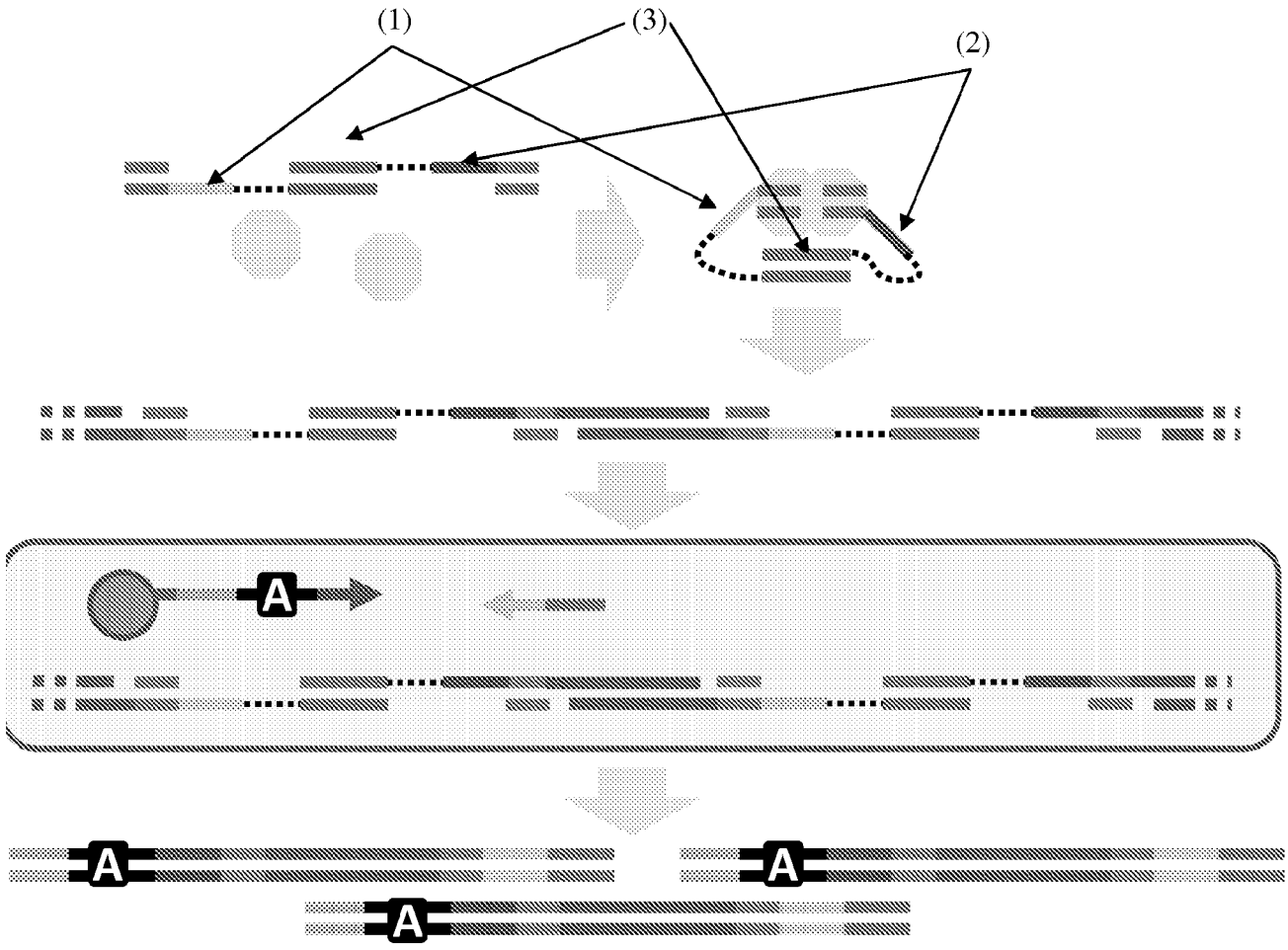


Figure 9

PARSE EX.1007.264

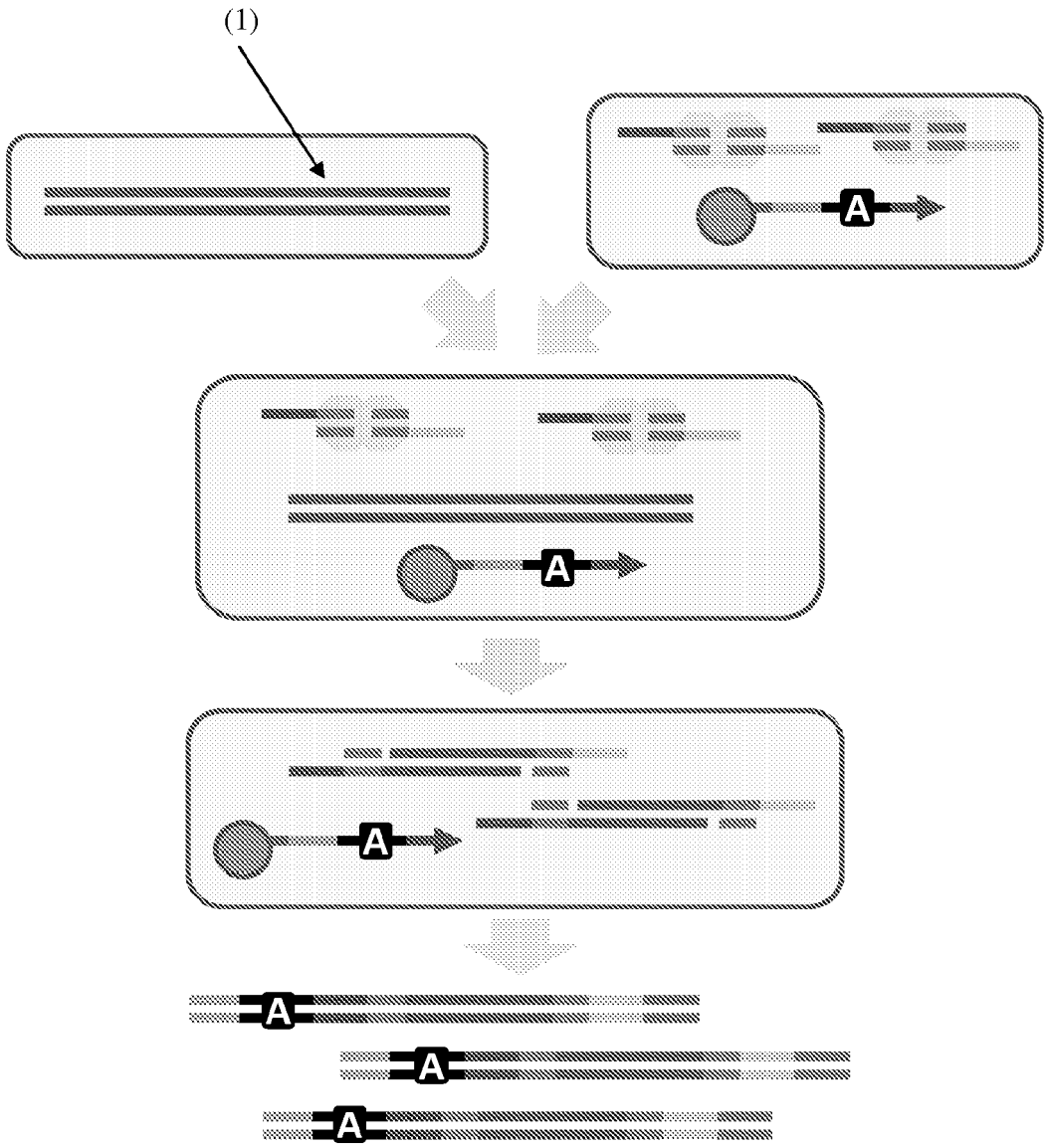


Figure 10

PARSE EX.1007.265

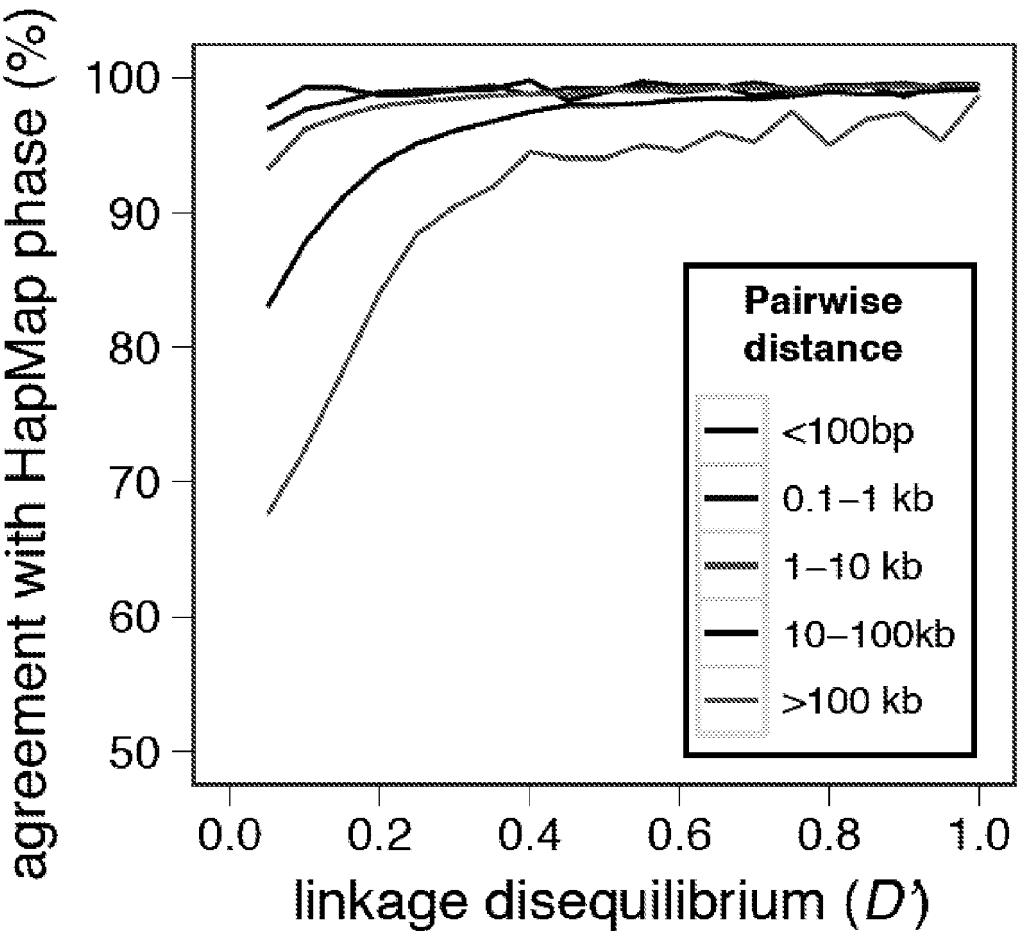


Figure 11

PARSE EX.1007.266

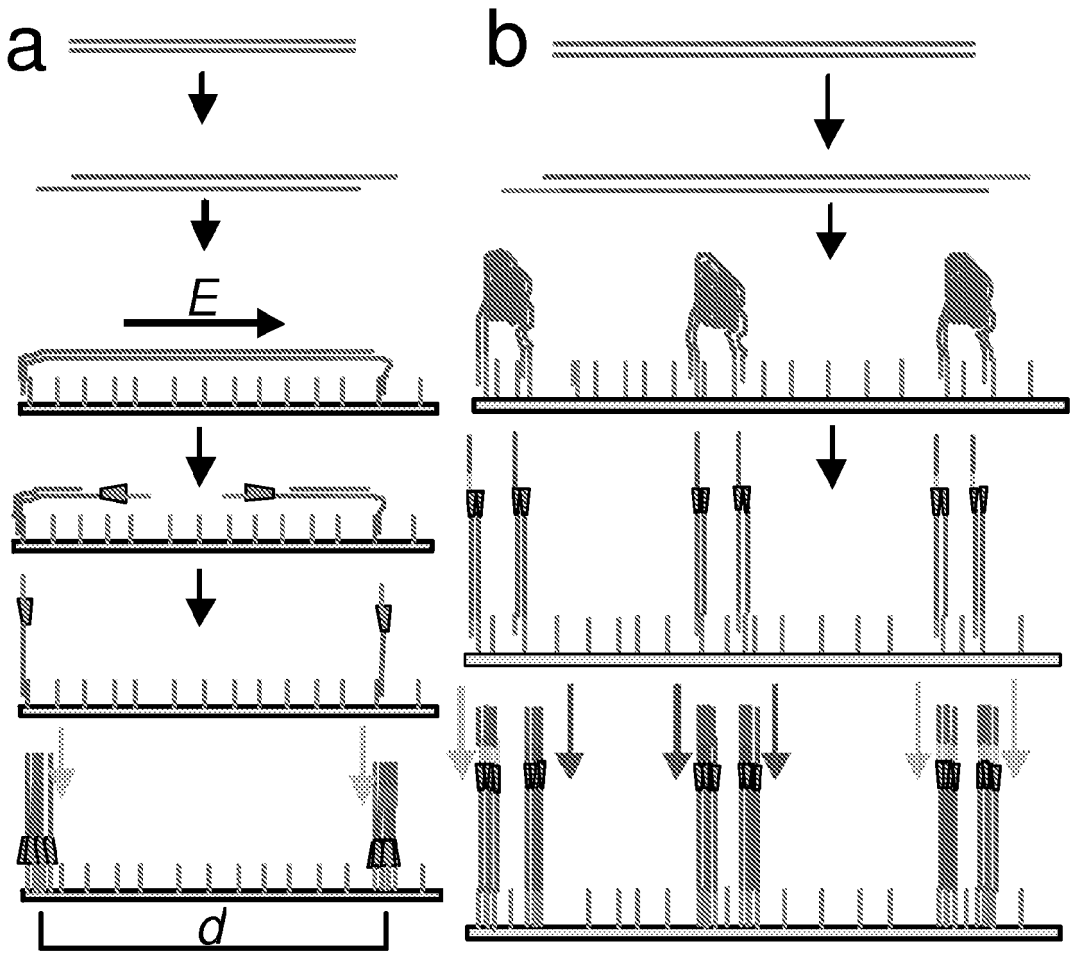


Figure 12

PARSE EX.1007.267

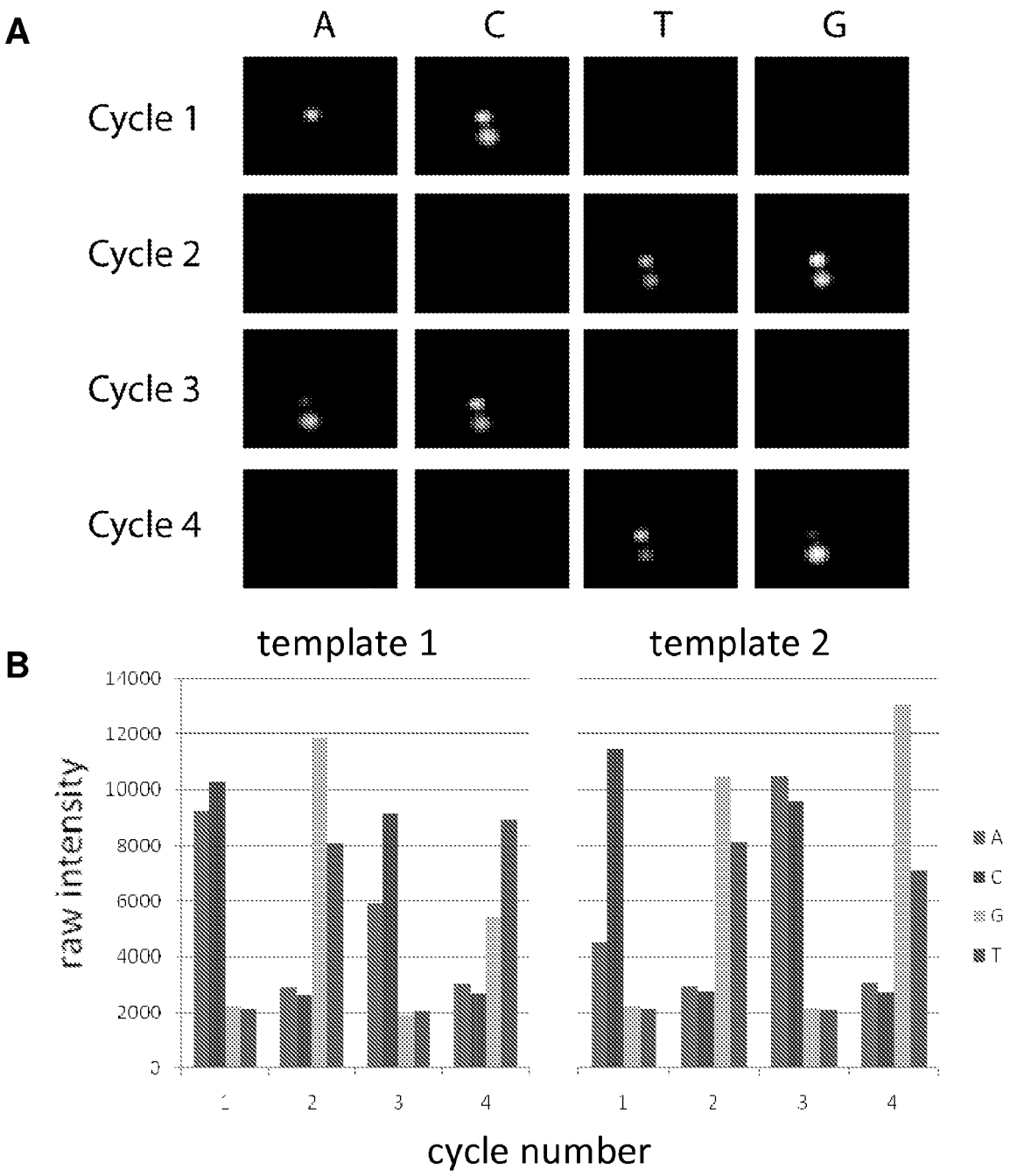


Figure 13

PARSE EX.1007.268

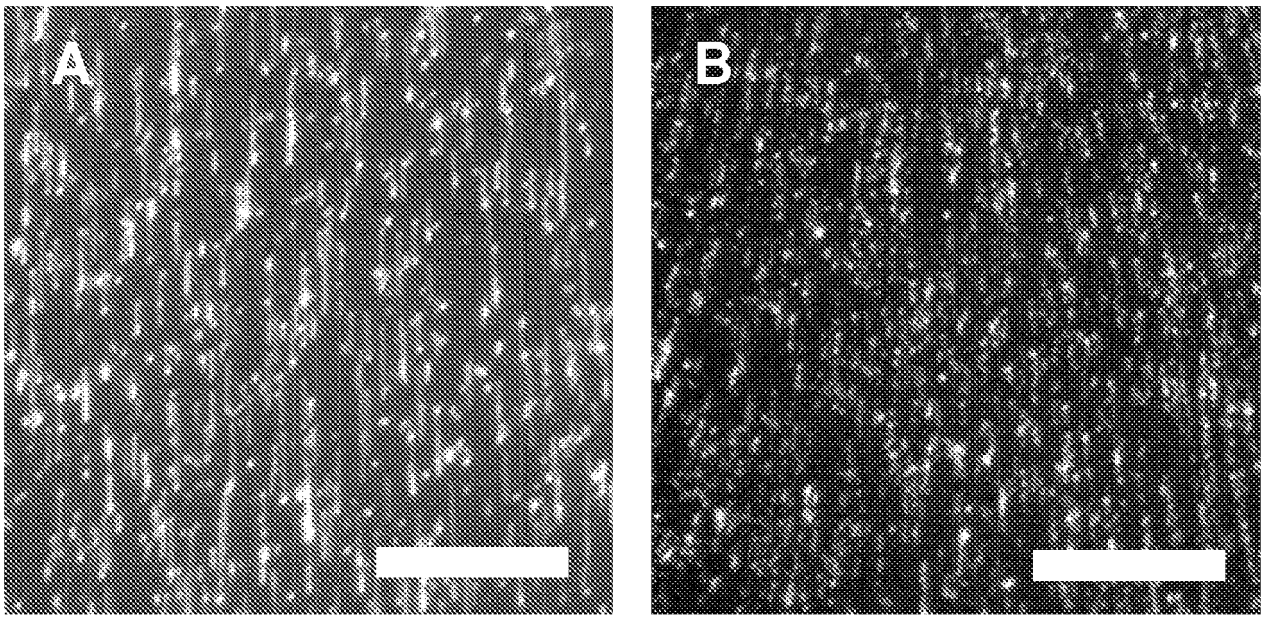


Figure 14

PARSE EX.1007.269

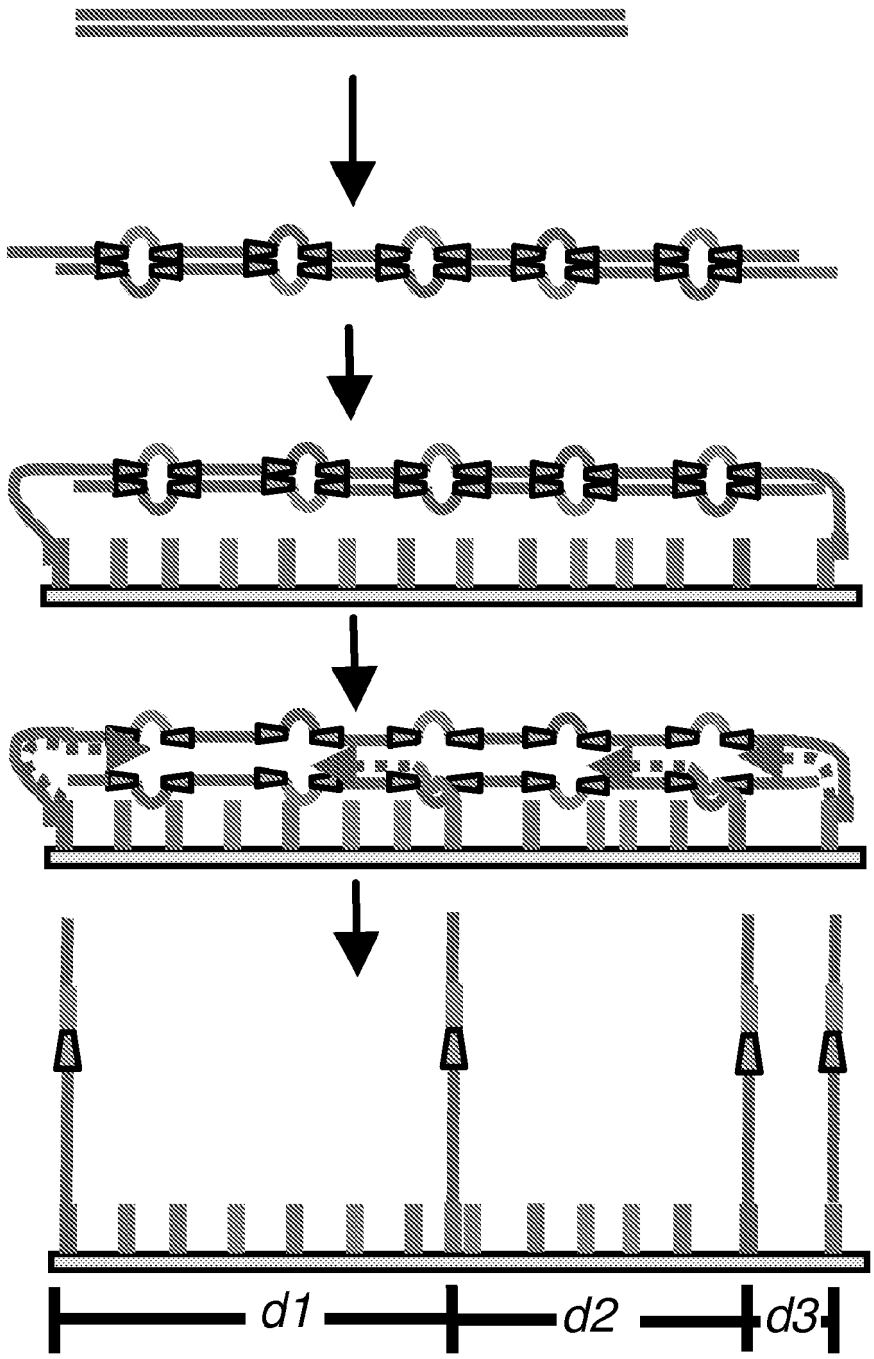


Figure 15

PARSE EX.1007.270

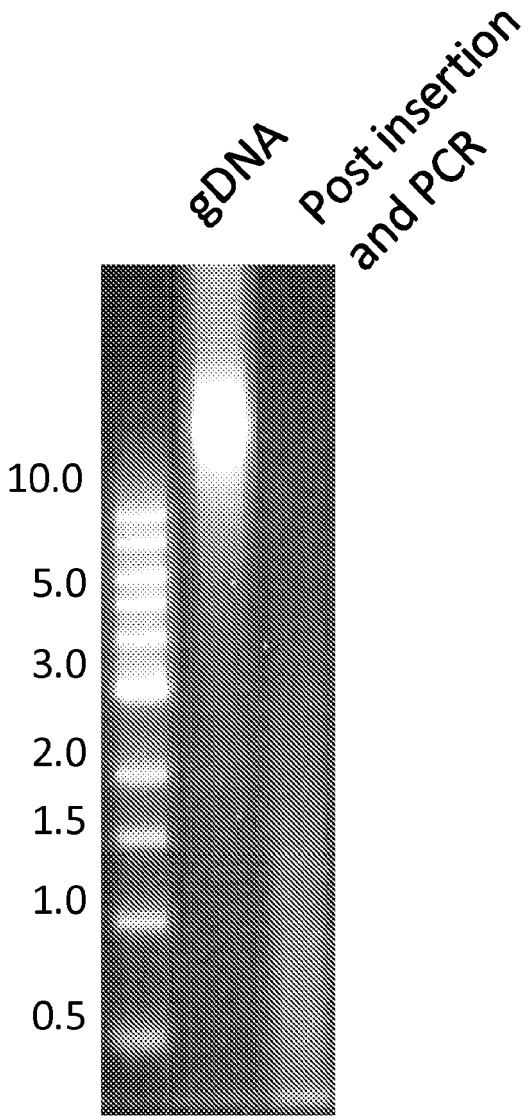


Figure 16

PARSE EX.1007.271

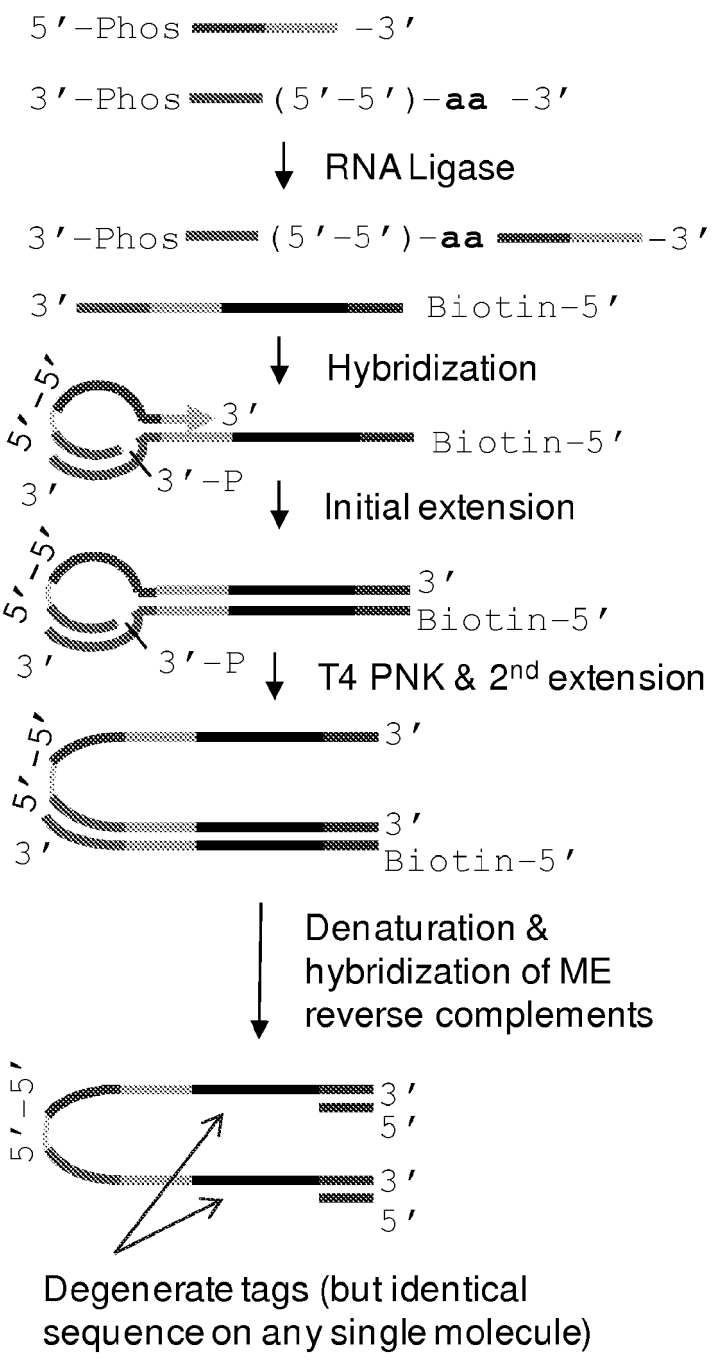


Figure 17

PARSE EX.1007.272

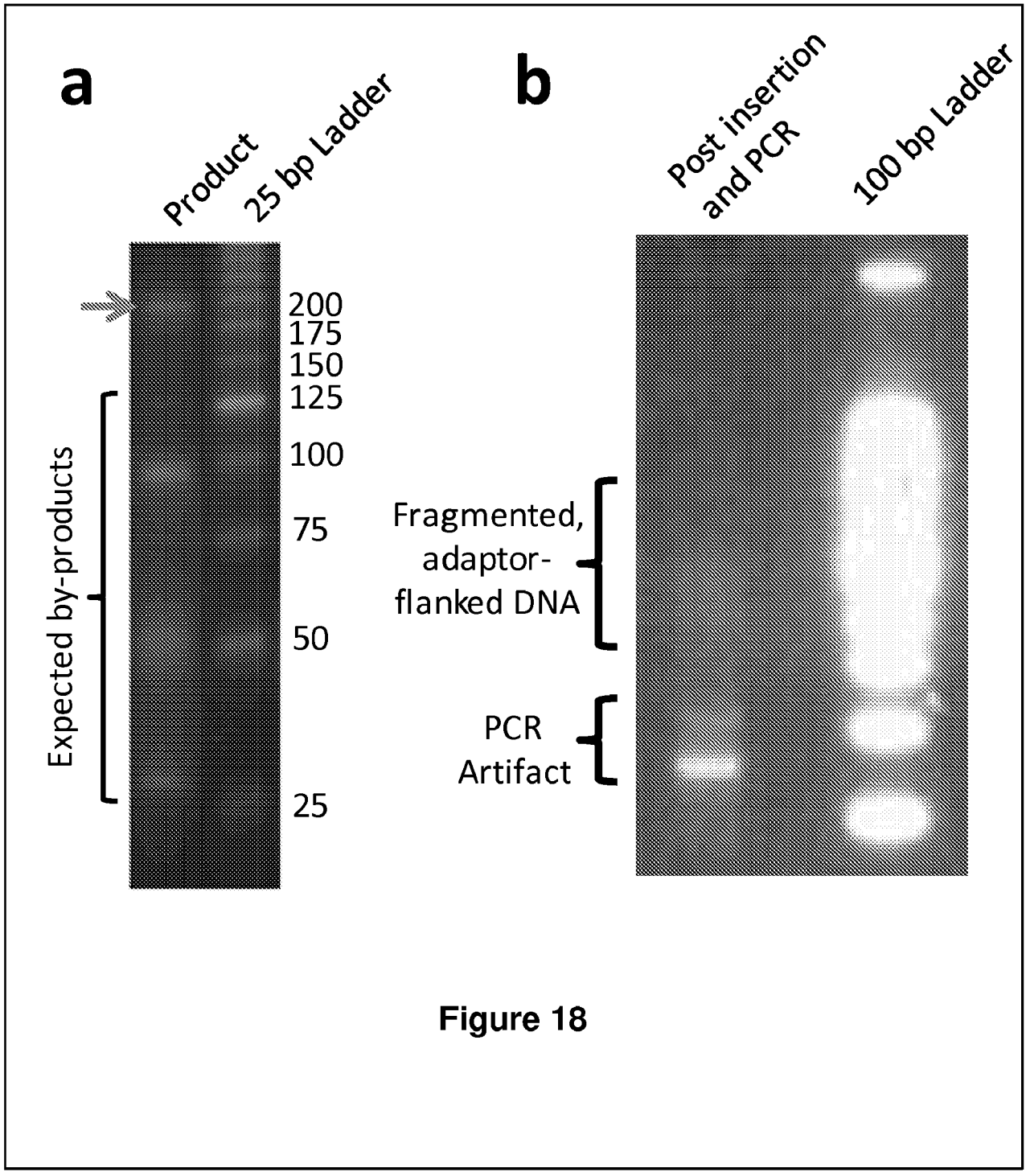


Figure 18

PARSE EX.1007.273

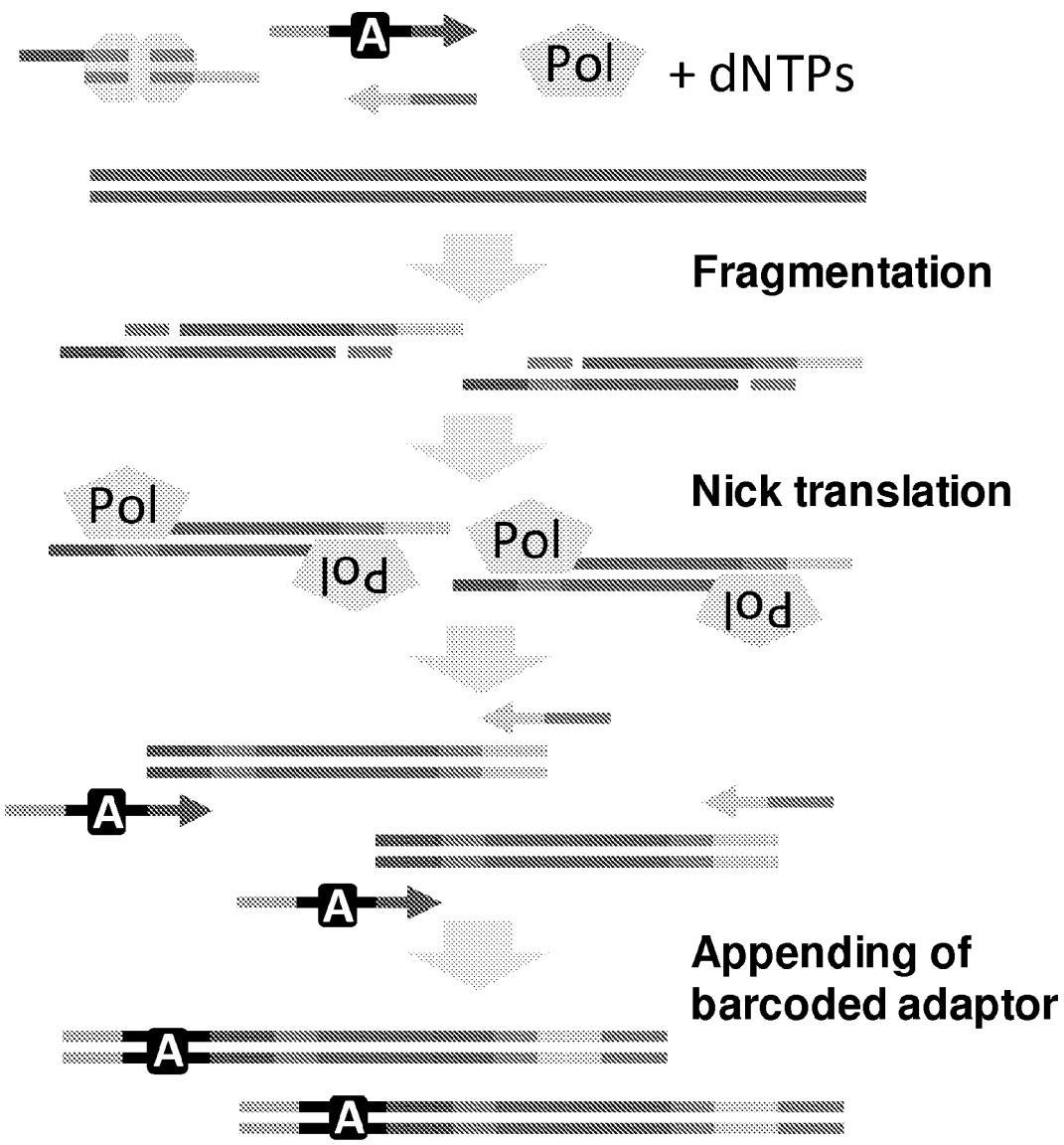


Figure 19

PARSE EX.1007.274

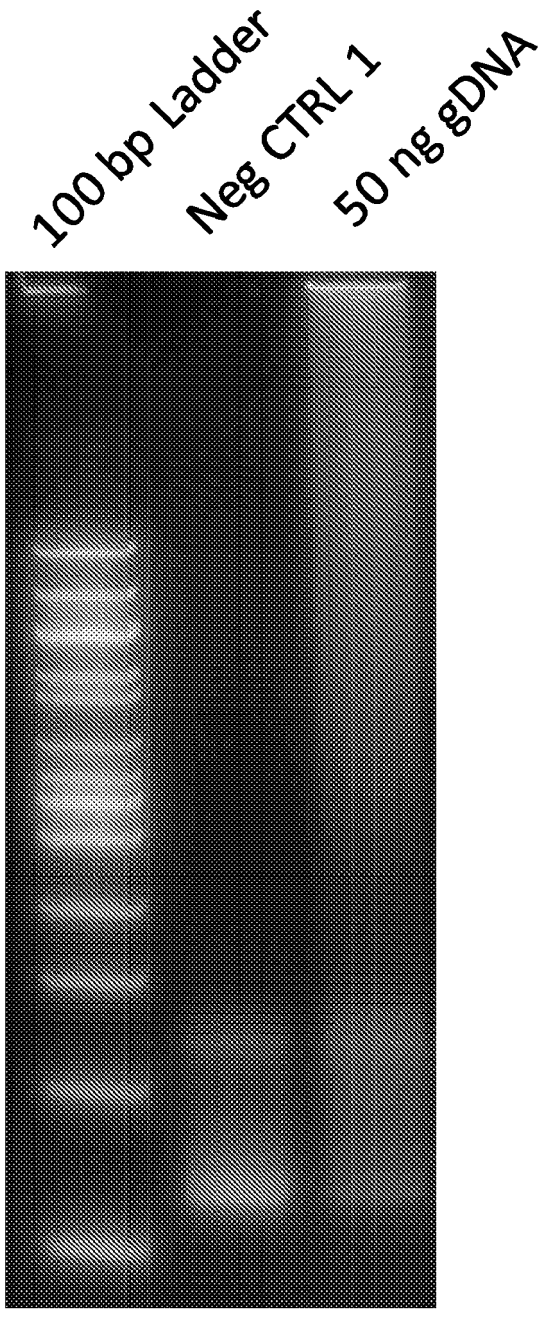


Figure 20

PARSE EX.1007.275

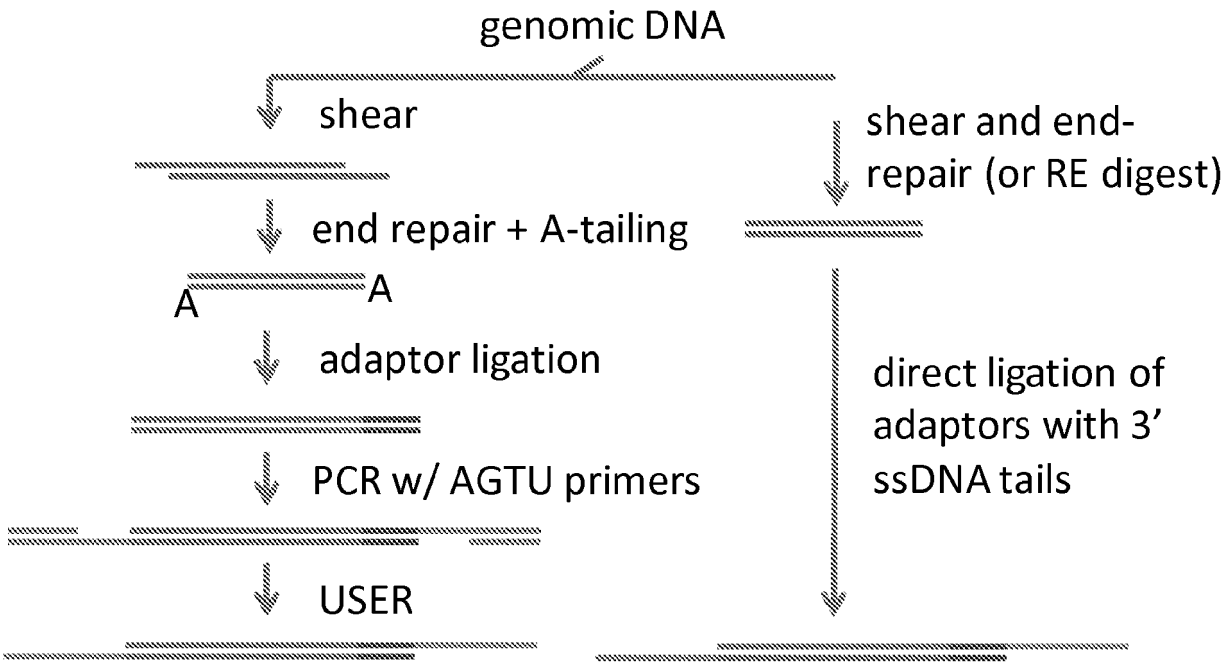


Figure 21

PARSE EX.1007.276

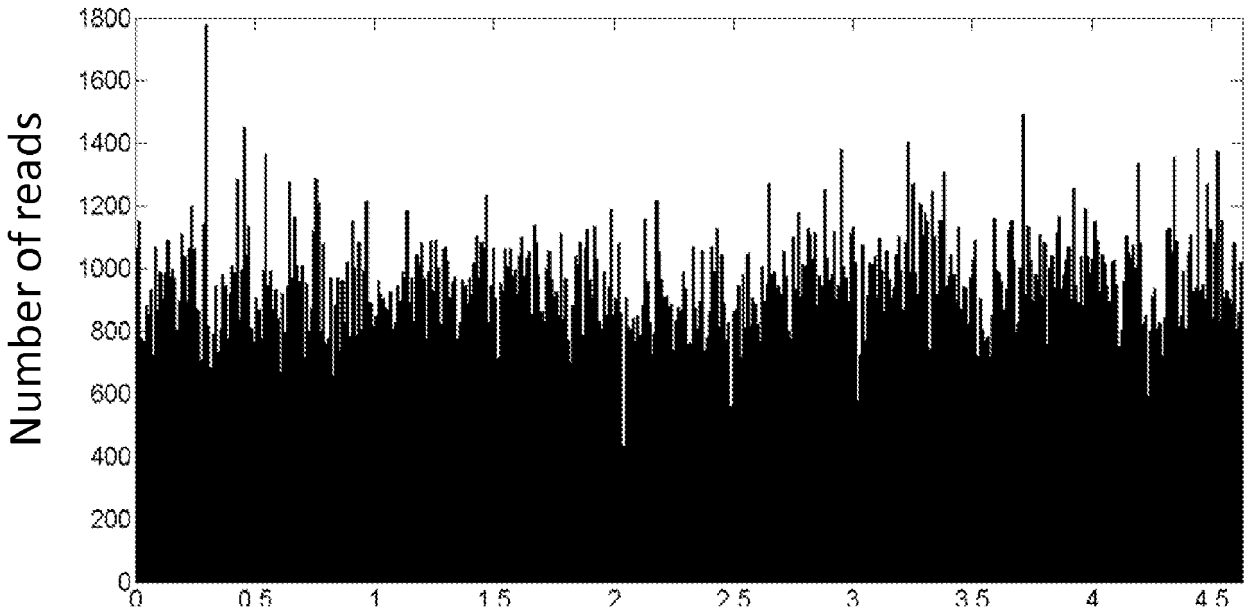


Figure 22

PARSE EX.1007.277

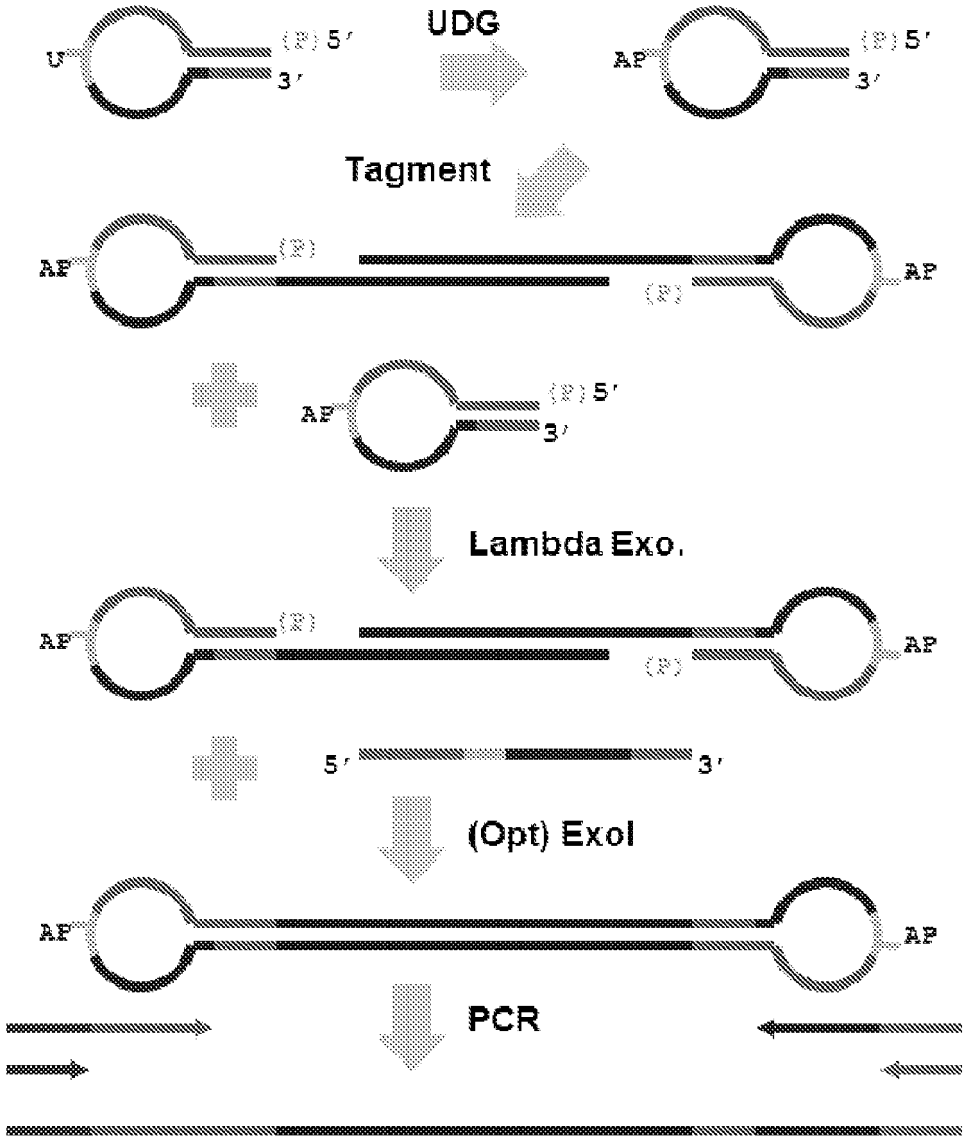


Figure 23

PARSE EX.1007.278

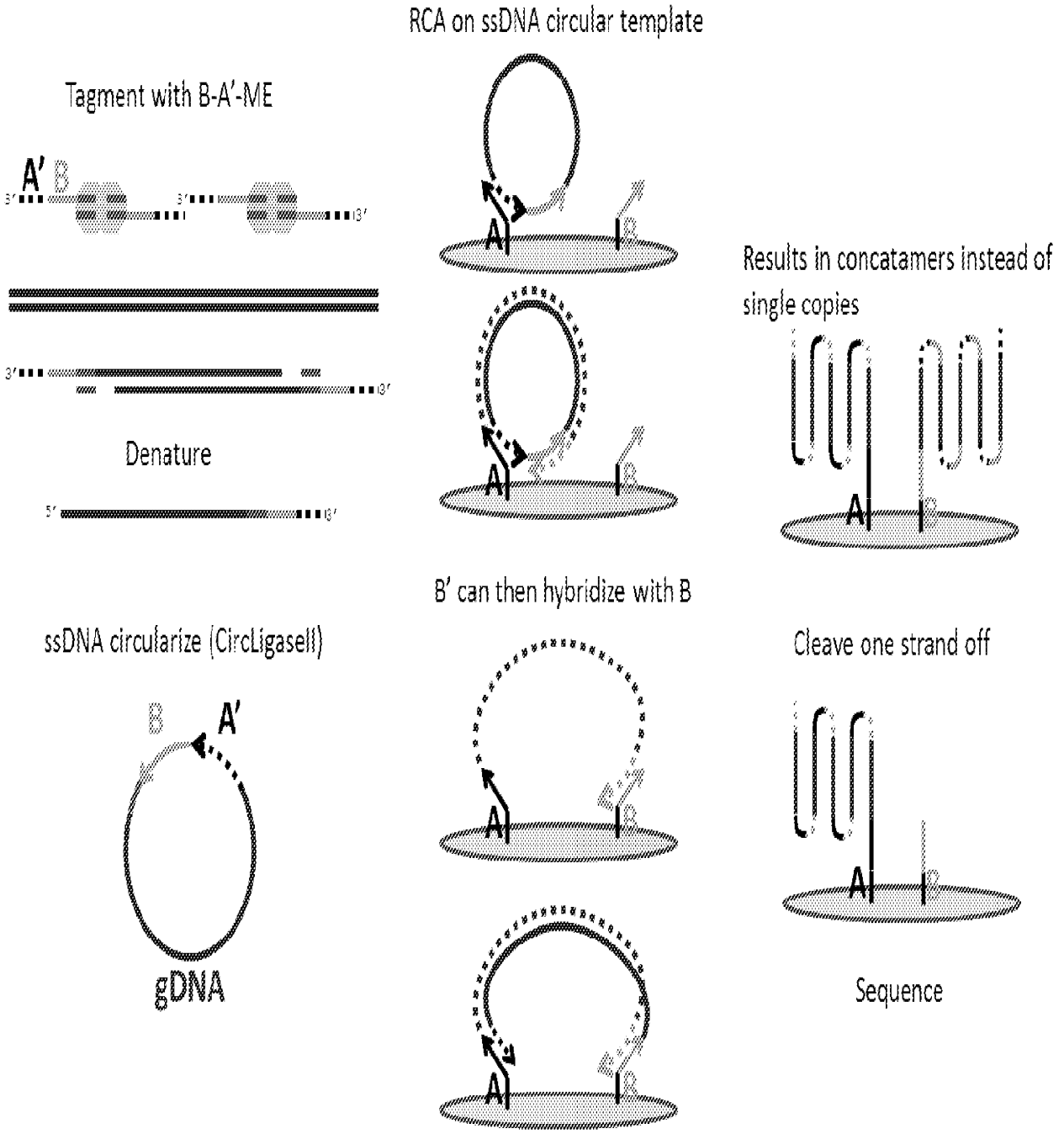


Figure 24

PARSE EX.1007.279

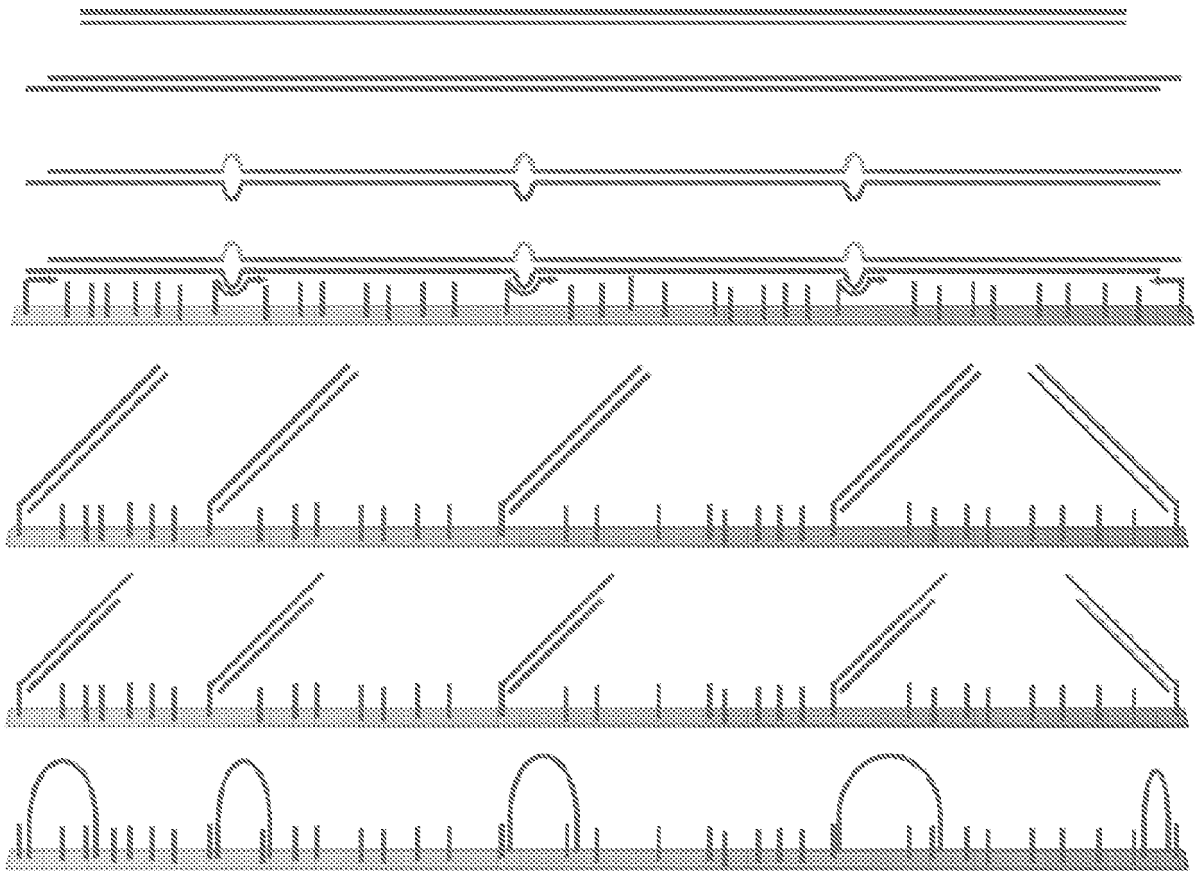


Figure 25

PARSE EX.1007.280

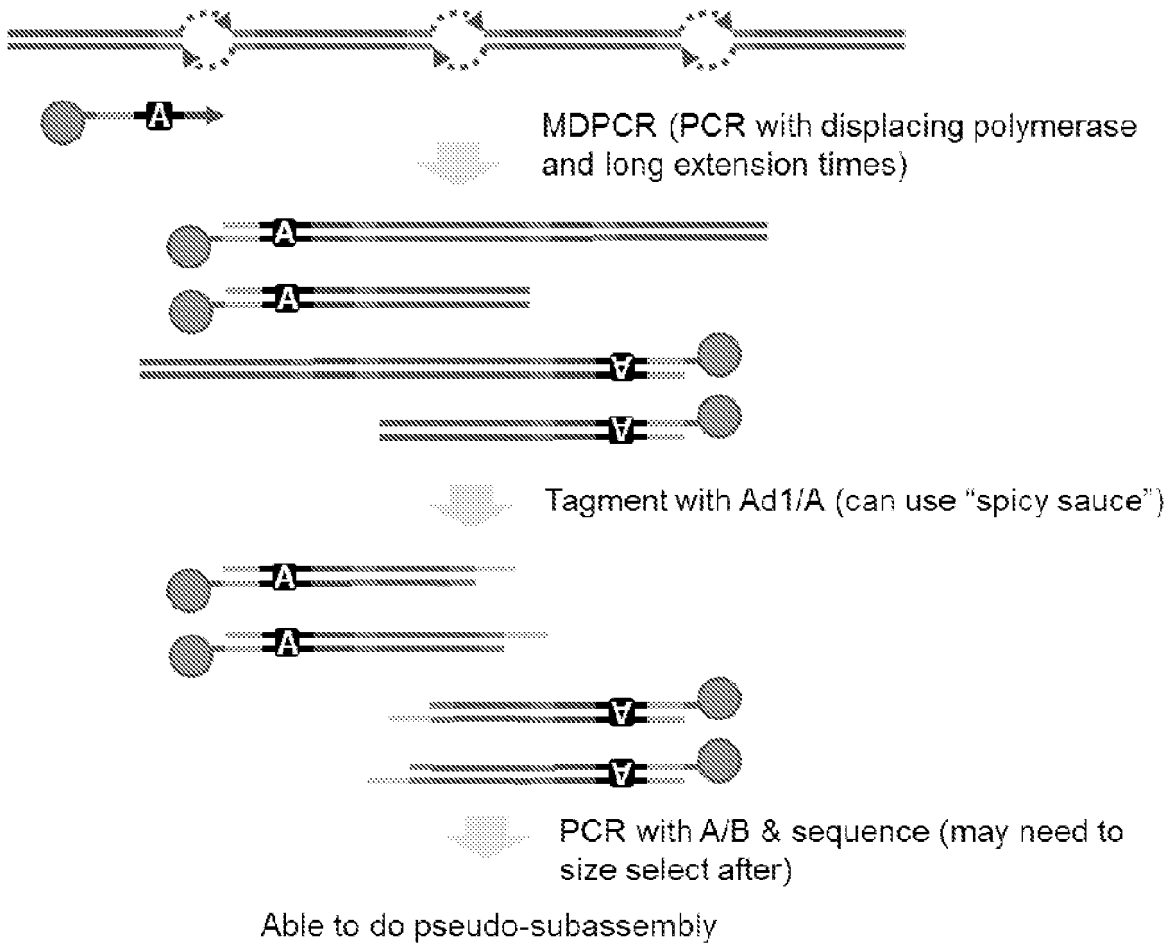


Figure 26

PARSE EX.1007.281

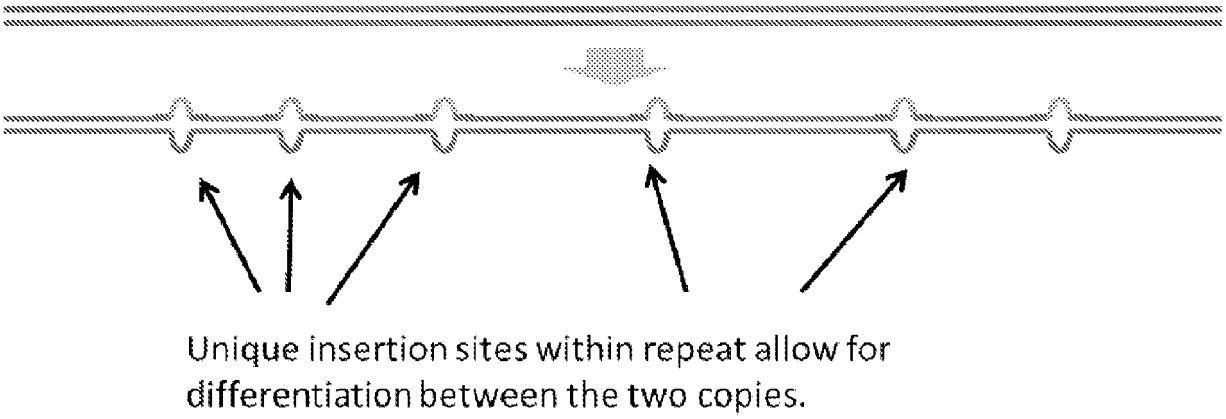


Figure 27

PARSE EX.1007.282

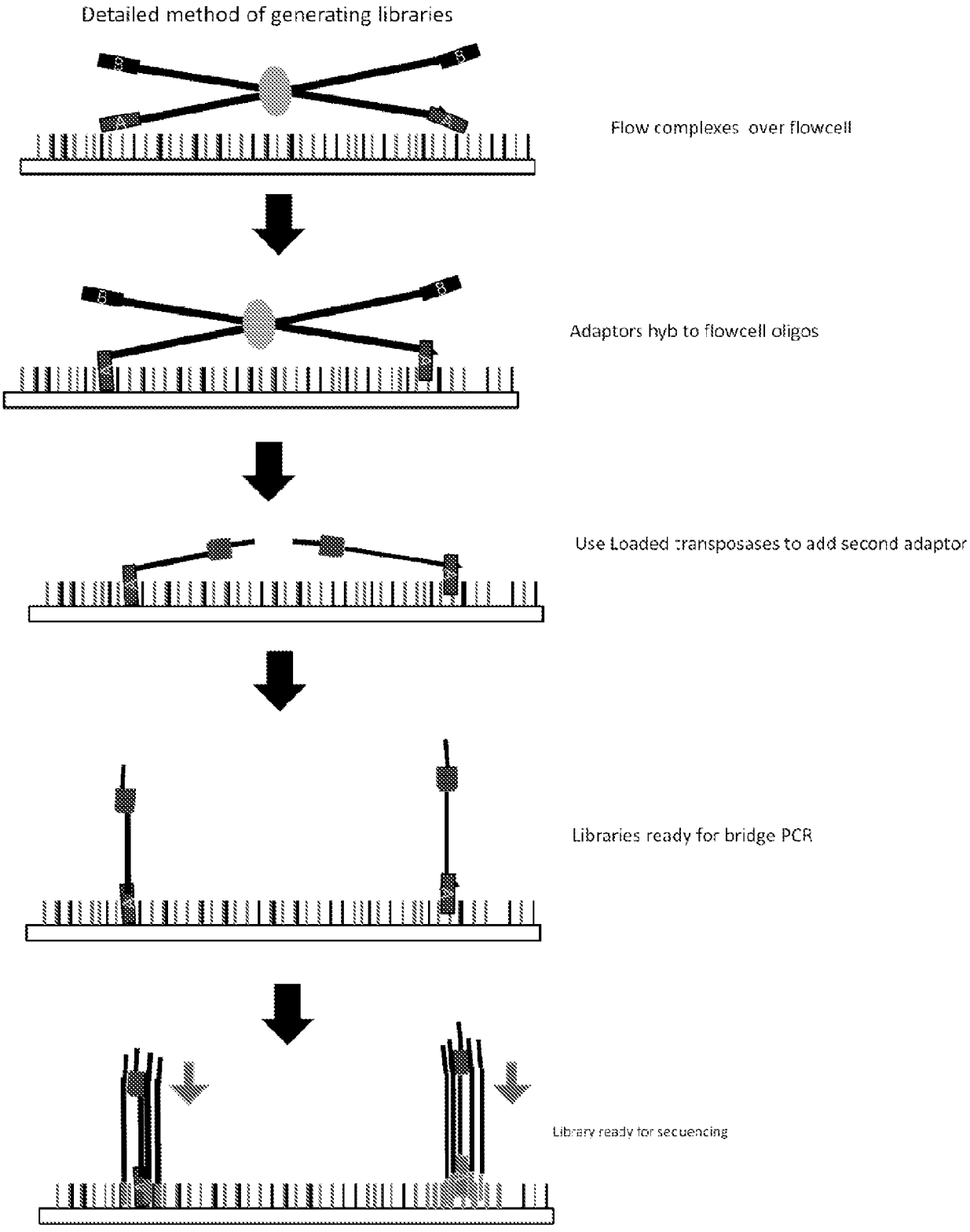


Figure 28

PARSE EX.1007.283

chr10
chr5
Immunoprecipitate DNA with Antibody
chr10
chr5
Remove other binding partners by sonication
Ligate adaptors
A + B
Nextera fragmentation followed by making clusters
"Clusters" of clusters
"Cluster" reads based on proximity (e.g. nearest 2 or 3 neighbors)
Identify new cis- and trans interactions

Figure 29

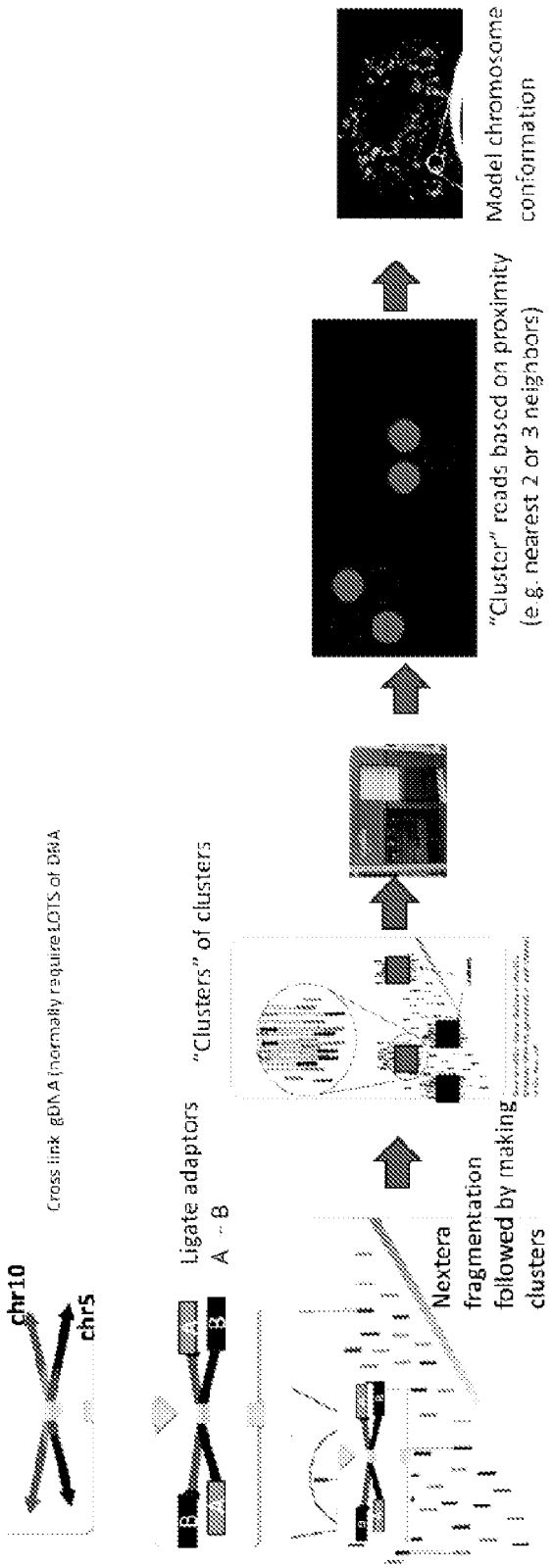


Figure 30

PARSE EX.1007.285

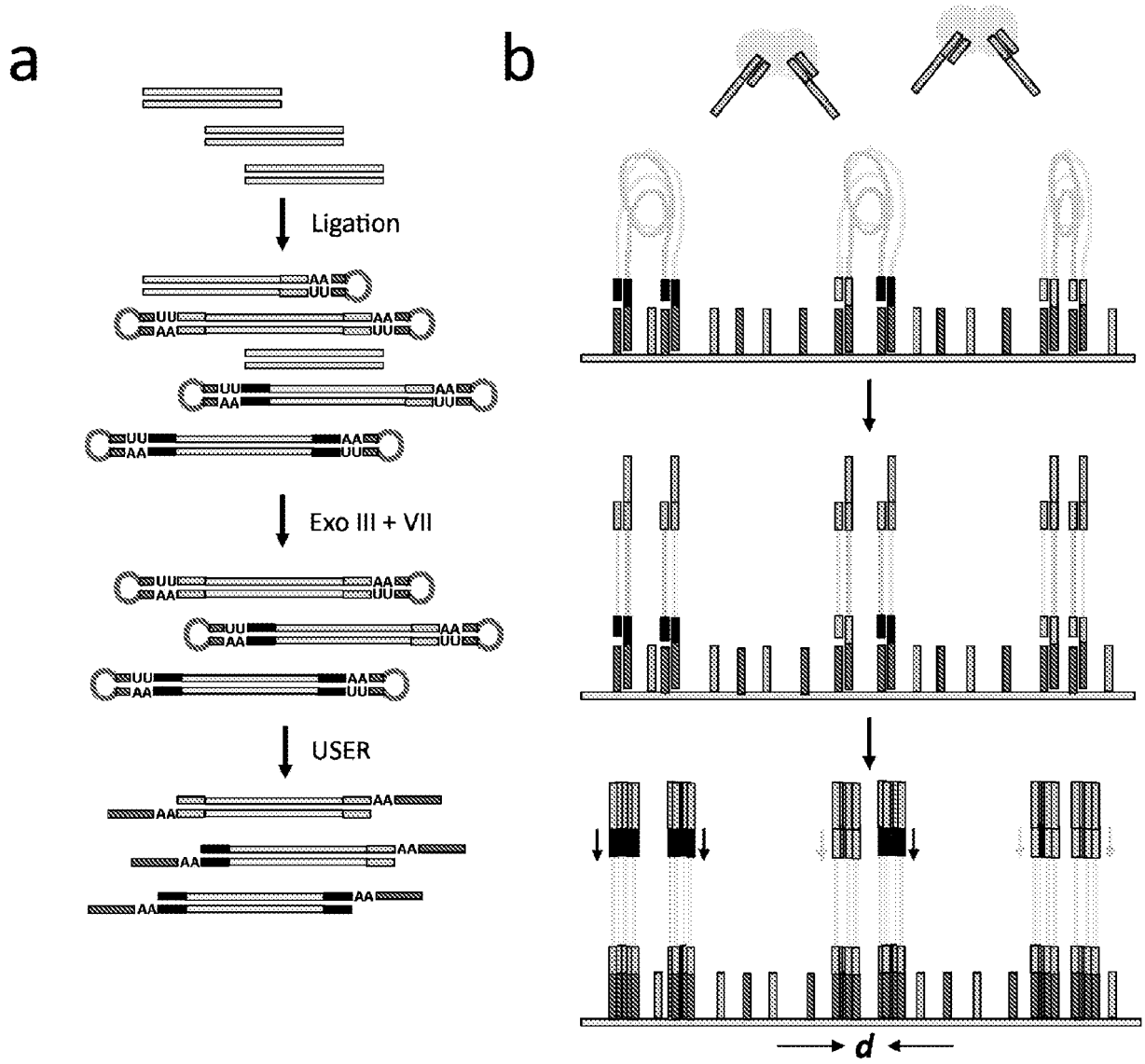


Figure 31

PARSE EX.1007.286

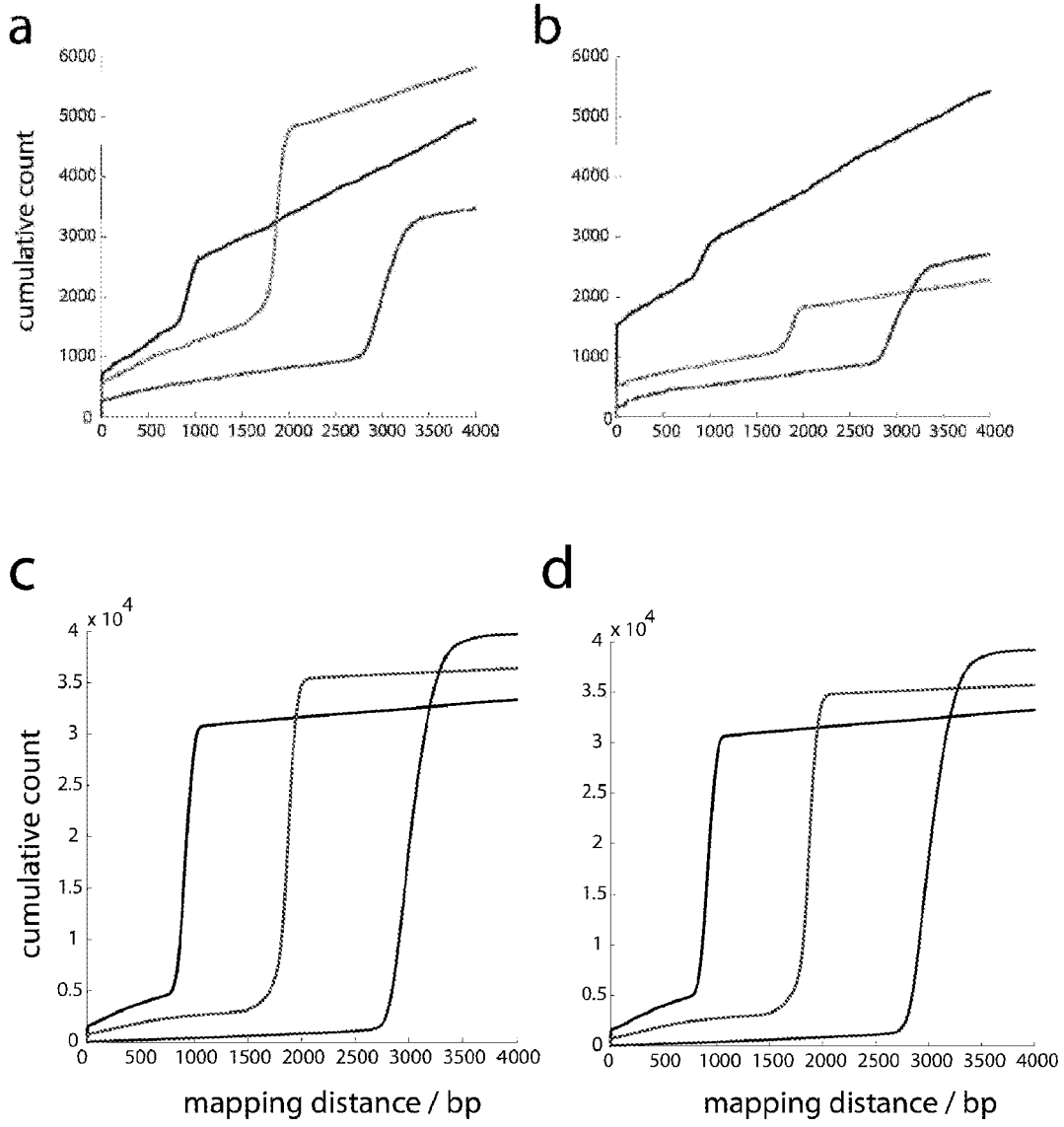


Figure 32

PARSE EX.1007.287

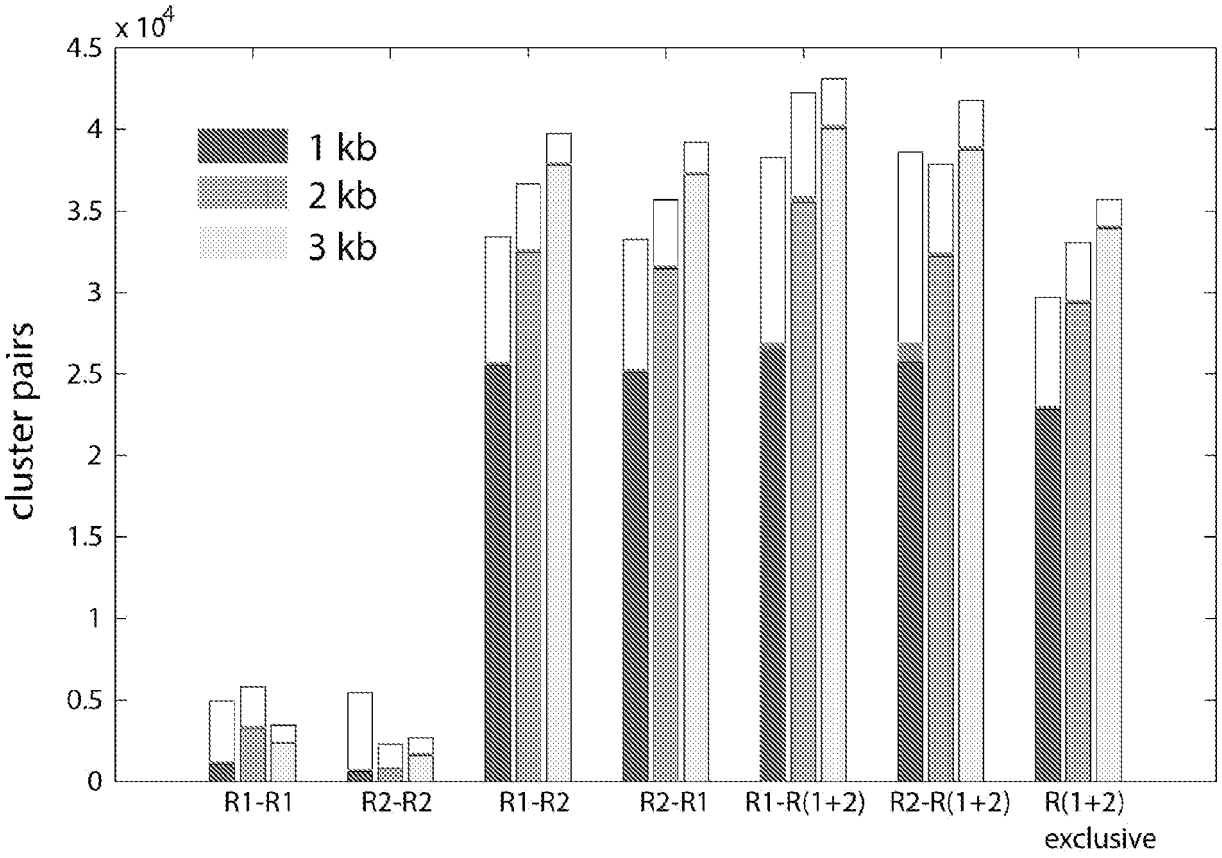


Figure 33

PARSE EX.1007.288

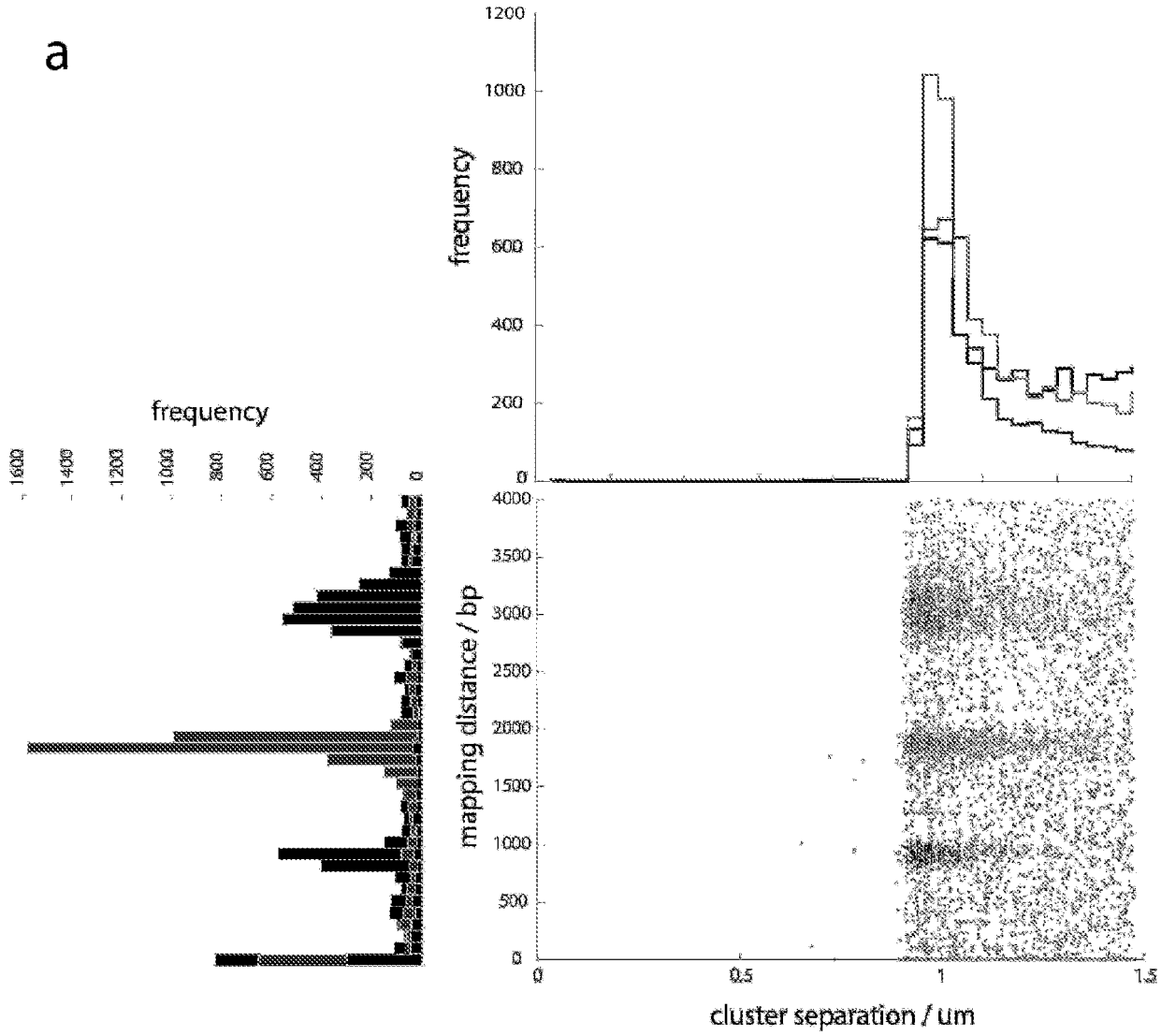


Figure 34a

PARSE EX.1007.289

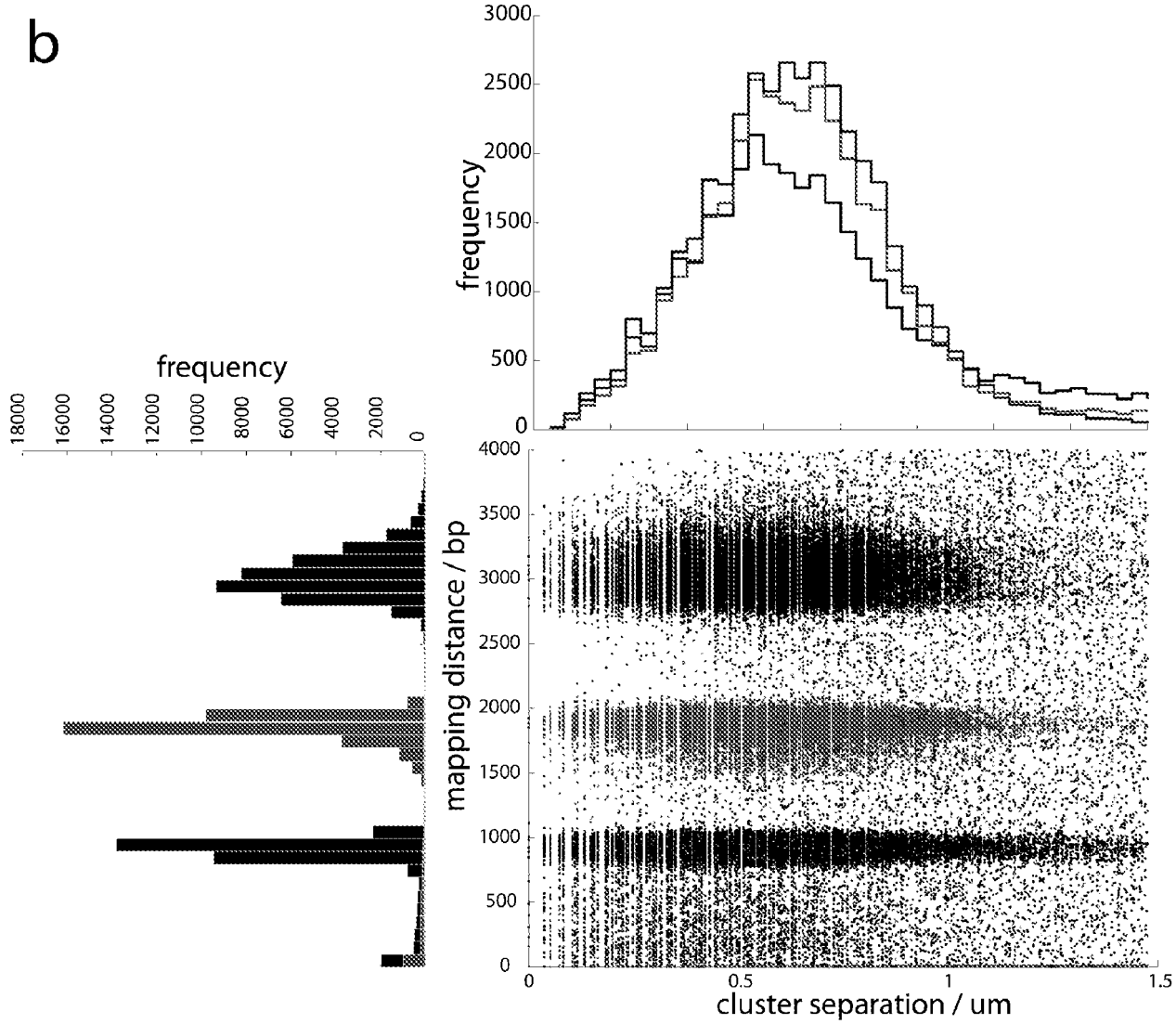


Figure 34b

PARSE EX.1007.290

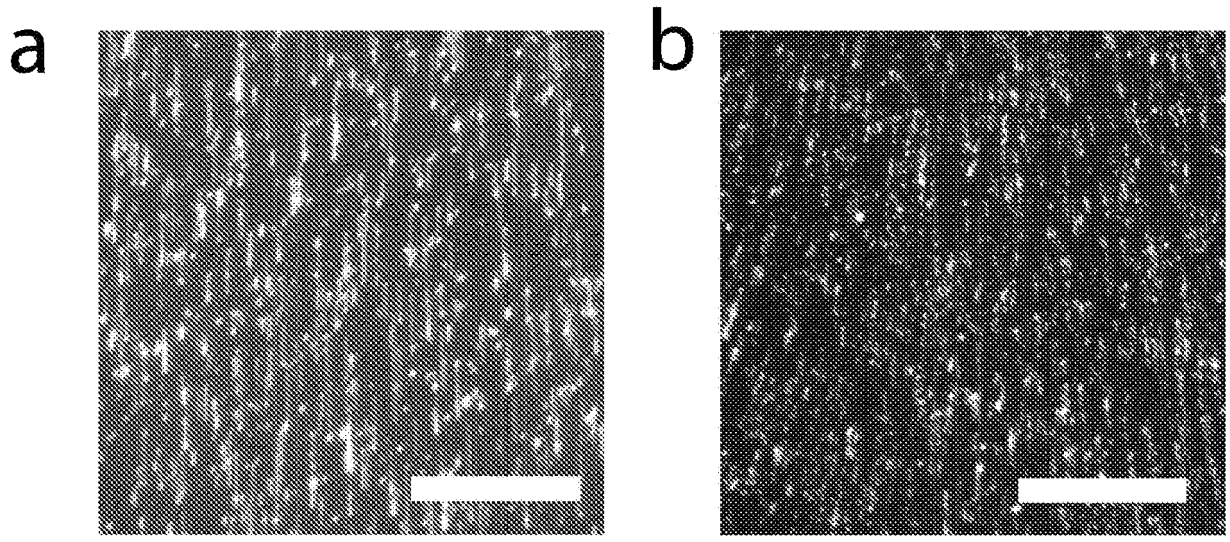


Figure 35

PARSE EX.1007.291

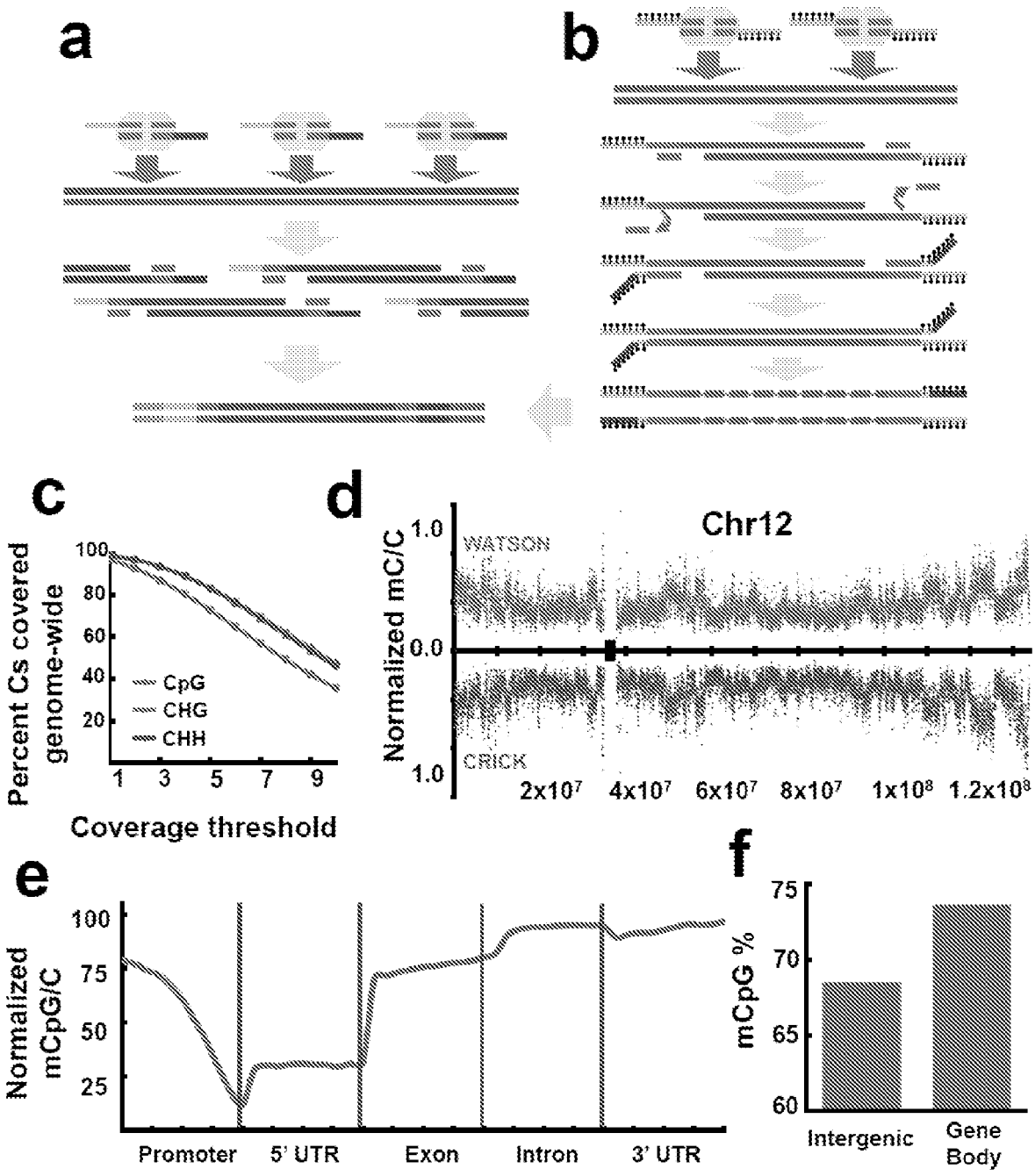


Figure 36

PARSE EX.1007.292

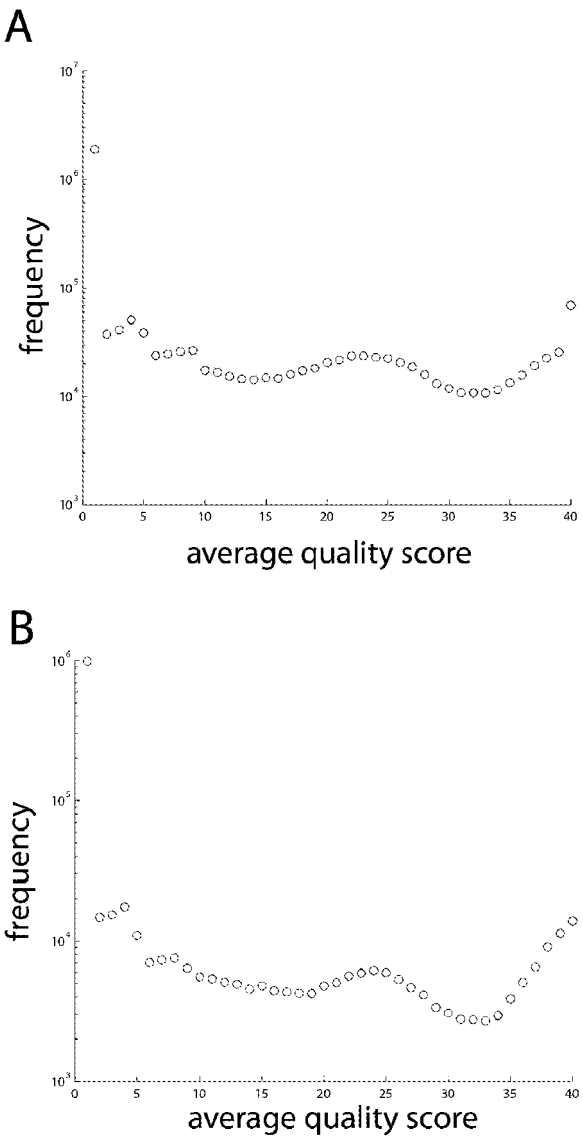


Figure 37

PARSE EX.1007.293

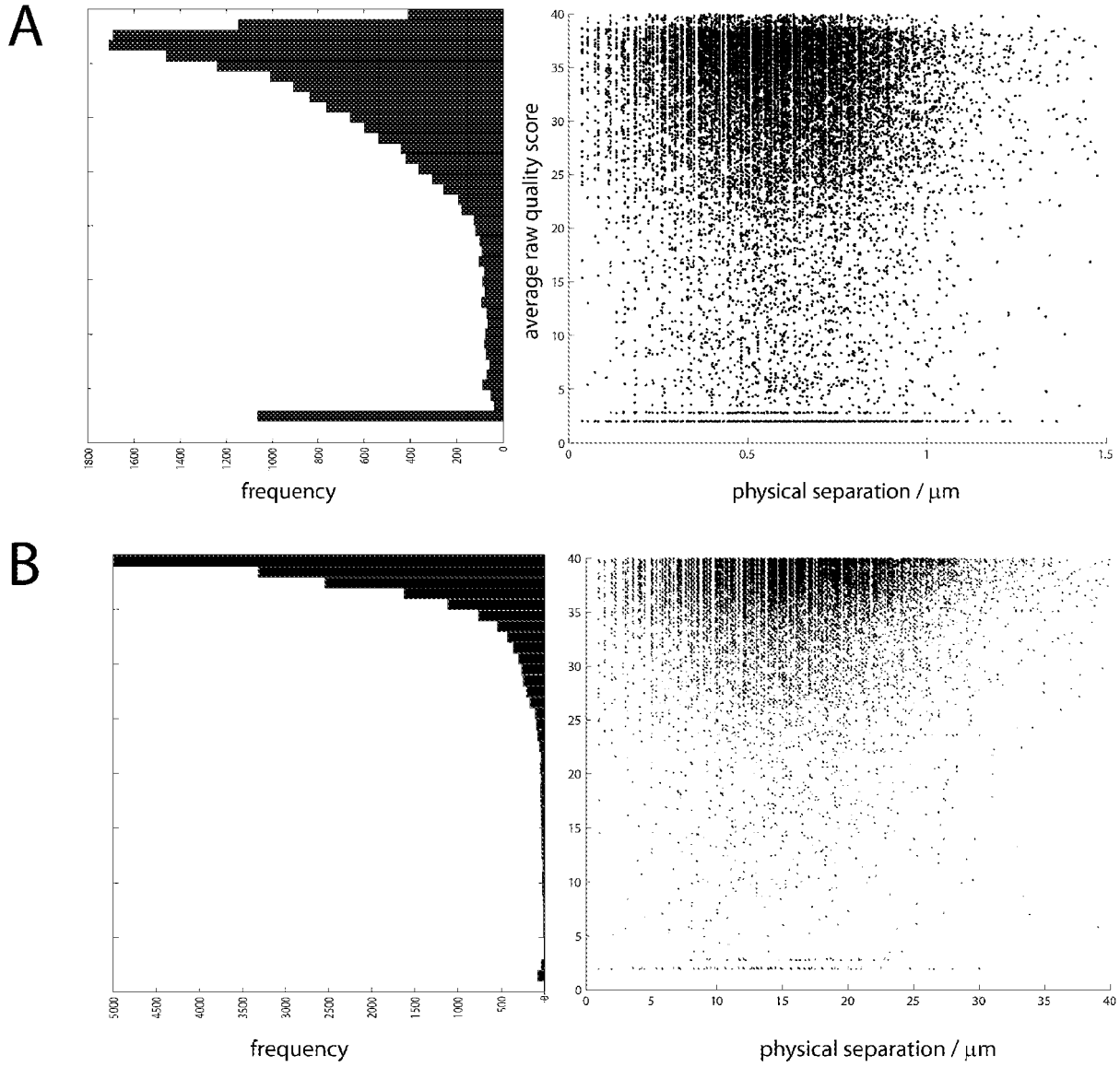


Figure 38

PARSE EX.1007.294

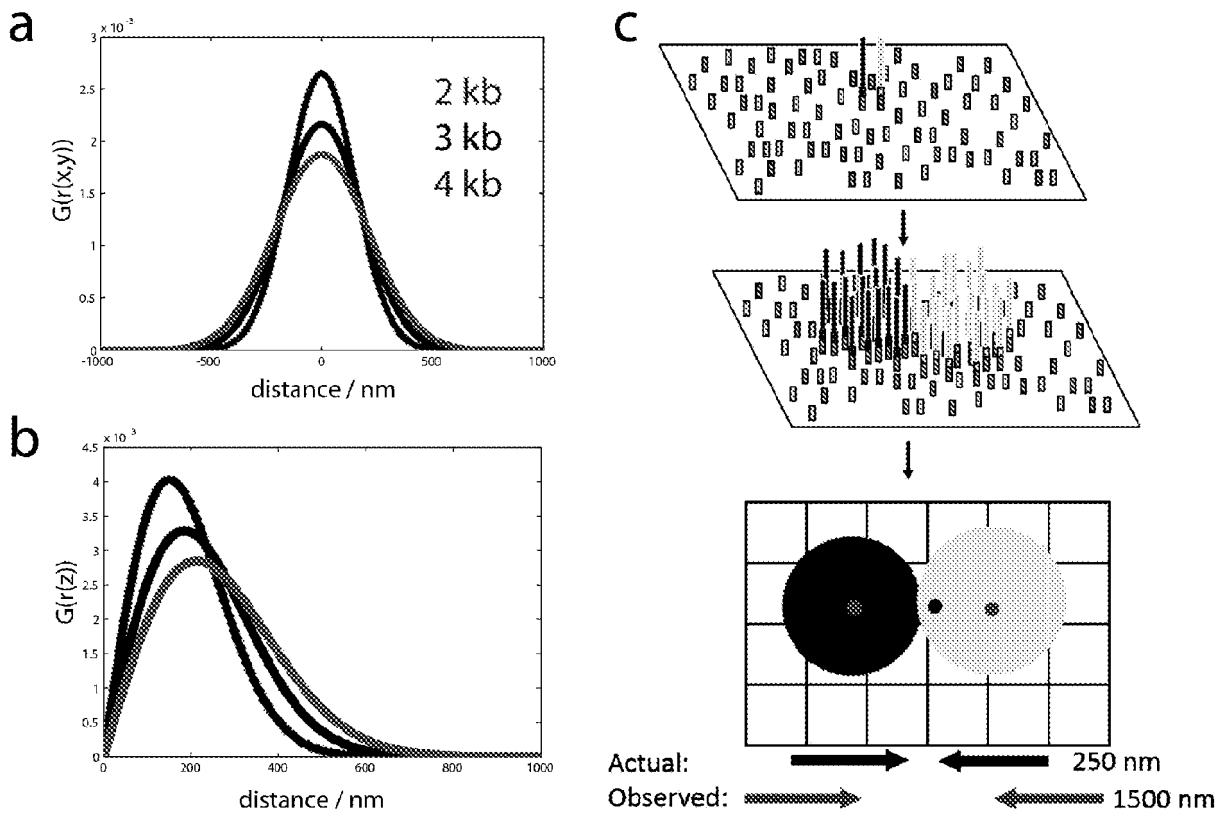


Figure 39

PARSE EX.1007.295

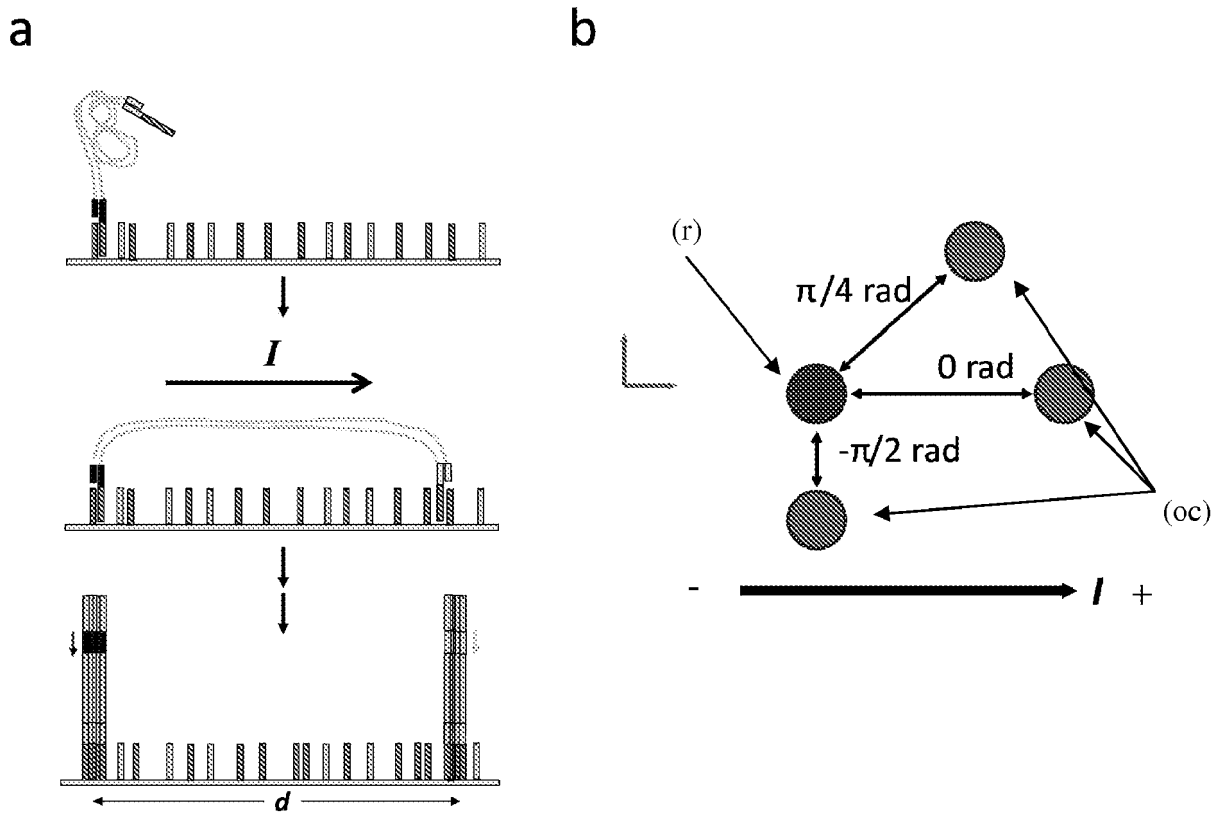


Figure 40

PARSE EX.1007.296

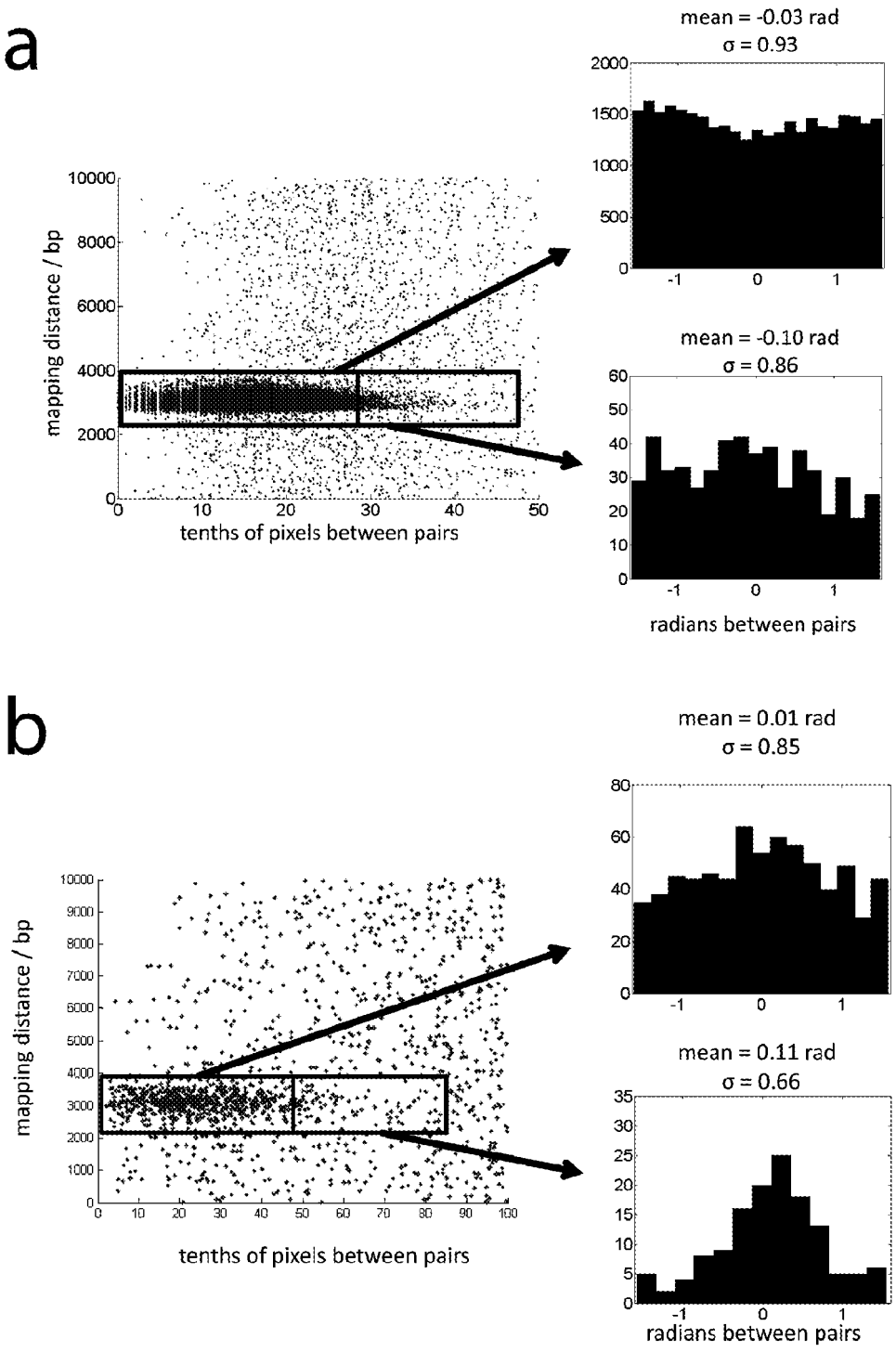


Figure 41

PARSE EX.1007.297

Espacenet

Bibliographic data: CN102292455 (A) — 2011-12-21

CHROMATIN STRUCTURE DETECTION

Inventor(s):

Applicant(s):

Classification:	**- international:**	***C12Q1/68***
	- cooperative:	**C12Q1/68** more
Application number:	CN20098155221 20091113	
Priority number(s):	WO2009US64425 20091113 ; US20080119280P 20081202	
Also published as:	CA2749570 (A1) EP2352852 (A1) EP2352852 (A4) JP2012510294 (A) JP5685195 (B2) more	

Abstract not available for CN102292455 (A)
Abstract of corresponding document: US2010136559 (A1)

The present application provides methods and compositions for determining accessibility of a DNA modifying agent in genomic DNA.

PARSE EX.1007.298

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102292455 A
(43)申请公布日 2011.12.21

(21)申请号 200980155221.3

(22)申请日 2009.11.13

(30)优先权数据
61/119,280 2008.12.02 US

(85)PCT申请进入国家阶段日
2011.07.22

(86)PCT申请的申请数据
PCT/US2009/064425 2009.11.13

(87)PCT申请的公布数据
WO2010/065266 EN 2010.06.10

(71)申请人 伯乐实验室公司
地址 美国加利福尼亚州

(72)发明人 史蒂夫·奥基诺 程曼

(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 王旭

(51)Int.Cl.
C12Q 1/68(2006.01)

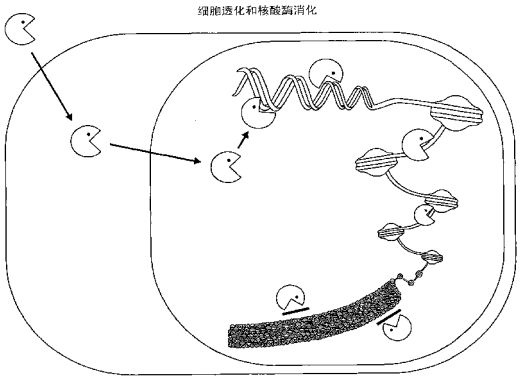
权利要求书 3 页 说明书 24 页 附图 21 页

(54)发明名称

染色质结构检测

(57)摘要

本发明提供了用于测定 DNA 修饰剂在基因组 DNA 中的可接近性的方法和组合物。



CN 102292455 A

PARSE EX.1007.299

1. 一种用于分析染色体 DNA 的方法，该方法包括：

a. 同时地：

i. 透化或破坏细胞的细胞膜；和

ii 使所述细胞与 DNA 切割剂或修饰剂在所述试剂切割或修饰所述细胞中基因组 DNA 的条件下接触，其中所述基因组 DNA 不同的区被所述试剂切割或修饰至不同的程度，从而产生切割的和完整的 DNA 区，或修饰的和未修饰的 DNA 区；和

b. 检测至少一个完整的或未修饰的或修饰的 DNA 区的量或对至少一个完整的或未修饰的或修饰的 DNA 区进行克隆、分离或核苷酸测序。

2. 权利要求 1 所述的方法，包括将所述至少一个完整的或未修饰的或修饰的 DNA 区的量与所述细胞中所述 DNA 区的染色质结构相关联。

3. 权利要求 1 所述的方法，包括

检测至少第一染色体完整的或未修饰的或修饰的 DNA 区和第二染色体完整的或未修饰的或修饰的 DNA 区的量；和

比较所述第一和第二 DNA 区的量。

4. 权利要求 1 所述的方法，包括定量第一染色体 DNA 区和第二染色体 DNA 区的完整拷贝数，从而评估所述第一和第二 DNA 区与所述 DNA 切割剂的相对可接近性。

5. 权利要求 1 所述的方法，其中在所述接触步骤之后和所述检测步骤之前分离所述基因组 DNA。

6. 权利要求 1 所述的方法，其中在所述细胞直接或间接地附着于人工培养表面时，所述细胞被透化并与所述 DNA 切割剂或修饰剂接触。

7. 权利要求 1 所述的方法，其中所述透化步骤包括使所述细胞与透化细胞膜的试剂接触。

8. 权利要求 7 所述的方法，其中所述透化细胞膜的试剂是溶血脂质。

9. 权利要求 8 所述的方法，其中所述溶血脂质是溶血卵磷脂。

10. 权利要求 1 所述的方法，其中所述透化或破坏步骤包括用非离子型洗涤剂破坏细胞膜。

11. 权利要求 10 所述的方法，其中所述非离子型洗涤剂选自 NP40、Tween20 和 Triton X-100 的组。

12. 权利要求 1 所述的方法，其中所述细胞与 DNA 切割剂接触，且所述 DNA 切割剂是酶。

13. 权利要求 1 所述的方法，其中所述细胞与 DNA 修饰剂接触，且所述 DNA 修饰剂选自由甲基转移酶和修饰 DNA 的化学制剂组成的组。

14. 权利要求 12 所述的方法，其中

所述 DNA 切割剂是 DNA 切割酶；

所述基因组 DNA 中的修饰是这样的位点，在所述位点处 DNA 已经被切割；和

所述检测步骤包括定量切割后 DNA 区的完整拷贝的量。

15. 权利要求 14 所述的方法，其中所述检测步骤包括定量扩增。

16. 权利要求 15 所述的方法，其中定量扩增选自由定量聚合酶链反应和定量连接介导聚合酶链反应组成的组。

17. 权利要求 12 所述的方法，其中所述 DNA 切割剂选自 DNase 和限制酶。

PARSE EX.1007.300

18. 权利要求 13 所述的方法，其中：

所述 DNA 修饰剂是甲基转移酶；

所述修饰是不被所述细胞的天然甲基化酶甲基化的核苷酸序列中核苷酸上的甲基；和

所述检测步骤包括检测所述修饰存在或不存在的方法。

19. 权利要求 18 所述的方法，其中所述甲基转移酶是 DAM 甲基转移酶并且所述定量步骤包括用限制酶切割所述修饰的 DNA，所述限制酶识别被 DAM 甲基转移酶甲基化的 DNA 序列。

20. 权利要求 18 所述的方法，其中所述甲基转移酶将甲基部分添加至 DNA 中的胞嘧啶，并且所述检测步骤包括用甲基化特异性限制酶切割所述修饰的 DNA 和 / 或用亚硫酸氢盐处理所述修饰的 DNA。

21. 权利要求 1 所述的方法，其中所述 DNA 修饰剂是相对于染色质结构具有空间位阻的分子，其中所述分子连接修饰 DNA 的化学制剂。

22. 权利要求 21 所述的方法，其中所述修饰 DNA 的化学制剂选自由硫酸二甲酯和肼组成的组。

23. 权利要求 1 所述的方法，其中

所述检测步骤包括定量靶 DNA 区的至少一部分和对照 DNA 区的至少一部分，其中所述对照 DNA 区包含下列序列，所述序列

i. 在动物的基本上所有细胞中是可接近的；或

ii 在动物的大多数细胞中是不可接近的。

24. 权利要求 1 所述的方法，其中所述对照 DNA 区使用下列每一种进行定量扩增：

引发所述对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分不包括所述 DNA 修饰剂的可能修饰位点；和

引发所述对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分确实包括所述 DNA 修饰剂的至少一个可能修饰位点。

25. 权利要求 1 所述的方法，其中确定所述 DNA 区的 DNA 甲基化。

26. 权利要求 25 所述的方法，其中步骤 a 包括使所述 DNA 与 DNA 切割剂接触，并且步骤 b 包括使所述完整的 DNA 与亚硫酸氢盐接触。

27. 权利要求 26 所述的方法，还包括确定亚硫酸氢盐处理的 DNA 的解链温度，并且将所述解链温度与 DNA 甲基化的存在、不存在或程度相关联。

28. 权利要求 1 所述的方法，其中步骤 b 包括制备下列的文库：

完整 DNA 区；或

或修饰的 DNA 区；或

未修饰的 DNA 区。

29. 权利要求 1 所述的方法，其中步骤 b 包括使至少一个完整的或未修饰的或修饰的 DNA 区与多核苷酸文库在允许所述 DNA 区与所述文库的一个或多个成员杂交的条件下接触，并且检测所述 DNA 区与所述一个或多个成员的杂交。

30. 权利要求 29 所述的方法，其中所述文库组织在微阵列上。

31. 权利要求 1 所述的方法，其中步骤 b 包括扩增至少一个完整的或未修饰的或修饰的 DNA 区。

PARSE EX.1007.301

32. 权利要求 1 所述的方法，其中步骤 b 包括对至少一个完整的或未修饰的或修饰的 DNA 区进行核苷酸测序。

33. 一种用于确定染色体上的基因座与 DNA 修饰剂的可接近性的试剂盒，所述试剂盒包含

细胞膜透化剂或破坏剂；

限制酶或 DNase。

34. 权利要求 33 所述的试剂盒，其中所述细胞膜透化剂或破坏剂和所述限制酶或 DNase 处于同一容器中的同一缓冲液中。

35. 权利要求 33 所述的试剂盒，其中所述细胞膜透化剂或破坏剂和所述限制酶或 DNase 处于单独的容器中。

36. 权利要求 33 所述的试剂盒，还包含亚硫酸氢盐。

37. 权利要求 33 所述的试剂盒，还包含用于扩增真核生物的基因组 DNA 的区的引物对。

38. 权利要求 33 所述的试剂盒，包含：

a. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分不包括所述 DNA 修饰剂的可能修饰位点；和 / 或

b. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分确实包括所述 DNA 修饰剂的至少一个可能修饰位点。

39. 权利要求 33 所述的试剂盒，包含：

a. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分不能接近所述修饰剂；和 / 或

b. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分可接近所述修饰剂。

PARSE EX.1007.302

染色质结构检测

[0001] 相关专利申请的交叉引用

[0002] 本申请要求于 2008 年 12 月 2 日提交的美国临时专利申请号 61/119,280 的优先权，该申请通过引用结合用于所有目的。

[0003] 发明背景

[0004] 在已知为核小体的结构中细胞中的大多数 DNA 包装在一组组蛋白周围。此核小体 DNA 可以进一步被包装进紧密压紧 DNA 的卷曲结构中。此紧密包装可以限制 DNA 接触转录因子和转录机器。以此方式包装的基因组 DNA 有时被称为染色质 (chromatin)。

[0005] 染色质分为两类，常染色质和异染色质，在常染色质中 DNA 疏松地包装，是可接近的并且通常但不总是有转录活性的，在异染色质中 DNA 是紧密包装的，不可接近的且通常但不总是转录沉默的。

[0006] 控制这两种染色质状态之间转变的是表观遗传学。存在两种主要的表观遗传学事件 :DNA 甲基化和组蛋白修饰。这些事件影响 DNA 如何被包装以及 DNA 是否是有转录活性的或转录沉默的。

[0007] 发明的简述

[0008] 本发明提供了分析染色体 DNA 的方法，包括但不限于，确定染色体上的 DNA 区与 DNA 修饰剂的可接近性，任选地将所述可接近性与染色质结构相关联。在一些实施方案中，该方法包括 :

[0009] a. 同时地 :

[0010] i. 透化或破坏细胞的细胞膜 ;和

[0011] ii 使所述细胞与 DNA 切割剂或修饰剂在一定的条件下接触，所述条件使得所述试剂切割或修饰所述细胞中的基因组 DNA，其中所述基因组 DNA 的不同区域被所述试剂切割或修饰至不同的程度，从而产生切割的和完整的 DNA 区，或修饰的和未修饰的 DNA 区 ;和

[0012] b. 检测至少一个完整的或未修饰的或修饰的 DNA 区的物理特性或量或对至少一个完整的或未修饰的或修饰的 DNA 区进行克隆、分离或核苷酸测序。

[0013] 在一些实施方案中，该方法还包括将所述量与所述细胞中所述 DNA 区的染色质结构相关联。

[0014] 在一些实施方案中，该方法包括检测至少第一染色体完整的或未修饰的或修饰的 DNA 区和第二染色体完整的或未修饰的或修饰的 DNA 区的物理特性或量 ;和比较所述第一和第二 DNA 区的物理特性或量。

[0015] 在一些实施方案中，该方法包括定量第一染色体 DNA 区和第二染色体 DNA 区的完整拷贝数，从而评估所述第一和第二 DNA 区与所述 DNA 切割剂的相对可接近性。

[0016] 在一些实施方案中，在所述接触步骤之后和检测步骤之前分离所述基因组 DNA。

[0017] 在一些实施方案中，所述细胞被透化并与所述 DNA 切割剂或修饰剂接触，同时所述细胞直接或间接地附着于人工培养表面。

[0018] 在一些实施方案中，所述透化步骤包括使所述细胞与透化细胞膜的试剂接触。在一些实施方案中，透化细胞膜的所述试剂是溶血脂质 (lysolipid)。在一些实施方案中，所

PARSE EX.1007.303

述溶血脂质是溶血卵磷脂。

[0019] 在一些实施方案中，所述透化或破坏步骤包括用非离子型洗涤剂破坏细胞膜。在一些实施方案中，所述非离子型洗涤剂选自由 NP40、Tween20 和 Triton X-100 组成的组。

[0020] 在一些实施方案中，所述细胞与 DNA 切割剂接触，且所述 DNA 切割剂是酶。

[0021] 在一些实施方案中，所述细胞与 DNA 修饰剂接触，且所述 DNA 修饰剂选自由甲基转移酶和修饰 DNA 的化学制剂组成的组。在一些实施方案中，所述 DNA 切割剂是 DNA 切割酶；所述基因组 DNA 中的修饰是切割位点；并且所述检测步骤包括定量切割后 DNA 区的完整拷贝的量。在一些实施方案中，所述检测步骤包括定量扩增。在一些实施方案中，定量扩增选自由定量聚合酶链反应（任选地实时的）和定量连接介导聚合酶链反应（任选地实时的）组成的组。

[0022] 在一些实施方案中，所述 DNA 切割剂选自 DNase 和限制酶组成的组。在一些实施方案中，所述 DNA 修饰剂是甲基转移酶；所述修饰是不会被所述细胞的天然甲基化酶甲基化的核苷酸序列中核苷酸上的甲基；并且所述检测步骤包括检测所述修饰存在与否的方法。在一些实施方案中，所述甲基转移酶是 DAM 甲基转移酶并且所述定量步骤包括用限制酶切割所述修饰的 DNA，所述限制酶识别被 DAM 甲基转移酶甲基化的 DNA 序列。在一些实施方案中，所述甲基转移酶将甲基结构部分添加至 DNA 中的胞嘧啶，并且所述检测步骤包括用甲基化特异性限制酶切割所述修饰的 DNA 和 / 或用亚硫酸氢盐处理所述修饰的 DNA。

[0023] 在一些实施方案中，所述 DNA 修饰剂是具有相对于染色质结构的空间位阻的分子，其中所述分子连接修饰 DNA 的化学制剂。一些实施方案中，所述修饰 DNA 的化学制剂选自由硫酸二甲酯和肼组成的组。

[0024] 在一些实施方案中，所述检测步骤包括定量靶 DNA 区的至少一部分和对照 DNA 区的至少一部分，其中所述对照 DNA 区包含这样的序列，所述序列

[0025] i. 在动物的基本上所有细胞中是可接近的；或

[0026] ii 在动物的大多数细胞中是不可接近的；或

[0027] iii. 根据细胞的类型或生长环境具有可变的可接近性。

[0028] 在一些实施方案中，所述对照 DNA 区使用下列的每一种定量扩增：

[0029] 引发所述对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分不包括所述 DNA 修饰剂的可能修饰位点；和

[0030] 引发所述对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分确实包括所述 DNA 修饰剂的至少一个可能修饰位点。

[0031] 在一些实施方案中，所述物理特性是 DNA 甲基化。

[0032] 一些实施方案中，步骤 a 包括使所述 DNA 与 DNA 切割剂接触，并且步骤 b 包括使所述完整的 DNA 与亚硫酸氢盐接触。

[0033] 在一些实施方案中，该方法还包括确定亚硫酸氢盐处理的 DNA 的解链温度，并且将所述解链温度与 DNA 甲基化的存在、不存在或程度相关联。

[0034] 在一些实施方案中，步骤 b 包括制备下列的文库：完整 DNA 区；或修饰的 DNA 区；或未修饰的 DNA 区。

[0035] 在一些实施方案中，步骤 b 包括使至少一个完整的或未修饰的或修饰的 DNA 区与多核苷酸文库在允许所述 DNA 区与所述文库的一个或多个成员杂交的条件下接触，并且检

PARSE EX.1007.304

测所述 DNA 区与所述一个或多个成员的杂交。在一些实施方案中，所述文库组织在微阵列上。

[0036] 在一些实施方案中，步骤 b 包括扩增至少一个完整的或未修饰的或修饰的 DNA 区。

[0037] 在一些实施方案中，步骤 b 包括对至少一个完整的或未修饰的或修饰的 DNA 区进行核苷酸测序。

[0038] 本发明还提供了用于确定染色体上的基因座与 DNA 修饰剂的可接近性的试剂盒。在一些实施方案中，该试剂盒包含细胞膜透化剂或破坏剂；和限制酶、DNase 或其他 DNA 修饰剂。

[0039] 在一些实施方案中，所述细胞膜透化剂或破坏剂和所述限制酶或 DNase 处于同一容器中的同一缓冲液中。

[0040] 在一些实施方案中，所述细胞膜透化剂或破坏剂和所述限制酶或 DNase 处于单独的容器中。

[0041] 在一些实施方案中，该试剂盒还包含亚硫酸氢盐。

[0042] 在一些实施方案中，该试剂盒还包含用于扩增真核生物的基因组 DNA 的区的引物对。

[0043] 在一些实施方案中，该试剂盒还包含：

[0044] a. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分不包括所述 DNA 修饰剂的可能修饰位点；和 / 或

[0045] b. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分确实包括所述 DNA 修饰剂的至少一个可能修饰位点。

[0046] 在一些实施方案中，该试剂盒包含：

[0047] a. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分不能接近所述修饰剂；和 / 或

[0048] b. 引发对照 DNA 区的一部分扩增的引物，所述对照 DNA 区的一部分可接近所述修饰剂。

[0049] 通过阅读本文件的其余部分，本发明的其他实施方案将是明显的。

[0050] 定义

[0051] 本文使用的“透化 (Permeabilizing)”细胞膜是指降低细胞膜的完整性从而允许修饰剂进入细胞中。具有透化的细胞膜的细胞通常保留细胞膜使得细胞的结构仍然基本上是完整的。相反，本文使用的“破坏”细胞膜是指降低细胞膜的整合性使得细胞的结构不保持完整性。例如，使细胞膜与非离子型洗涤剂接触将去除和 / 或溶解细胞膜，从而允许修饰剂进入保留了至少一些染色体结构的基因组 DNA 中。

[0052] 本文使用的“DNA 修饰剂”是指以可检测到的方式改变 DNA 的分子。示例性的修饰包括 DNA 切口形成或 DNA 切割或将化学结构部分引入至 DNA 或从 DNA 去除化学结构部分。不导致 DNA 切割的 DNA 修饰剂包括，但不限于 DNA 甲基化酶。

[0053] 本文使用的“DNA 区”是指基因组 DNA 内的目的靶序列。所述 DNA 区可以具有任意的目的长度，所述长度可以被使用的 DNA 修饰剂所接近。在一些实施方案中，所述 DNA 区可以包括单个碱基对，但也可以是基因组 DNA 内的短序列节段（例如，2-100，2-500，50-500bp）或较大的节段（例如，100-10,000，100-1000，或 1000-5000bp）。DNA 区中 DNA 的量有时通

PARSE EX.1007.305

过要在 PCR 反应中扩增的序列的量来确定。例如，标准 PCR 反应通常可以扩增约 35-5000 个碱基对。

[0054] 不同“程度”的修饰是指样品间或一个或多个样品中两个或多个 DNA 区之间一个或多个 DNA 区的修饰的拷贝数不同（实际不同或相对不同）。例如，如果 100 拷贝的两个 DNA 区（为了方便起见标为“A 区”和“B 区”）各自存在于细胞中的染色体 DNA 中，则修饰至不同的程度的实例为如果 10 个拷贝的 A 区被修饰而 70 个拷贝的 B 区被修饰。

[0055] 术语“寡核苷酸”或“多核苷酸”或“核酸”可互换使用，是指单体的聚合物，其可以对应于核糖核酸 (RNA) 或脱氧核糖核酸 (DNA) 聚合物，或其类似物。这包括核苷酸诸如 RNA 和 DNA，以及其修饰形式，肽核酸 (PNAs)，锁核酸 (LNA™) 和类似物的聚合物。在某些应用中，核酸可以是包括多种单体类型的聚合物，例如，包括 RNA 和 DNA 亚单位两者。

[0056] 核酸典型是单链的或双链的并且通常包含磷酸二酯键，尽管在一些情况下，如本文所概述的，包括这样的核酸类似物，其可以具有备选的主链，包括，例如和不限于，磷酰胺 (Beaucage 等 (1993)Tetrahedron（四面体）49(10)：1925 和本文的参考文献；Letsinger(1970)J.Org.Chem.（有机化学杂志）35：3800；Sprinzl 等 (1977)Eur.J.Biochem.（欧洲生物化学杂志）81：579；Letsinger 等 (1986)Nucl.Acids Res.（核酸研究）14：3487；Sawai 等 (1984)Chem.Lett.（化学快报）805；Letsinger 等 (1988)J.Am.Chem.Soc.（美国化学会会志）110：4470；和 Pauwels 等 (1986)Chemica Scripta 26：1419)，硫代磷酸盐 (phosphorothioate)(Mag 等 (1991)Nucleic Acids Res.（核酸研究）19：1437 和美国专利申请号 5,644,048)，二硫代磷酸盐 (Briu 等 (1989)J.Am.Chem.Soc.（美国化学会会志）111：2321)，O-methylphophoroamidite 键 (Eckstein，Oligonucleotides and Analogues：A Practical Approach，（寡核苷酸和类似物：实践方法）Oxford University Press（牛津大学出版社）(1992))，和肽核酸主链和键 (Egholm(1992)J.Am.Chem.Soc.（美国化学会会志）114：1895；Meier 等 (1992)Chem.Int.Ed.Engl.（化学英文版）31：1008；Nielsen(1993)Nature（自然）365：566；和 Carlsson 等 (1996)Nature（自然）380：207)，这些参考文献的每一篇均通过引用结合。其他类似物核酸包括带正电荷主链的那些 (Denpcy 等 (1995)Proc.Natl.Acad.Sci.USA（美国国家科学院学报）92：6097)；带非离子主链的那些（美国专利号 5,386,023,5,637,684,5,602,240,5,216,141 和 4,469,863；Angew(1991)Chem.Intl.Ed.English（应用化学英文版）30：423；Letsinger 等 (1988)J.Am.Chem.Soc.（美国化学会会志）110：4470；Letsinger 等 (1994)Nucleoside & Nucleotide（核苷和核苷酸）13：1597；第 2 章和第 3 章，ASC Symposium Series 580(ASC 研讨会系列 580)，″ Carbohydrate Modifications in Antisense Research（反义研究中的碳水化合物修饰）″，Y.S.Sanghvi 和 P.Dan Cook 编辑；Mesmaeker 等 (1994)Bioorganic & Medicinal Chem.Lett.（生物有机化学和医药化学通讯）4：395；Jeffs 等 (1994)J.Biomolecular NMR（生物分子 NMR 杂志）34：17；Tetrahedron Lett.（四面体通讯）37：743(1996)) 和带非核糖主链的那些，包括美国专利号 5,235,033 和 5,034,506，以及第 6 章和第 7 章，ASC Symposium Series 580(ASC 研讨会系列 580)，Carbohydrate Modifications in Antisense Research（反义研究中的碳水化合物修饰），Y.S.Sanghvi 和 P.Dan Cook 编辑中所述的那些，这些参考文献的每一篇均通过引用结合。包含一种或多种碳环糖的核酸也包括在核酸的定义内 (Jenkins 等 (1995)Chem.Soc.Rev.

PARSE EX.1007.306

（化学会评论）169-176 页，该文献通过引用结合）。几种核酸类似物也描述在例如，Rawls，C & E News(C&E 新闻）6 月 2 日，1997，35 页中，该文献通过引用结合。核糖 - 磷酸盐主链的这些修饰可以被用来促进附加的结构部分诸如标记结构部分的添加，或用来改变这些分子在生理环境中的稳定性和半衰期。

[0057] 除了天然存在的杂环碱基（它们典型地存在于核酸中）（例如，腺嘌呤、鸟嘌呤、胸腺嘧啶、胞嘧啶和尿嘧啶）以外，核酸类似物也包括具有非天然存在的杂环或其他修饰的碱基的那些，它们中的许多记载在本文中，或另外提及。具体地，许多非天然存在的碱基还记载在，例如，Seela 等 (1991)Helv.Chim.Acta（瑞士化学学报）74 :1790，Grein 等 (1994) Bioorg.Med.Chem.Lett.（生物有机化学与医药化学通讯）4 :971-976，和 Seela 等 (1999) Helv.Chim.Acta（瑞士化学学报）82 :1640，它们的每一篇均通过引用结合。为了进一步说明，任选地包括在核苷酸中使用的充当解链温度 (Tm) 修饰剂的某些碱基。

[0058] 例如，这些中的一些包括 7- 去氮杂嘌呤 (deazapurines)（例如，7- 去氮杂鸟嘌呤，7- 去氮杂腺嘌呤等），吡唑并 [3,4-d] 嘧啶，丙炔基 -dN（例如，丙炔基 -dU，丙炔基 -dC 等），等等。参见，例如，美国专利号 5,990,303，题目为“7- 去氮杂 -2'- 脱氧鸟苷核苷酸的合成”，该专利于 1999 年 11 月 23 日授权给 Seela，并通过引用结合。其他代表性的杂环碱基包括，例如，次黄嘌呤，肌苷，黄嘌呤 ;2- 氨基嘌呤，2,6- 二氨基嘌呤，2- 氨基 -6- 氯嘌呤，次黄嘌呤，肌苷和黄嘌呤的 8- 氮杂衍生物 ;腺嘌呤，鸟嘌呤，2- 氨基嘌呤，2,6- 二氨基嘌呤，2- 氨基 -6- 氯嘌呤，次黄嘌呤，肌苷和黄嘌呤的 7- 去氮杂 -8- 氮杂衍生物 ;6- 氮杂胞嘧啶 ;5- 氟胞嘧啶 ;5- 氯胞嘧啶 ;5- 碘胞嘧啶 ;5- 溴胞嘧啶 ;5- 甲基胞嘧啶 ;5- 丙炔基胞嘧啶 ;5- 溴乙烯基尿嘧啶 ;5- 氟尿嘧啶 ;5- 氯尿嘧啶 ;5- 碘尿嘧啶 ;5- 溴尿嘧啶 ;5- 三氟甲基尿嘧啶 ;5- 甲氧基甲基尿嘧啶 ;5- 乙炔基尿嘧啶 ;5- 丙炔基尿嘧啶等。

[0059] 本文使用的 DNA 区与 DNA 修饰剂的“可接近性 (accessibility)”是指特定的 DNA 修饰剂接触和修饰细胞的染色体中的特定 DNA 区的能力。不希望限制本发明的范围，相信包含该 DNA 区的特定染色质结构将影响 DNA 修饰剂修饰该特定 DNA 区的能力。例如，该 DNA 区可以缠绕在组蛋白周围并且可以进一步具有其他的核小体结构，所述结构防止或减少 DNA 修饰剂接近目的 DNA 区。

[0060] “II-S 型限制酶”使用其本领域中的通常含义，指识别 DNA 中的特定识别序列，然后切割该识别序列外部的 DNA 分子的限制酶。示例性的 II-S 型限制酶包括，但不限于，MnII，FokI，和 AlwI。

[0061] 附图简要说明

[0062] 图 1 图示了染色质的示意性图示。真核生物 DNA 可以分成两种一般状态，常染色质和异染色质，在常染色质中 DNA 疏松地包装，是可接近的和有转录活性的，在异染色质中 DNA 是紧密包装的，不可接近的和转录沉默的。表观遗传学控制这两种状态之间的转变。本文所述的测定可以评估染色质结构，表观遗传学事件的功能结果。

[0063] 图 2 图示了测定的示意性图示。吸取培养基并且加入透化 / 消化缓冲液。核酸酶扩散至细胞中，进入细胞核中并消化可接近的染色质，但不消化不可接近的染色质（表示为朝向图底部的粗线）。

[0064] 图 3 图示了示例性工作流程的示意性图示。

[0065] 图 4 图示了一种评估当 Mnl 1 用作探针时生成的数据的分析方法。在每个测定中

PARSE EX.1007.307

分析两个 DNA 样品，一个 DNA 样品是对照，其分离自用不含核酸酶的缓冲液处理的细胞，另一个样品分离自用缓冲液和核酸酶处理的细胞。两个样品使用总引物组和完整引物组通过实时 PCR 分析。(A) 使用对照 DNA 样品数据通过从总 Ct 中减去完整的 Ct 计算 ΔCt（未切割的）。(B) 使用核酸酶处理的 DNA 样品数据通过从总 Ct 中减去完整的 Ct 计算 ΔCt（切割的）。可接近的基因在 MnlI 位点处被切割，其具有较少的完整 DNA 和较大的 ΔCt（切割的），该 ΔCt 与消化的程度成正比并反映较大的染色质可接近性。不可接近的或“锁定的(locked-down)”DNA 的量通过公式 2^{Δ}（ΔCt（未切割的）-ΔCt（切割的））来计算。

[0066] 图 5 是分析的靶基因的图。灰色区域代表转录的区域。指出了 Mnl 1 位点和 TATA 盒的位置。GAPDH 是一种管家基因，其在大多数细胞系中表达并且预期位于可接近的染色质结构中。血红蛋白 β 链 (HBB) 在大多数细胞中是沉默的并且预期位于不可接近的染色质结构中。谷胱甘肽 -s- 转移酶 pi(GSTP1) 在前列腺癌中表观遗传地失活，其在非癌性 RWPE-1 前列腺细胞中表达，但在癌性 LNCaP 前列腺细胞中是沉默的。

[0067] 图 6 图示了分析的靶基因和引物组的位置。使用总引物组和完整引物组分析所有基因，所述总引物组不跨越任何 MnlII 位点并且扩增所有 DNA 链，所述完整引物组跨越 2-4 个 MnlI 位点的完整引物组并仅扩增在这些位点处未被切割的 DNA 链。

[0068] 图 7 提供了对使用琼脂糖凝胶电泳分析 gDNA 完整性的概述。Hela 细胞用不含核酸酶的透化 / 消化缓冲液 (M，MnlI ；D，DNase I)，不含溶血卵磷脂的透化 / 消化缓冲液，或含有核酸酶和溶血卵磷脂两者的透化 / 消化缓冲液处理。通过琼脂糖凝胶电泳分离和分析 gDNA。结果表明核酸酶和溶血卵磷脂两者对于显著的 gDNA 消化均是必需的。

[0069] 图 8 图示了作为 MnlI 处理的 Hela 细胞中 HBB 的循环数的函数的荧光。Hela 细胞如指示的那样处理。用扩增 HBB 基因的引物分离和分析 gDNA。当使用 HBB 完整引物组时完全反应 (+，x) 的 Ct 值类似于对照反应。这提示在 Hela 细胞中 HBB 处于不可接近的染色质构型中。

[0070] 图 9 图示了作为 MnlI 处理的 Hela 细胞中 GAPDH 的循环数的函数的荧光。Hela 细胞如指示的那样处理。用扩增 GAPDH 基因的引物分离和分析 gDNA。当使用 GAPDH 完整引物组时完全反应 (+，x) 的 Ct 值高于对照反应（曲线右移）。这提示在 Hela 细胞中 GAPDH 处于可接近的染色质构型中。

[0071] 图 10 图示了作为 MnlI 处理的 Hela 细胞的循环数的函数的荧光。Hela 细胞如指示的那样处理。用扩增 GSTP1 基因的引物分离和分析 gDNA。当使用 GSTP1 完整引物组时完全反应 (+，x) 的 Ct 值高于对照反应（曲线右移）。这提示在 Hela 细胞中 GSTP1 处于可接近的染色质构型中。

[0072] 图 11 以图表的形式图示了三个 DNA 区 (HBB，GAPDH 和 GSTP1) 的数据，输出值计算为不可接近的总 DNA 的百分比。

[0073] 图 12 图示了作为 MnlI 处理的前列腺细胞中 HBB 的循环数的函数的荧光。在两种细胞系中，使用 HBB 完整引物组 (x) 的 + 酶反应的 Ct 值类似于使用 HBB 总引物组 (o) 的 + 酶反应。这提示在 RWPE-1 和 LNCaP 细胞中 HBB 处于不可接近的染色质构型中。

[0074] 图 13 图示了作为 MnlI 处理的前列腺细胞中 GAPDH 的循环数的函数的荧光。在两种细胞系中，使用 GAPDH 完整引物组 (x) 的 + 酶反应的 Ct 值高于（曲线右移）使用 GAPDH 总引物组 (o) 的 + 酶反应。这提示在 RWPE-1 和 LNCaP 细胞中 GAPDH 处于可接近的染色质

PARSE EX.1007.308

构型中。

[0075] 图 14 图示了作为 MnlI 处理的前列腺细胞中 GSTP1 的循环数的函数的荧光。在 RWPE-1 细胞中，使用 GSTP1 完整引物组 (x) 的 + 酶反应的 Ct 值高于（曲线右移）使用 GSTP1 总引物组 (o) 的 + 酶反应。这提示在 RWPE-1 细胞中 GSTP1 处于可接近的染色质构型中。相反，在 LNCaP 细胞中，使用 GSTP1 完整引物组 (x) 的 + 酶反应的 Ct 值类似于使用 GSTP1 总引物组 (o) 的 + 酶反应。这提示在 LNCaP 细胞中 GSTP1 处于不可接近的染色质构型中。

[0076] 图 15 图示了来自前列腺细胞的数据的概述。如材料和方法章节中所述计算不可接近的 DNA 的百分比。结果表明在两种细胞系中，HBB 处于不可接近的染色质结构中，GAPDH 处于可接近的染色质结构中。GSTP1 在 RWPE-1 细胞中处于可接近的染色质中，并在 LNCaP 细胞中处于不可接近的染色质结构中。

[0077] 图 16 图示了一种评估当 DNase I 用作染色质结构探针时生成的数据的分析方法。在每个测定中分析两个 DNA 样品，一个 DNA 样品是对照，其分离自用不含核酸酶的缓冲液处理的细胞，另一个样品分离自用缓冲液和核酸酶处理的细胞。两个样品使用扩增不可接近的参照基因 (HBB) 的引物组通过实时 PCR 分析。ΔCt（参照）代表不可接近的染色质的核酸酶消化程度，并且通过从与 + 核酸酶曲线相关的 Ct 中减去与无核酸酶曲线相关的 Ct 来计算。ΔCt（靶）代表靶染色质区的核酸酶消化程度，并且通过从与 + 核酸酶曲线相关的 Ct 减去与无核酸酶曲线相关的 Ct 来计算。不可接近的或“锁定的 (locked-down)”DNA 的量通过公式 2^{Δ}（ΔCt（参照）-ΔCt（靶））来计算。

[0078] 图 17 是分析的靶基因和引物组的位置的图。灰色区域代表转录的区域。带有箭头的黑色区域代表每个引物的位置和方向。

[0079] 图 18 图示了作为 DNase I 处理的细胞系中 HBB 的循环数的函数的荧光。在所有细胞系中，+ 核酸酶反应（灰线）的 Ct 值类似于无核酸酶反应（黑线）。这提示在所有细胞系中 HBB 处于不可接近的染色质构型中。

[0080] 图 19 图示了作为 DNase I 处理的细胞系中 GAPDH 的循环数的函数的荧光。在所有细胞系中，+ 核酸酶反应（灰线）的 Ct 值明显高于无核酸酶反应（黑线）的 Ct 值。这提示在所有细胞系中 GAPDH 处于可接近的染色质构型中。

[0081] 图 20 图示了作为 DNase I 处理的细胞系中 GSTP1 的循环数的函数的荧光。我们观察到 ΔCt(GSTP1) 值的细胞系依赖性差异。在 Hela 和 HCT15 细胞中，ΔCt(GSTP1) 很大，表明在这些细胞系中，GSTP1 处于可接近的构型中。在 LNCaP 细胞中，ΔCt(GSTP1) 很小，表明 GSTP1 处于不可接近的染色质中。在 PC3 细胞中，ΔCt(GSTP1) 具有中等值，表明 GSTP1 处于部分可接近的染色质结构中。这些结果与在不同细胞系中检测到的 GSTP1mRNA 表达的水平高度一致（参见图 21）。

[0082] 图 21 图示了来自使用 DNase I 作为染色质结构探针的实验的数据的概述。不可接近的 DNA 的百分比如图 16 中所述进行计算。GAPDH 和 GSTP1RNA 表达的相对水平通过 qRT-PCR 分析分离自不同细胞系的 mRNA 来计算。结果表明靶基因染色质结构与靶基因表达的水平的相关性良好。这提示所述的测定是评估对基因调控的表观遗传作用的有用工具。

[0083] 详述

[0084] I. 介绍

[0085] 本发明允许通过透化或破坏细胞膜和修饰细胞中的基因组 DNA，然后定量不同基

PARSE EX.1007.309

因座中的修饰程度来分析染色质结构。特定基因座处的修饰程度反映了染色体的该部分与修饰剂的可接近性，并且因此反映了染色质的状态。

[0086] 本发明的一个优点是发现了可以同时地透化和修饰细胞中的完整染色质，例如，通过使细胞与一种缓冲液接触，所述缓冲液包含透化剂和 DNA 修饰剂或切割剂。这允许非常快速地产生结果。此外，此方法消除了繁琐的和可能人工施加的步骤诸如细胞核的分离，等等，这些步骤在一些以前的染色质分析方法中是必需的。

[0087] II. 一般方法

[0088] 本发明的方法涉及同时透化细胞并使细胞与 DNA 修饰剂在一定的条件下接触，所述条件使得细胞中的基因组 DNA 与所述修饰剂具有不同的可接近性（由于染色质结构差异所导致），然后定量 DNA 区中修饰的量。DNA 的不同的可接近性可以反映基因组 DNA 的核小体结构。例如，在一些实施方案中，较易接近 DNA 修饰剂的 DNA 区有可能处于较“疏松”的染色质结构中。

[0089] 在本发明中可以使用多种真核细胞。在一些实施方案中，所述细胞是动物细胞，包括但不限于，人，或非人哺乳动物细胞。非人哺乳动物细胞包括但不限于，灵长类细胞，小鼠细胞，大鼠细胞，猪细胞和牛细胞。在一些实施方案中，所述细胞是植物细胞。细胞可以是，例如，培养的原代细胞，永生化培养细胞或可以来自活组织检查或组织样品，所述样品在测定前任选地被培养和刺激以发生分裂。培养细胞在透化和 / 或 DNA 修饰步骤之前和 / 或期间可以处于悬液中或是粘附的。细胞可以来自动物组织、活组织检查等。例如，所述细胞可以来自肿瘤活组织检查。

[0090] 本发明的方法可以包括将 DNA 区的可接近性与来自该相同区域的转录相关联。在一些实施方案中，进行实验以确定可接近性和基因表达之间的相关性，并且随后 DNA 修饰剂与特定 DNA 区的可接近性可以用来预测来自该 DNA 区的转录。在一些实施方案中，来自 DNA 区的转录和该区与 DNA 修饰剂的可接近性两者均被确定。已知有大量测量转录的方法，所述方法包括但不限于 northern 印迹和 RT-PCR 的应用。

[0091] 在一些实施方案中，一个区域的 DNA 甲基化状态可以与 DNA 区与 DNA 修饰剂的可接近性相关联。在一些实施方案中，进行实验以确定该区域的可接近性和 DNA 甲基化之间的相关性，随后 DNA 修饰剂与特定 DNA 区的可接近性可以用来预测来自该 DNA 区的 DNA 甲基化。在一些实施方案中，DNA 区的甲基化和该区与 DNA 修饰剂的可接近性两者均被确定。已知有大量测量 DNA 甲基化的方法，所述方法包括但不限于应用亚硫酸氢盐（例如，用于测序中和 / 或与甲基化敏感的限制酶结合使用（参见，例如，Eads 等，Nucleic Acids Research（核酸研究）28(8) :E32(2002)）和高分辨率熔解测定 (HRM)（参见。例如，Wodjacz 等，Nucleic Acids Research（核酸研究）35(6) :e41(2007)）。

[0092] 本发明在透化 /DNA 修饰后，提供了细胞基因组中第一 DNA 区与第二 DNA 区的量或其他物理特性的比较。备选地，或另外地，可以比较两种不同的细胞中第一 DNA 区的量或其他物理特性。例如，所述两种细胞可以代表患病的和健康的细胞或组织，不同的细胞类型，发育的不同阶段（包括但不限于干细胞或祖细胞）等。因此，通过使用本发明的方法，可以检测细胞之间染色质结构的差异和 / 或确定一个细胞内两个或多个 DNA 区（例如，基因）之间相对的染色质结构。另外，可以确定药物、化学制剂或环境刺激物对相同细胞或不同细胞中的特定区域的染色质结构的作用。

PARSE EX.1007.310

[0093] III. 透化和破坏细胞

[0094] 细胞膜可以以本领域中已知的任何方式被透化或破坏。如本文所解释的那样，本发明的方法包括在分离基因组 DNA 之前接触所述基因组 DNA，因此透化或破坏细胞膜的方法将不会破坏细胞的基因组 DNA 的结构从而破坏核小体或染色质结构。

[0095] 在一些实施方案中，细胞膜与透化或破坏细胞膜的试剂接触。溶血磷脂是示例类型的透化细胞膜的试剂。示例性的溶血磷脂包括，但不限于，溶血卵磷脂（在本领域中也称为溶血卵磷脂）或单棕榈酰磷脂酰胆碱。多种溶血磷脂也记载在，例如，WO/2003/052095 中。

[0096] 非离子型洗涤剂是示例类型的破坏细胞膜的试剂。示例的非离子型洗涤剂包括但不限于，NP40、Tween 20 和 Triton X-100。

[0097] 本发明的一个优势是同时递送透化剂和 DNA 切割剂或 DNA 修饰剂。因此，在一些实施方案中，包含两种试剂的缓冲液与细胞接触。缓冲液应当适于保持两种试剂的活性，同时维持细胞染色质的结构。

[0098] 备选地，可以使用电穿孔或基因枪法来透化细胞膜，使得 DNA 修饰剂引入至细胞中，并且因此接触基因组 DNA。大量的电穿孔法是公知的，并且可以适于递送如本文所述的 DNA 修饰剂。示例性的电穿孔法包括但不限于，WO/2000/062855 中所述的那些。基因枪法包括但不限于美国专利号 5,179,022 中所述的那些。

[0099] IV. DNA 修饰剂

[0100] 在透化后，或与透化同时（例如，在电穿孔期间或与透化剂温育期间），引入 DNA 修饰剂，使得该试剂接触基因组 DNA，从而在 DNA 中引入修饰作用。根据本发明可以使用大量的 DNA 修饰剂。

[0101] 在一些实施方案中，在去除透化剂后，DNA 修饰剂与被透化的细胞接触，任选地更换缓冲液。备选地，在一些优选的实施方案中，DNA 修饰剂与基因组 DNA 接触，而没有一个或多个中间步骤（例如，不更换缓冲液，不洗涤细胞等）。如上所述，此后一种方法可以便于减少必需的劳力和时间的量，并且还去除了测定中潜在的误差和污染的来源。

[0102] 使用的 DNA 修饰剂的量，以及与 DNA 修饰剂反应的时间将取决于使用的试剂。本领域技术人员将认识到如何根据使用的试剂来调整条件。通常，DNA 修饰步骤的条件被调节为使得不达到“完全的”消化。因此，例如，在一些实施方案中，设定修饰步骤的条件使得阳性对照（即，其中修饰是可接近的并发生修饰的对照）以高水平但小于 100％的水平，例如，80-95％，80-99％，85-95％，90-98％等之间的水平出现。

[0103] A. 限制酶

[0104] 在一些实施方案中，DNA 修饰剂是限制酶。因此，在这些实施方案中，引入至基因组 DNA 中的修饰是序列特异的单链（例如，切口）或双链切割剂。大量的限制酶是已知的并且可以用于本发明中。

[0105] 可以使用任何类型的限制酶。I 型酶远离其识别序列随机地切割 DNA。II 型酶在靠近其识别序列或其识别序列内的限定位置处切割 DNA。一些 II 型酶在其识别序列内切割 DNA。II-S 型酶切割其识别序列的外侧成一侧。第三种主要类型的 II 型酶，更确切地称为“IV 型”，切割其识别序列的外侧。例如，识别连续序列（例如，AcuI :CTGAAG）的那些仅在一侧上切割；识别不连续序列（例如，BcgI :CGANNNNNNTGC）的那些在两侧上切割，释放包含识

PARSE EX.1007.311

别序列的小片段。III 型切割其识别序列的外侧，并且需要同一 DNA 分子内处于相反方向上的两个这样的序列以完成切割。

[0106] 本发明的方法可以适于使用任何类型的限制酶或其他 DNA 切割酶。在一些实施方案中，该酶是相对接近地（例如，5、10 或 20 个碱基对内）切割识别序列的一种或多种酶。这样的酶可以特别地用于测定染色质结构，因为必须是可接近的从而实现切割的 DNA 的跨度比识别序列本身更大，因此可能涉及较宽跨度的 DNA，所述 DNA 不处于“紧密的”染色质结构中。序列特异的限制酶可以提供部分改善的定量结果，因为基于同一 DNA 区的对照可以如本文所述那样设计（例如，在实施例中）。因此，与例如用序列非特异的核酸内切酶（" DNase"）消化相比，可以更精确地确定总的和消化的拷贝数。不象 DNase I，II-S 型限制酶的切割可以对组蛋白与一个或多个目标 DNA 区的结合敏感。示例性的切割其识别序列外侧的酶包括，例如，II-S 型、III 型和 IV 型酶。II-S 型限制酶包括，但不限于，MnII、FokI 和 AlwI。

[0107] 在一些实施方案中，使用超过一种的（例如，两种、三种、四种等）限制酶。酶的组合可以包括所有均来自一种类型的酶的组合或可以是不同类型的混合物。

[0108] 随后可以单独地检测和定量完整的或切割的 DNA，并且可以如本文所述确定 DNA 区的完整和 / 或切割拷贝数。

[0109] 在一些实施方案中，透化剂或膜破坏剂在加入限制酶之前加入。在一些实施方案中，限制酶和透化剂或破坏剂同时加入（例如，在适当的缓冲液中或与适当的缓冲液一起）。即使两种试剂开始不在同一时刻与细胞接触，仍然可以获得同时透化和与 DNA 修饰剂的接触，因为透化可以是一个一直进行的过程。因此，例如，在添加透化剂稍后（在透化基本完成之前）添加 DNA 修饰剂可以被认为“同时”透化细胞和将细胞与 DNA 修饰剂接触。“同时”是指在添加透化剂和修饰剂之间没有中间操作发生（包括但不限于更换缓冲液，离心等）。

[0110] 在一些实施方案中，使用 0.5％溶血卵磷脂 (w/v)，50mM NaCl，10mM Tris-HCl pH7.4，10mM MgCl2，1mM DTT，100μg/ml BSA 和 0-500 单位 /ml MnlI（或其他限制酶）。在一些实施方案中，使用 0.25％溶血卵磷脂 (w/v)，50mM NaCl，10mM Tris-HCl pH7.4，10mM MgCl2，1mM DTT，100μg/ml BSA 和 0-500 单位 /ml MnlI（或其他限制酶）。在一些实施方案中，使用 0.75％溶血卵磷脂 (w/v)，50mM NaCl，10mM Tris-HCl pH7.4，10mM MgCl2，1mM DTT，100μg/ml BSA 和 0-500 单位 /ml MnlI（或其他限制酶）。在一些实施方案中，使用 1％溶血卵磷脂 (w/v)，50mM NaCl，10mM Tris-HCl pH7.4，10mM MgCl2，1mM DTT，100μg/ml BSA 和 0-800 单位 /ml MnlI（或其他限制酶）。

[0111] 在透化和消化后，任选地停止消化并且溶解细胞，所述溶解细胞任选地通过同时地添加溶解 / 终止缓冲液和 / 或增加温度。示例性的溶解 / 终止缓冲液可以包含足量的螯合剂和洗涤剂以终止反应和溶解细胞。例如，在一些实施方案中，溶解 / 终止缓冲液包含 100mM Tris-HCl pH8，100mM NaCl，100mM EDTA，5％ SDS(w/v) 和 3mg/ml 蛋白酶 K。在一些实施方案中，溶解 / 终止缓冲液包含 100mM Tris-HCl pH8，100mM NaCl，100mM EDTA，1％ SDS(w/v) 和 3mg/ml 蛋白酶 K。在一些实施方案中，溶解 / 终止缓冲液包含 200mM Tris-HCl pH8，100mM NaCl，500mM EDTA，5％ SDS(w/v) 和 5mg/ml 蛋白酶 K。

[0112] B. DNase

PARSE EX.1007.312

[0113] 在一些实施方案中，使用以序列非特异的方式切割或切口 DNA 的酶作为 DNA 修饰剂。因此，在一些实施方案中，DNA 修饰剂是序列非特异的核酸内切酶（本文中也称为“DNase”）。

[0114] 根据本发明可以使用任何序列非特异的核酸内切酶（例如，DNase I，II，III，IV，V，VI，VII 中的任一种）。例如，可以使用任何 DNase，包括但不限于 DNase I。使用的 DNase 可以包括天然存在的 DNase 以及修饰的 DNase。修饰的 DNase 的一个实例是 TURBO DNase(Ambion)，其包括允许“高活性 (hyperactivity)”和耐盐性的突变。示例性的 DNase 包括但不限于牛胰腺 DNase I（可获自，例如，New England Biolabs）。

[0115] 随后可以单独地检测和定量完整的 DNA，并且可以如本文所述确定 DNA 区的完整和 / 或切割拷贝数。

[0116] 在一些实施方案中，透化剂或膜破坏剂在加入 DNase 之前加入。在一些实施方案中，DNase 和透化剂或破坏剂同时加入（例如，与适当的缓冲液一起）。在一些实施方案中，透化 / 消化缓冲液包含 0.25％溶血卵磷脂 (w/v)，10mM Tris-HCl pH7.4，2.5mM $MgCl_2$，0.5mM CaCl2 和 0-200 单位 /ml DNase I。在一些实施方案中，透化 / 消化缓冲液包含 0.5％溶血卵磷脂 (w/v)，10mM Tris-HCl pH7.4，2.5mM $MgCl_2$，0.5mM CaCl2 和 0-200 单位 /ml DNase I。在一些实施方案中，透化 / 消化缓冲液包含 0.75％溶血卵磷脂 (w/v)，10mM Tris-HCl pH7.4，2.5mM $MgCl_2$，0.5mM CaCl2 和 0-500 单位 /ml DNase I。在一些实施方案中，透化 / 消化缓冲液包含 0.25％溶血卵磷脂 (w/v)，10mM Tris-HCl pH7.4，2.5mM $MgCl_2$，0.5mM CaCl2 和 0-500 单位 /ml DNase I。透化和溶解可以例如，如上面对限制酶所述的那样被终止。

[0117] 如其他地方所讨论的那样，通过监测至少两个不同的 DNA 区之间的切割程度来增强 DNase 或其他一般 DNA 切割剂的使用，所述一个 DNA 区是靶标，另一个是在任何测试条件下通常总是可接近的 DNA 区或通常总是不可接近的 DNA 区。这样的基因的实例在本文其他地方讨论，并且是已知的或可以被鉴定。例如，包括“管家”基因的 DNA 区通常总是可接近的。其余靶标与对照相比的相对量然后可以用来确定靶 DNA 区处的相对染色质结构。

[0118] C. 甲基转移酶

[0119] 在本发明的一些实施方案中，DNA 修饰剂对 DNA 产生共价修饰。例如，在一些实施方案中，本发明的 DNA 修饰剂是甲基转移酶。

[0120] 许多甲基转移酶在本领域中是已知的，并且可以在本发明中使用。在一些实施方案中，使用的甲基转移酶向 DNA 中的腺苷添加甲基结构部分。这样的甲基转移酶的实例包括，但不限于，DAM 甲基转移酶。因为在真核细胞中腺苷是被未甲基化，因此特定 DNA 区中甲基化腺苷的存在表明 DAM 甲基转移酶（或具有类似活性的其他甲基转移酶）能够接近该 DNA 区。腺苷甲基化可以被检测到，例如，使用其识别序列包含甲基化腺苷的限制酶。此类酶的实例包括，但不限于，DpnI。然后可以如本文所述定量由限制酶进行的切割（例如，其中完整的 DNA 被扩增，但切割的 DNA 不被扩增，或使用 LM-PCR 来扩增切割的 DNA 但不扩增完整的 DNA）。

[0121] 在一些实施方案中，甲基转移酶甲基化 GC 序列中的胞嘧啶。此类甲基转移酶的实例包括但不限于 MCviPI。参见，例如，Xu 等，Nuc.Acids Res.（核酸研究）26(17)：3961-3966(1998)。因为在真核细胞中 GC 序列是未甲基化的，因此特定 DNA 区中甲基化 GC

PARSE EX.1007.313

序列的存在表明 DNA 修饰剂（即，甲基化 GC 序列中的胞嘧啶的甲基转移酶）能够接近该 DNA 区。可以使用任何数量的技术来鉴定甲基化的 GC 序列。在一些实施方案中，检测甲基化 GC 序列的方法包括亚硫酸氢盐转变。亚硫酸氢盐转变包括使 DNA 与足量的亚硫酸氢盐接触从而将未甲基化的胞嘧啶转变为尿嘧啶。甲基化胞嘧啶未被转变。因此，包含 GC 序列的 DNA 区可以与甲基化 GC 序列中的胞嘧啶的甲基转移酶接触，分离，然后与亚硫酸氢盐接触。如果 GC 序列中的 C 未被甲基化，则 C 将被转变为 U（如果随后被扩增，则转变为 T），而甲基化的 C 将仍然为 C。任何数量的方法，包括但不限于，核苷酸测序和涉及引物延伸或基于引物的扩增和 / 或甲基化敏感的限制酶消化的方法可以用来检测亚硫酸氢盐转变的 C 的存在与否（例如，MSnuPE，MSP 或 Methyllight，高分辨率熔解分析；焦磷酸测序，等）。参见，例如，Fraga 等，Biotechniques（生物技术）33 ：632，634，和 636-649(2002) ；El-Maarri O Adv Exp Med Biol（实验医学与生物学进展）544 ：197-204(2003) ；Laird，Nat Rev Cancer（自然评论：癌症）3 ：253-266(2003) ；和 Callinan，Hum Mol Genet（人类分子遗传学）15 Spec No 1 ：R95-101(2006)。

[0122] 在一些实施方案中，甲基转移酶甲基化 CG（也称为“CpG”）序列中的胞嘧啶。此类甲基转移酶的实例包括但不限于 M. SssI。此类甲基转移酶的使用通常限于用于未典型地甲基化的那些 DNA 区。这是因为在真核细胞中 CG 序列被内源性地甲基化，因此除了在其中甲基化很少见的 DNA 区中以外，通常不可能假定 CG 序列被修饰剂甲基化而不是被内源性甲基转移酶甲基化。如 GC 序列一样，CG 序列的甲基化可以通过任何数量的方法来检测，包括涉及亚硫酸氢盐转变的方法。

[0123] D. 化学制剂

[0124] 在一些实施方案中，DNA 修饰剂包括修饰 DNA 的化学制剂。因为大多数修饰 DNA 的化学制剂与染色质相比相对较少，因此在不使用融合伴侣的条件下使用修饰 DNA 的化学制剂在一些情况下可能无效，因为如果不同 DNA 区的可接近性程度有任何差异也将是很小的。因此，在一些实施方案中，DNA 修饰剂包含与修饰 DNA 的化学制剂连接的具有位阻的分子。具有位阻的分子可以是根据染色质结构导致 DNA 修饰剂的可接近性产生差别的任何蛋白或其他分子。这可以例如，如本文所述通过将结果与使用 DNase 或限制酶的结果相比较来测试。

[0125] 在一些实施方案中，具有位阻的分子的大小为至少 5，7，10 或 15kD。本领域技术人员将有可能发现使用多肽作为具有位阻的分子是很适宜的。可以使用不显著地干扰 DNA 修饰剂修饰 DNA 的能力的任何多肽。在一些实施方案中，所述多肽是下面进一步详述的双链的序列非特异的核酸结合结构域。

[0126] 本发明的修饰 DNA 的化学制剂可以与具有位阻的分子直接连接或经由接头连接。多种同型和异型双功能接头是已知的，并且可以用于此目的。

[0127] 示例性的修饰 DNA 的化学制剂包括但不限于肼（及其衍生物，例如，如 Mathison 等，Toxicology and Applied Pharmacology（毒理学和实用药理学）127(1) ：91-98(1994) 中所述）和硫酸二甲酯。在一些实施方案中，肼将甲基引入至 DNA 中的鸟苷或否则损坏 DNA。在一些实施方案中，硫酸二甲酯甲基化鸟嘌呤或通过破坏鸟嘌呤中存在的咪唑环导致 DNA 中鸟嘌呤的碱基特异性切割。

[0128] 检测修饰 DNA 的化学制剂的修饰作用将取决于发生的 DNA 修饰的类型。在一些实

PARSE EX.1007.314

施方案中，为了检测硫酸二甲酯或肼修饰，在高温（90℃）下用哌啶处理 DNA。DNA 在 DNA 修饰的位点处断裂，并且可以以与本文所述的检测核酸酶切割的相同方式检测到该断裂。

[0129] E. 改善 DNA 修饰剂的 DNA 结合结构域

[0130] 在一些实施方案中，本发明的 DNA 修饰剂或切割剂与双链的序列非特异的核酸结合结构域（例如，DNA 结合结构域）融合或以其他的方式连接。在 DNA 修饰剂为多肽的情况下，双链的序列非特异的核酸结合结构域可以通过重组 DNA 技术合成为，例如，与 DNA 修饰剂的蛋白融合体。双链的序列非特异的核酸结合结构域是以序列非依赖性方式结合双链核酸的蛋白或蛋白的指定区域，即，对于特定序列结合不显示明显的偏好性。在一些实施方案中，双链的核酸结合蛋白显示对双链核酸的亲和力为对单链核酸的亲和力的 10 倍以上。在本发明的一些实施方案中双链的核酸结合蛋白是热稳定的。此类蛋白的实例包括，但不限于，古细菌的小碱性 DNA 结合蛋白 Sac7d 和 Sso7d（参见，例如，Choli 等，Biochimica et Biophysica Acta（生物化学与生物物理学报）950 :193-203，1988 ；Baumann 等，Structural Biol.（结构生物学）1 :808-819，1994 ；和 Gao 等，Nature Struc.Biol.（自然结构生物学）5 :782-786，1998），古细菌 HMf- 样蛋白（参见，例如，Starich 等，J.Molec.Biol.（分子生物学杂志）255 :187-203，1996 ；Sandman 等，Gene（基因）150 :207-208，1994），和 PCNA 类似物（参见，例如，Cann 等，J.Bacteriology（细菌学杂志）181 :6591-6599，1999 ；Shamoo 和 Steitz，Cell（细胞）:99，155-166，1999 ；De Felice 等，J.Molec.Biol.（分子生物学杂志）291，47-57，1999 ；和 Zhang 等，Biochemistry（生物化学）34 :10703-10712，1995）。对于关于 DNA 结合结构域的其他信息还参见欧洲专利 1283875B1。

[0131] Sso7d 和 Sac7d

[0132] Sso7d 和 Sac7d 是分别来自极端嗜热的古细菌（hyperthermophilic archaeabacteria）硫磺矿硫化叶菌（Sulfolobus solfataricus）和嗜硫菌（S.acidocaldarius）的小的（约 7,000kd MW），碱性染色体蛋白。这些蛋白富含赖氨酸并且具有高度的热、酸和化学制剂稳定性。它们以序列非依赖性方式结合 DNA，并且当结合时，在一些条件下将 DNA 的 T_M 增加至至多 40℃（McAfee 等，Biochemistry（生物化学）34 :10063-10077，1995）。这些蛋白和它们的同源物一般被认为在高温下参与稳定基因组 DNA。

[0133] HMF- 样蛋白

[0134] HMf- 样蛋白是古细菌的组蛋白，其在氨基酸序列和在结构方面均与真核生物 H4 组蛋白共享同源性，真核生物 H4 组蛋白被认为与 DNA 直接相互作用。HMf 蛋白家族在溶液中形成稳定的二聚体，并且已经从热稳定的物种（例如，炽热甲烷嗜热菌（Methanothermus fervidus）和焦球菌属（Pyrococcus）菌株 GB-3a）中鉴定了几种 HMf 同源物。HMf 蛋白家族一旦与 Taq DNA 聚合酶或任何具有低的固有持续合成能力（intrinsic processivity）的 DNA 修饰酶结合，就可以增强该酶沿 DNA 底物滑动的能力，并且因此增加其持续合成能力。例如，二聚体 HMf- 样蛋白可以与 Taq DNA 聚合酶的 N 末端共价连接，例如，通过化学修饰，并且因此改善了聚合酶的持续合成能力。

[0135] 本领域技术人员将认识到其他双链的序列非特异的核酸结合结构域在本领域中是已知的并且也可以如本文所述的那样使用。

[0136] F. 在 DNA 修饰步骤后分离 DNA

[0137] 在一些实施方案中，在 DNA 修饰 / 切割步骤后，根据可利用的任何方法将基因组

PARSE EX.1007.315

DNA 从细胞中分离。可以使用基本上任何 DNA 纯化方法只要其产生了可接受纯度的 DNA 用于随后的定量步骤。例如，可以使用标准的细胞溶解试剂来溶解细胞。任选地可以使用蛋白酶（包括但不限于蛋白酶 K）。如本领域中已知的那样，DNA 可以从混合物中分离。在一些实施方案中，使用苯酚 / 氯仿萃取，并且 DNA 可以随后被沉淀（例如，通过乙醇）和纯化。在一些实施方案中，如果需要，去除或降解 RNA（例如，使用 RNA 酶或使用 DNA 纯化柱）。

[0138] 任选地，扩增基因组 DNA 或否则从细胞溶胞产物中直接检测基因组 DNA，无需中间的纯化步骤。

[0139] 在一些实施方案中，将完整的、修饰的或未修饰的 DNA 分离并克隆至文库中。在一些情况下，分离和 / 或克隆一个或多个特异的完整的、修饰的或未修饰的序列。备选地，具有完整的、修饰的或未修饰的 DNA 区的样品用来制备为这些区富集的文库。在与 DNA 切割剂接触后完整的 DNA 代表了不太容易接近该试剂的 DNA。同样，在与 DNA 修饰剂接触后未修饰的 DNA 代表不太容易接近的 DNA。相反，修饰的 DNA 代表较容易接近修饰剂的 DNA。在上面的一些实施方案中，从切割的 DNA 中纯化（例如，分离）完整的 DNA 和 / 或在克隆前从未修饰的 DNA 中纯化修饰的 DNA，从而为一类 DNA 富集克隆池。对修饰的 / 未修饰的 DNA 的富集将根据修饰作用的性质而变化。在一些实施方案中，与修饰的（或未修饰的 DNA）特异性结合的亲合试剂用来从未修饰的 DNA 中分离修饰的 DNA。

[0140] 在一些实施方案中，产生差减文库 (subtractive libraries)。例如，可以产生对患病细胞 DNA 区富集的文库，在本发明的方法中所述 DNA 区是完整的、修饰的或未修饰的，随后该文库用来自健康细胞的相应文库差减，从而产生差异 DNA 序列的文库，所述 DNA 序列均是完整的、修饰的或未修饰的，并且对于特定疾病是特异的。可以使用任何患病细胞，包括但不限于，癌细胞。也可以采用可选的差减策略，例如，在不同的细胞类型之间，细胞期，治疗期间，等等。

[0141] G. 检测 DNA 的物理特性

[0142] 在使细胞与 DNA 修饰剂或 DNA 切割剂接触后可以检测 DNA 的任何数量的物理特性。物理特性包括，但不限于，DNA 甲基化，解链温度，GC 含量，核苷酸序列和与多核苷酸杂交的能力。多种方法已知用于检测此类特性，并且可以采用。在一些实施方案中，在 DNA 修饰 / 切割步骤后，确定的物理特性不包括确定 DNA 足迹（例如，一种或多种特定蛋白对 DNA 的特定区域的能力）。例如，在非限制性的实施中，完整 DNA 的定量，例如，使用 qPCR，不包括 DNA 足迹。

[0143] 在一些实施方案中，所述物理特性是 DNA 甲基化。例如，一旦相对可接近的 DNA 被 DNA 切割剂切割，则可以分离其余的完整 DNA（代表不太容易接近的 DNA），然后可以分析甲基化状态。大量 DNA 甲基化检测方法是已知的。在一些实施方案中，在与 DNA 修饰剂或切割剂接触后，DNA 与亚硫酸氢盐接触，从而将 DNA 中未甲基化的胞嘧啶转变为尿嘧啶。然后可以通过任何数量的甲基化检测方法，包括本文中讨论的那些确定特定 DNA 区的甲基化。在一些实施方案中，采用高分辨率熔解测定 (HR) 来检测亚硫酸氢盐转变后的甲基化状态。在此方法中，在亚硫酸氢盐转变后扩增 DNA 区，并测定得到的扩增子的解链温度。因为解链温度根据胞嘧啶是否被亚硫酸氢盐转变（并且随后在扩增反应中被拷贝为“T”）而不同，因此扩增子的解链温度可与甲基化含量相关。

[0144] V. 靶 DNA 区

PARSE EX.1007.316

[0145] DNA 区是基因组 DNA 内的目的靶序列。可以如本文所述评价细胞的基因组 DNA 中的任何 DNA 序列的 DNA 修饰剂可接近性。可以筛选 DNA 区以鉴定目的 DNA 区，所述目的 DNA 区例如在不同的细胞类型中，在未处理的细胞和暴露于药物、化学制剂或环境刺激物的细胞之间，或正常和患病组织之间展示不同的可接近性。因此，在一些实施方案中，本发明的方法用来鉴定 DNA 区，所述 DNA 区的可接近性变化用作疾病的标志物（或缺少该变化）。示例性的疾病包括但不限于癌症。已经描述了大量基因，与非癌细胞相比，它们在癌细胞中改变 DNA 甲基化和 / 或染色质结构。

[0146] 在一些实施方案中，所述 DNA 区已知根据特定细胞的疾病或发育状态而具有不同的可接近性。在这些实施方案中，本发明的方法可以用作诊断或预后工具。一旦使用本发明的方法确定了诊断或预后，则可以确定治疗方案或考虑该诊断或预后而改变现有的治疗方案。例如，根据本发明的方法检测癌细胞可以导致向检测到癌细胞的个体施用化疗剂和 / 或放射线。

[0147] 为了研究目的和 / 或作为对照 DNA 区可以检测许多 DNA 区从而确认试剂正如期望的那样发生作用。例如，在一些实施方案中，测定 DNA 区，该 DNA 区在动物的基本上所有细胞中均是可接近的。这样的 DNA 区，例如，可用作可接近性的阳性对照。这样的 DNA 区可以见于，例如，为组成性的或几乎组成性的基因内或邻近该基因。这样的基因包括通常被称作“管家”基因的那些，即，其表达对于维持基本细胞功能是必需的基因。这样的基因的实例包括，但不限于甘油醛 -3- 磷酸脱氢酶 (GAPDH) 和 β 肌动蛋白 (ACTB)。DNA 区可以包括此类基因的全部或一部分，任选地包括启动子的至少一部分。

[0148] 在一些实施方案中，DNA 区包含在动物的大多数细胞中是不可接近的 DNA 的至少一部分。这样的 DNA 区例如，可用作可接近性的阴性对照。在此背景中“不可接近的”是指这样的 DNA 区，该 DNA 区的拷贝中的不超过约 20％的拷贝被修饰。此类基因序列的实例包括通常被认为是“异染色质”的那些，并且包括仅在非常特殊的细胞类型中表达的基因（例如，以组织或器官特异性方式表达）。通常不可接近的示例性基因（除了特殊细胞类型外）包括，但不限于，血红蛋白 -β 链 (HBB) 和免疫球蛋白轻链 κ (IGK)。

[0149] 在一些实施方案中，DNA 区这样的基因序列，其根据细胞的疾病状态具有不同的可接近性，或否则根据细胞的类型或生长环境具有可变的可接近性。例如，一些基因通常在非癌细胞中是不可接近的，但在癌细胞中是可接近的。具有可变的可接近性的基因的实例包括，例如，谷胱甘肽 -s- 转移酶 pi(GSTP1)。

[0150] 在一些实施方案中，本发明的 DNA 区选自来自下列基因中的一种或多种的基因序列（例如，启动子序列）：1 型钙粘蛋白 (cadherin)1(E- 钙粘蛋白)，细胞色素 P450-1A1(CYP1A1)，Ras 相关结构域家族 1A(RASSF1A)，p15，p16，死亡相关蛋白激酶 1(DAPK)，结肠的腺瘤性息肉 (APC)，甲基鸟嘌呤 -DNA 甲基转移酶 (MGMT)，乳腺癌 1 基因 (BRCA1) 和 hMLH。

[0151] 在一些实施方案中，随机地选择 DNA 区，例如，以鉴定在不同的细胞类型、不同的条件、正常细胞相对于患病细胞，等之间具有的差别的可接近性的区域。

[0152] VI. 定量靶基因座的拷贝

[0153] 定量 DNA 修饰的方法将取决于引入至基因组 DNA 中的 DNA 修饰的类型。例如，使用设计用来产生包含目的 DNA 区的扩增子的扩增反应可以方便地检测到双链 DNA 切割事件

PARSE EX.1007.317

(例如,通过限制酶或 DNase 引入或在修饰后引入,例如,在甲基转移酶处理后,或在如本文所述被修饰 DNA 的化学制剂修饰后通过甲基化敏感的或甲基化依赖的限制酶引入)。在确定的位点处发生切割事件的情况下,诸如当使用序列特异的限制酶时,设计产生跨越潜在切割位点的扩增子的引物。仅完整的 DNA 会被扩增。如果还知道总 DNA 的量,则可以将总 DNA 和完整 DNA 之间的差值计算为切割的 DNA 的量。可以根据本领域中已知的任何 DNA 定量方法来确定 DNA 的总量。在一些实施方案中,可以通过设计一组引物来方便地确定总 DNA 的量,所述引物扩增 DNA 无论其是否存在修饰。这可以例如,通过设计同一基因区内或另一 DNA 区内的不跨越潜在切割位点的引物来实现。在不确定的位点处发生切割事件的情况下,诸如当使用非序列特异性核酸酶,诸如使用 DNase I 时,使用不可接近的参照基因应当加入作为内部对照。

[0154] 如下面所更详细地讨论的那样,定量扩增(包括,例如,实时 PCR)法允许确定 DNA 区的完整拷贝量,并且当用多种对照使用时,可以用来确定与细胞中的总拷贝数相比完整 DNA 的相对量。因此可以计算出该 DNA 区的修饰的或未修饰的实际或相对拷贝数(例如,相对于总拷贝数或相对于第二 DNA 区的修饰的或未修饰的拷贝数)。

[0155] 在本发明的一些实施方案中,直接测定 DNA 区的修饰的拷贝数。例如,也可以例如,通过检测特定的连接事件,例如仅在存在特定粘端或钝端时发生的事件来检测和定量限制酶切割。例如,包含粘端的核酸适体可以与切割的基因组 DNA 连接,所述粘端与由限制酶产生的粘端是互补的。然后可以检测和定量连接事件的次数(例如,通过定量扩增方法)。

[0156] 在一些实施方案中,采用连接介导的 PCR(LM-PCR)来定量 DNA 区的切割拷贝数。LM-PCR 的方法在本领域中是已知的,并且最初记载在 Pfeifer 等,Science(科学)246:810-813(1989)中。如果需要,LM-PCR 可以实时进行以获得定量结果。

[0157] 定量扩增方法(例如,定量 PCR 或定量线性扩增)涉及扩增核酸模板,直接或间接地(例如,确定 Ct 值)确定扩增的 DNA 的量,然后基于扩增的循环次数计算初始模板的量。使用反应扩增 DNA 基因座是公知的(参见美国专利号 4,683,195 和 4,683,202;PCR PROTOCOLS:A GUIDE TO METHODS AND APPLICATIONS(PCR 实验技术:方法和应用指南)(Innis 等,编辑,1990))。典型地,PCR 用来扩增 DNA 模板。然而,已经描述了可选的扩增方法并且也可以采用,只要该可选方法将完整 DNA 扩增至大于扩增切割的 DNA 的方法的程度。定量扩增的方法公开在,例如,美国专利号 6,180,349;6,033,854;和 5,972,602, 以及在,例如,Gibson 等,Genome Research(基因组研究)6:995-1001(1996);DeGraves 等,Biotechniques(生物技术)34(1):106-10,112-5(2003);Deiman B 等,Mol Biotechnol.(分子生物技术)20(2):163-79(2002)。扩增可以“实时地”监测。

[0158] 在一些实施方案中,定量扩增基于对信号(例如,探针的荧光)的监测,所述信号代表在扩增(例如,PCR)反应的循环中模板的拷贝。在 PCR 的初始循环中,观察到非常低的信号,因为形成的扩增子的量不支持从测定中输出可测量到的信号。在初始循环后,当形成的扩增子的量增加时,信号强度增加至可测量到的水平,并且当 PCR 进入非对数期时在后面的循环中达到平台期。通过信号强度相对于循环次数的作图,可以推算出从 PCR 反应获得可测量到的信号时的特定循环,并且用来逆算出 PCR 开始前靶标的量。通过此方法确定的特定循环的次数典型地称为循环阈值(Ct)。示例性方法记载在,例如,关于水解探针的

PARSE EX.1007.318

Heid 等 Genome Methods(基因组方法)6:986-94(1996) 中。

[0159] 一种检测扩增产物的方法是 5′ -3′ 核酸外切酶″水解″ PCR 测定(也称为 TaqMan™ 测定)(美国专利号 5,210,015 和 5,487,972;Holland 等,PNAS USA88:7276-7280(1991);Lee 等,Nucleic AcidsRes.(核酸研究)21:3761-3766(1993))。此测定通过扩增反应期间双标记的荧光探针(″ TaqMan™ 探针)的杂交和断裂来检测特定 PCR 产物的积累。该荧光探针由用荧光报告染料和淬灭剂染料两者标记的寡核苷酸组成。在 PCR 期间,如果,并且只有如果此探针与被扩增的区段杂交,则该探针被 DNA 聚合酶的 5′ -核酸外切酶活性切割。探针的断裂产生报告染料的荧光强度增加。

[0160] 另一种检测扩增产物的方法依赖于能量转移的使用,其是由 Tyagi 和 Kramer,Nature Biotech.(自然生物技术)14:303-309(1996) 描述的″灯塔探针(beacon probe)″法,该方法也是美国专利号 5,119,801 和 5,312,728 的主题。此方法采用了可以形成发夹结构的寡核苷酸杂交探针。在该杂交探针的一端上(5′ 或 3′ 端),存在给体荧光团,在另一端上为受体结构部分。在 Tyagi 和 Kramer 方法的情况下,此受体结构部分是淬灭剂,即,该受体吸收给体释放的能量,但然后其本身不发荧光。因此,当灯塔处于开放构型时,给体荧光团的荧光是可检测到的,而当灯塔处于发夹(闭合)构型时,给体荧光团的荧光被淬灭。当在 PCR 中使用时,与 PCR 产物的一条链杂交的分子灯塔探针处于开放构型,并且检测到荧光,同时未杂交的那些不会发出荧光(Tyagi 和 Kramer,Nature Biotechnol.(自然生物技术)14:303-306(1996))。结果,当 PCR 产物的量增加时荧光的量将增加,并且因此荧光量可以用作 PCR 继续进行的量度。本领域技术人员将认识到也可利用其他定量扩增方法。

[0161] 用于进行核酸的定量扩增的多种其他技术也是已知的。例如,一些方法学采用一种或多种探针寡核苷酸,所述探针寡核苷酸被构造为当一种或多种寡核苷酸与靶核酸杂交时产生荧光变化。例如,一种此类方法包括一种双荧光团法,该方法利用荧光共振能量转移(FRET),例如,LightCycler™ 杂交探针,其中两个寡核苷酸探针复性成扩增子。寡核苷酸被设计为以头对尾方向与荧光团杂交,所述荧光团以与有效能量转移相容的距离相隔。被构造成当与核酸结合或掺入至延伸产物中时发出信号的标记寡核苷酸的其他实例包括:Scorpions™ 探针(例如,Whitcombe 等,Nature Biotechnology(自然生物技术)17:804-807,1999,和美国专利号 6,326,145),Sunrise™(或 Amplifluor™) 探针(例如,Nazarenko 等,Nuc.Acids Res.(核酸研究)25:2516-2521,1997,和美国专利号 6,117,635),和形成二级结构的探针,所述二级结构在无淬灭剂的条件下导致信号减小并且当与靶标杂交时发出增加的信号(例如,Lux 探针 ™)。

[0162] 在其他实施方案中,可以使用嵌入剂,所述嵌入剂当嵌入至双链 DNA 中时产生信号。示例性的试剂包括 SYBR GREEN™,SYBR GOLD™ 和 EVAGREEN™。由于这些试剂不是模板特异的,因此假定基于模板特异的扩增产生信号。这可以通过监测作为温度的函数的信号来证实,因为模板序列的熔点通常比例如,引物 - 二聚体等高得多。

[0163] 在一些实施方案中,DNA 区的量通过对样品中的拷贝进行核苷酸测序,然后确定样品中具有相同序列的拷贝的相对数或绝对数来确定。

[0164] 在一些实施方案中,根据本发明的方法定量修饰的(或未修饰的)DNA 区可以通过确定该 DNA 区的修饰的或未修饰的拷贝与该同一区的总拷贝数相比的相对量(例如,归一

PARSE EX.1007.319

化值诸如比率或百分比）来进一步改进。在一些实施方案中，一个 DNA 区的修饰的或未修饰的拷贝的相对量与第二个（或更多个）DNA 区的修饰的或未修饰的拷贝数相比较。在一些实施方案中，当在两个或多个 DNA 区之间比较时，每个 DNA 区的修饰的或未修饰的拷贝的相对量可以首先针对该 DNA 区的总拷贝数归一化。备选地，在一些实施方案中，当获自同一样品时，可以假定每个 DNA 区的总拷贝数大致相同，并且因此，当在两个或多个 DNA 区之间比较时，确定每个 DNA 区之间修饰的或未修饰的拷贝的相对量（例如，比率或百分比）而不用首先将每个值针对总拷贝数进行归一化。

[0165] 在一些实施方案中，修饰的或未修饰的拷贝的实际或相对（例如，相对于总 DNA）量与对照值进行比较。对照值可以适宜地使用，例如，在想知道特定 DNA 区的可接近性是否超过或低于特定值的情况下。例如，在特定 DNA 区在正常细胞中典型地是可接近的，但在患病细胞中是不可接近的（或反之亦然）的情况下，可以简单地将修饰的或未修饰的拷贝的实际数或相对数与对照值进行比较（例如，大于或小于 20％修饰的或未修饰的，大于或小于 80％修饰的或未修饰的，等）。备选地，对照值可以代表关于对照 DNA 区的历史数据或预期数据。在这些情况下，确定对照 DNA 区的实际量或相对量（任选地确定多次），得到的数据用来产生对照值，该对照值可以与对目的 DNA 区确定的修饰的或未修饰的拷贝的实际数或相对数进行比较。

[0166] 对本文所述的方法的计算可以涉及基于计算机的计算和工具。所述工具有利地以计算机程序的形式提供，所述计算机程序可由常规设计的通用计算机系统（本文中称为“主机 (host computer)”）执行。该主机可以配置以许多不同的硬件元件并且可以以许多尺寸和样式制造（例如，台式 PC，膝上型计算机，平板 PC，掌上电脑，服务器，工作站，大型电脑）。可以包括标准元件，诸如监视器、键盘、磁盘驱动器、CD 和 / 或 DVD 驱动器等。在主机与网络连接时，所述连接可以通过任何合适的传输介质（例如，有线的、光纤的、和 / 或无线的介质）和任何合适的通信协议（例如，TCP/IP）来提供；主机可以包括合适的网络硬件（例如，调制解调器，以太网卡，WiFi 卡）。主机可以执行多种操作系统中的任一种，包括 UNIX，Linux，Microsoft Windows，MacOS，或任何其他操作系统。

[0167] 用于实现本发明的各个方面的计算机代码可以以多种语言编写，包括 PERL，C，C++，Java，JavaScript，VBScript，AWK 或任何其他脚本或编程语言，它们可以在主机上执行或可以被编译从而在主机上执行。代码还可以以低级语言编写或分配，所述低级语言诸如汇编语言或机器语言。

[0168] 主机系统有利地提供界面，通过所述界面用户控制所述工具的操作。在本文所述的实施例中，软件工具作为脚本来实现（例如，使用 PERL），其执行可以由用户从操作系统诸如 Linux 或 UNIX 的标准命令行界面来启动。本领域技术人员将理解命令可以根据需要被改编以适应操作系统。在其他实施方案中，可以提供图形用户界，该界面允许用户使用点击设备控制操作。因此，本发明不限于任何特殊的用户界面。

[0169] 加入了本发明的各种特征的脚本或程序可以编码在多种计算机可读介质上用于保存和 / 或传输。合适的介质的实例包括磁盘或磁带，光存储介质诸如光盘 (CD) 或 DVD（多功能数码光盘），快闪存储器，和适于通过符合多种协议的有线、光纤和 / 或无线网络包括互联网 (Internet) 传输的载波信号。

[0170] VII. 反应混合物

PARSE EX.1007.320

[0171] 本发明还提供了包含本文所述的一种或多种试剂的反应混合物，任选地含有真核细胞（其染色质状态待确定）。在一些实施方案中，该反应混合物包含，例如，DNA 修饰剂（例如，限制酶，DNase，甲基转移酶或修饰 DNA 的化学制剂）和细胞透化剂和 / 或细胞破坏剂和真核细胞。本文所述的其他试剂（包括但不限于亚硫酸氢盐）也可以包含在本发明的反应混合物中。

[0172] VIII. 试剂盒

[0173] 本发明还提供了用于进行本发明的可接近性测定的试剂盒。试剂盒可以任选地包括书面说明书或电子说明书（例如，在 CD-ROM 或 DVD 上）。本发明的试剂盒可以包括，例如，DNA 修饰剂和细胞透化剂和 / 或细胞破坏剂。DNA 修饰剂可以包括本文详述的那些，包括，例如，限制酶，DNase，甲基转移酶或修饰 DNA 的化学制剂。在一些实施方案中，DNA 修饰剂是 II-S 型限制酶，包括但不限于，MnlI，FokI 和 AlwI。本发明的试剂盒可以在相同的瓶 / 容器中包含透化剂和 DNA 修饰剂（并且因此在相同的缓冲液中）。备选地，透化剂和 DNA 修饰剂可以处于单独的瓶 / 容器中。

[0174] 本发明的试剂盒还可以包含一种或多种对照细胞和 / 或核酸。示例性的对照核酸包括，例如，包含在动物的基本上所有细胞中均是可接近的（例如，管家基因序列或其启动子）或在动物的大多数细胞中均是不可接近的基因序列的那些。在一些实施方案中，该试剂盒包括一个或多个用于扩增所述基因序列的引物组（无论该基因序列或细胞是否实际包括在试剂盒中）。例如，在一些实施方案中，该试剂盒包括 DNA 修饰剂，和细胞透化剂和 / 或细胞破坏剂，和一个或多个用于扩增对照 DNA 区的引物组，和任选地一个或多个用于扩增第二个 DNA 区（例如，靶 DNA 区）的引物组。在其中 DNA 修饰剂是限制酶的实施方案中，每个 DNA 区可以包括至少两个引物组：一个引物组用于扩增 DNA 区的一部分，所述部分包括限制酶的至少一个（例如，1 个，2 个，3 个，4 个等）潜在切割位点（例如，可用于计算未修饰的拷贝数），和至少第二引物组用于扩增 DNA 区的一部分，所述部分不包含任何潜在切割位点（例如，可用于计算总拷贝数）。

[0175] 在一些实施方案中，本发明的试剂盒包括下面各项中的一个或多个：

[0176] (i) 细胞膜透化剂或破坏剂；

[0177] (ii) 限制酶，DNase，或其他 DNA 修饰剂；

[0178] (iii) 能够防止修饰剂进一步修饰的“终止”溶液；

[0179] (iv) 用于提取和 / 或纯化核酸的材料（例如，用于纯化基因组 DNA 的离心柱 (spin column) 和 / 或去除非 DNA 成分诸如“终止”溶液的成分）；

[0180] (vi) 用于 PCR/qPCR 扩增 DNA 的试剂，任选地包含模板和 / 或聚合酶以外的 PCR 或 qPCR 所必需的所有成分的一种混合物；

[0181] (vii) 用于 PCR/qPCR 扩增特定靶基因的引物组。

实施例

[0182] 提供下面的实施例以说明，但不限于要求保护的本发明。

[0183] 实施例 1：一般方法

[0184] 本测定利用了 DNA 修饰剂对细胞中染色质的不同部分的可接近性差异。如图 2 中所示，DNA 修饰剂可以比其他部分更容易地接近染色质的某些部分。图 3 图示了该测定的

PARSE EX.1007.321

示例性工作流程。粘附细胞生长在 24 孔板中作为起始材料。在此实施方案中，对于每个实验使用两个孔：一个孔用不含核酸酶的透化缓冲液处理；另一孔用含有核酸酶的透化缓冲液处理。整个过程，从细胞至实时 PCR，需要约 3 个小时，并且在收获细胞的当天就可得到结果。

[0185] 图 4 图示了当 MnlI 用作染色质结构探针时分析分离自细胞的 DNA 的示意图。在此情况下该测定使用“无核酸酶”对照平行地进行。两个引物组用于每个基因。一个引物组不跨越任何 MnlI 消化位点，而另一个引物组跨越至少一个 MnlI 位点。用透化试剂处理但不用核酸酶处理的细胞导致从两个引物组完全扩增（即，所有拷贝被扩增）并且因此应该具有类似的 Ct 值，如图 4 左图中所图示。ΔCt 值（总 Ct 减去完整的 Ct）反映了基于引物特性，未切割的 DNA 中这些引物组之间 Ct 值的差异。

[0186] 相反，在已经用 MnlI 处理的细胞中，可接近的基因应该在 MnlI 位点处被切断。不跨越 MnlI 位点的引物组仍会扩增所有 DNA 分子（用作总拷贝数的量度），因此 Ct 将反映 DNA 的总拷贝的量。另一引物组（跨越至少一个 MnlI 位点）将仅能够扩增未切割的或完整的序列。取决于消化的程度，Ct 值将改变随着消化的量而成比例地移动和增加。然后 ΔCt 可以通过从总 Ct 中减去完整的 Ct 来计算，如果 DNA 是可接近的，则这将产生负数。为不可接近的 MnlI 位点的百分比通过图 4 中显示的方程来计算。

[0187] 将此技术应用于不同类型的 DNA 区图示在图 5-6 中。图 5 显示 GAPDH，HBB，和 GSTP1 基因的至少一部分的图示。显示了 Mnl I、TATA 盒、启动子区和外显子。GAPDH 用作阳性对照。GAPDH 是管家基因，其在大多数细胞系中表达并且应当是可接近的。血红蛋白 β 链基因被选择为阴性对照。血红蛋白在大多数细胞中不表达，是表观遗传沉默的，并且预期处于不可接近的染色质结构中。最后分析谷胱甘肽 -s- 转移酶 pi，或 GSTP1，其在前列腺癌中是表观遗传失活的。图 6 显示可以被靶向用于扩增的区。对于每种基因，使用不跨越任何 MnlI 位点并且将扩增所有 DNA 链的总引物组。还使用跨越至少一个（例如，2-4 个）MnlI 位点并且将仅扩增在这些位点处未被切割的 DNA 链的“完整”引物组。

[0188] 实施例 2：在 Hela 细胞中使用 MnlI 的数据

[0189] 此实施例显示了使用如上所概述的实验方法产生的数据的结果。

[0190] 材料和方法

[0191] 化学制剂 .L-α- 溶血磷脂酰胆碱（溶血卵磷脂）购自 Sigma-Aldrich。MnlI，DNase I，BSA 和蛋白酶 K 购自 New England Biolabs。RNA 酶 A 购自 Qiagen。组织培养板购自 VWR。iQ SYBR 购自 Bio-Rad。

[0192] 细胞的处理 . 生长在 24 孔板中的细胞当达到 90％汇合时被处理。吸取培养基并且将 100μl 透化 / 消化缓冲液轻轻地铺在细胞上。对于用 MnlI 处理的细胞，透化 / 消化缓冲液由溶血卵磷脂，NaCl，Tris-HCl，$MgCl_2$，DTT，BSA 和 MnlI 组成。对于用 DNase I 处理的细胞，透化 / 消化缓冲液由溶血卵磷脂，Tris-HCl，$MgCl_2$，CaCl2 和 DNase I 组成。然后透化的细胞在 37℃下孵育 1 小时。孵育后，25μl 溶解 / 终止溶液（100mM Tris-HCl pH7.4，100mM NaCl，100mM EDTA，5％ N- 月桂酰肌苷酸 (w/v)，80μg/ml RNA 酶 A 和 3mg/ml 蛋白酶 K）加入至透化 / 消化缓冲液中，并且细胞溶胞产物在 37℃下孵育 10 分钟。

[0193] DNA 的分离 . 收集细胞溶胞产物，并且使用市售的核酸纯化试剂盒 (Aurum，Bio-Rad) 遵照标准方法分离 DNA。DNA 以 100μl 洗脱，并且使用 NanoDrop 1000 分光光度

PARSE EX.1007.322

计 (Thermo Scientific) 测定 DNA 的浓度。然后将样品稀释至 1-10ng DNA/μl 的浓度。

[0194] 分析的基因 . 我们分析了三个人类基因的启动子区 :甘油醛 -3- 磷酸脱氢酶 (GAPDH),血红蛋白, β (HBB) 和谷胱甘肽 -s- 转移酶 π (GSTP1)。GAPDH 是一种管家基因 ; 其在大多数细胞中表达并且预期具有可接近的染色质结构。HBB 在大多数细胞中不表达,并且预期具有不可接近的染色质结构。GSTP1 在前列腺癌中是表观遗传失活的。在非癌性前列腺细胞系 RWPE-1 中, GSTP1 处于可接近的染色质结构中并且被表达,在癌性前列腺细胞系 LNCaP 中, GSTP1 处于不可接近的染色质结构中并且不表达 (Okino, S.T. 等, Chromatin changes on the GSTPI promoter associated with its mactivation in prostate cancer(GSTP1 启动子上的染色质改变与其在前列腺癌中的失活有关).Mol Carcinog(分子癌症发生),2007.46(10) :p.839-46)。GSTP1mRNA 在 Hela 和 HCT15 细胞系中高度表达 ; 在 PC3 细胞系中, GSTP1mRNA 表达的水平大约为 Hela 和 HCT15 细胞中发现的水平的一半。

[0195] 实时 PCR. 分离自用 MnlI 处理的细胞的 DNA 样品使用每种引物组在装有 Chromo4 连续荧光检测器 (MJ Research) 的 PTC-200 热循环仪中分析。每个反应的体积为 20μl,并且包含 50ng DNA,500nM 每种引物和 10μl iQ SYBR green supermix (Bio-Rad)。PCR 条件为 96℃ 10 分钟 ;96℃ 30 秒和 68℃ 1 分钟进行 40 个循环 ;在 72℃下 5 分钟 ;在每个步骤中从 72℃至 98℃的熔解曲线以 0.2℃的间隔保持 5 秒。分离自用 DNase I 处理的细胞的 DNA 样品在 CFX96 实时 PCR 检测系统 (Bio-Rad) 中分析。每个反应的体积为 20μl 并且包含 5ng DNA,500nM 每种引物和 10μl 含有 ROX 的 iTaq Fast SYBR green supermix(Bio-Rad)。PCR 条件为 95℃ 5 分钟 ;95℃ 30 秒和 72℃ 1 分钟进行 40 个循环 ;72℃下 5 分钟 ;在每个步骤中从 72℃至 95℃的熔解曲线以 0.2℃的间隔保持 5 秒。

[0196] 数据分析 . 为了使用 MnlI 作为核酸酶探针确定不可接近的染色质的水平,我们进行了下列的计算。当分析分离自用缺少 MnlI 的消化 / 透化缓冲液处理的细胞的 DNA 时, ΔCt(未切割的) 计算为 (完整的 Ct- 总 Ct)。当分析分离自用 MnlI 处理的细胞的 DNA 时, ΔCt(切割的) 计算为 (完整的 Ct- 总 Ct)。不可接近的染色质的水平计算为 2^{Δ}(ΔCt(未切割的)-ΔCt(切割的))(参见图 4)。

[0197] 对于 GAPDH 和 GSTP1 为了使用 DNase I 确定不可接近的染色质的水平,我们不能使用分析 MnlI 时使用的相同计算,因为 DNase I 在可接近的染色质中的总扩增区和完整扩增区内均进行切割。我们发现 HBB 启动子高度耐受 DNase I 消化,我们使用 HBB ΔCt 作为内参照标准,其反映了不可接近的染色质的低水平消化 (图示在图 16 中)。ΔCt(参照) 计算为 (HBB+ 核酸酶 Ct-HBB 无核酸酶 Ct)。ΔCt(靶) 计算为 GAPDH/GSTP1+ 核酸酶 Ct-GAPDH/GSTP1 无核酸酶 Ct。GAPDH 或 GSTP1 靶基因的不可接近的染色质的水平计算为 2^{Δ}(ΔCt(参照)-ΔCt(靶))。

[0198] 结果

[0199] Hela 细胞在下面任一种中孵育 :无核酸酶的消化缓冲液,缺少溶血卵磷脂 (细胞透化试剂) 和核酸酶的消化缓冲液,或含有溶血卵磷脂和核酸酶的完全消化缓冲液 (MnlI[M] 或 DNase I[D])。

[0200] 然后基因组 DNA(gDNA) 从细胞中分离并且进行琼脂糖凝胶电泳 (参见图 7)。数据表明溶血卵磷脂和核酸酶是消化 gDNA 必不可少的。在缺少溶血卵磷脂或核酸酶中任一种的反应中仅检测到轻微的 gDNA 消化。这表明在被溶血卵磷脂透化的细胞中,内源性核酸酶

PARSE EX.1007.323

对染色质的消化不明显，并且在缺少溶血卵磷脂的条件下消化缓冲液中存在的核酸酶不进入细胞中。

[0201] 随后，分离自 MnlI 处理的细胞的 DNA 样品使用图 6 中所示的引物组通过实时 PCR 分析。如图 8 中所示，阴性对照反应分析血红蛋白 β 链基因（其在 Hela 细胞中是沉默的）。对于总引物组和完整引物组两者，Ct 值是大致相同的。这表明在 Hela 细胞中此基因区处于不可接近的染色质结构中。

[0202] 图 9 显示了阳性对照 GAPDH 的结果。GAPDH 的 Ct 值与阴性对照中对 HBB 获得的 Ct 值大致是相同的，但使用“完整”引物组的扩增右移，表明与“总”DNA 相比不太完整。这意味着 MnlI 接近 GAPDH 启动子中的其识别位点并在该处消化。因此，我们推断在 Hela 细胞中 GAPDH 启动子处于可接近的染色质中。

[0203] 图 10 显示了分析 GST pi 的结果。GST pi 在前列腺癌中是失活的并且在 Hela 细胞中表达。在完整引物组反应中完全反应的 Ct 值右移，但在总引物组反应中不右移。这表明在 Hela 细胞中 GST pi 处于可接近的染色质结构中。

[0204] 图 11 在表中总结了上述数据。计算的值是与通过“总”引物组确定的总拷贝数相比完整的拷贝的百分比（并且因此对核酸酶是不可接近的）。

[0205] 实施例 3：在前列腺癌细胞中使用 MnlI 的数据

[0206] 上述的方法也应用于不同的细胞类型以确定可接近性是否与 GSTP1 的表达相关。使用两个细胞类型：

[0207] RWPE-1 细胞，其来自非癌性人前列腺并且高度表达 GSTP1；和 LNCaP 细胞，其来自人前列腺癌并且其中 GSTP1 被表观遗传修饰沉默。

[0208] 分析这些细胞的方法基本上如上所述。图 12 显示了对照反应（即，针对血红蛋白基因）。在两个细胞系中，无酶样品中的 ΔCt 与有酶样品的 ΔCt 大致相同。因此 ΔCt 大约为 0，表明在两个细胞系中此基因是不可接近的。

[0209] 图 13 显示分析 GAPDH 基因的阳性对照反应的结果。对于无酶的反应，总 Ct 和完整 Ct 两种大致是相同的。然而，对于有酶的反应，完整 Ct 右移，表明存在较少完整的 DNA，因此 GAPDH 基因处于可接近的染色质结构中。

[0210] 图 14 显示来自分析 GST pi 基因的数据。在非癌 RWPE-1 细胞系中，在酶的存在下完整 Ct 右移。相反，在癌性 LNCaP 细胞系中，完整 Ct 没有改变至可察觉的程度。这表明在 RWPE-1 细胞中 GSTP1 处于可接近的染色质中，在 LNCaP 细胞中处于不可接近的染色质中。

[0211] 这些实验重复 4 次。结果的总结显示在图 15 中，其再次使用不可接近（或“锁住”）的拷贝的百分比计算。

[0212] 实施例 4：使用 DNase I 的数据

[0213] 本实施例显示使用下述方法获得的数据，其中 DNase I 而非 MnlI 是核酸酶。在此实施例中，我们分析 4 种人癌细胞系：Hela（宫颈癌），PC3（前列腺癌），LNCaP（前列腺癌）和 HCT 15（结肠癌）。我们还分析在 MnlI 实施例中分析的相同基因：HBB，GAPDH 和 GSTP1。

[0214] 因为 DNase I 不在限定位点处消化 DNA，因此我们不能使用当 MnlI 为染色质结构探针时使用的相同数据分析策略。代之我们利用了表观遗传沉默的参照基因 (HBB) 作为内部对照，其反映了不可接近的染色质的消化程度。使用此策略，可以如图 16 中所详述确定特定靶基因的染色质结构。图 17 是分析的基因和用来评估染色质结构的引物的位置的示

PARSE EX.1007.324

意性图示。

[0215] 如图 18 中所示，在所有细胞系中 DNase I 均未将 HBB 消化至可察觉的程度（无合成酶和 + 核酸酶曲线几乎重叠）。此结果表明在所有细胞系中 HBB 处于不可接近的染色质构型中，并且验证了其作为参照基因对照的应用。这些结果也是预期的，因为 HBB mRNA 在这些细胞系的任一个中均不表达（未显示的数据）。

[0216] 我们对 GAPDH 基因的分析显示在图 19 中。在所有细胞系中，我们观察到 + 核酸酶曲线相对于无核酸酶曲线显著右移。这提示在所有细胞系中 GAPDH 处于可接近的染色质结构中。这些结果是预期的，因为 GAPDH 是“管家”基因，其在大多数细胞系（如果不在所有细胞系中表达的话）中表达（未显示的数据）。

[0217] GSTP1mRNA 以不同的水平在分析的细胞系中表达。Hela 和 HCT 15 细胞表达大量的 GSTP1mRNA，PC3 细胞表达中等量（约为 Hela 和 HCT 15 细胞中发现的一半），LNCaP 细胞不以可察觉的程度表达 GSTP 1mRNA（未显示的数据）。我们对 GSTP1 染色质结构的分析显示在图 20 中。在 Hela 和 HCT 15 细胞中，我们观察到 + 核酸酶曲线相对于无核酸酶曲线显著右移，提示 GSTP1 处于可接近的染色质结构中。在 PC3 细胞中，我们观察到 + 核酸酶曲线相对于无核酸酶曲线中等程度右移，提示 GSTP1 处于部分可接近的染色质结构中。最后，在 LNCaP 细胞中，我们观察到 + 核酸酶曲线和无核酸酶曲线几乎重叠。这提示在 LNCaP 细胞中，GSTP1 处于不可接近的染色质构型中。很显然，我们在分析的所有细胞系中观察到了 GSTP1 染色质结构和 GSTP1mRNA 表达之间的良好相关性。

[0218] 图 21 显示了结果的总结。如图 16 中所述计算不可接近的 DNA 的百分比，并且显示为“锁住的”染色质。通过标准方法通过 qRT-PCR 分析分离自不同细胞系的 mRNA 来计算 GAPDH 和 GSTP1RNA 表达的相对水平。结果表明靶基因染色质结构与靶基因表达水平的相关性良好。这提示所述的测定代表了一种用来评估对基因表达的表观遗传作用的有用工具。

[0219] 总结

[0220] 我们已经开发了一种快速的和灵敏的测定，该测定可以评估培养细胞中的染色质结构。我们的测定提供了几个优于其他染色质结构测定的优势 (Okino, ST. 等，Mol Carcinog（分子癌症发生）46(10)：839-46(2007)；Okino, S.T. 和 J.P.Whitlock, Jr.，Mol Cell Biol（分子细胞生物学）15(7)：p.3714-21(1995)；Rao, S.，E.Procko, 和 M.F.Shannon, J Immunol（免疫学杂志）167(8)：4494-503(2001)；Kilgore, J.A. 等，Methods（方法）41(3)：320-32(2007))。最显著地，在我们的方法中，我们在一个步骤中原位消化了透化的细胞中的染色质。其他方法需要分离细胞核，这需要较多的劳力、较多的时间和较多的细胞作为起始材料。另外，我们的引物组设计得到了对染色质可接近性的更准确的和内部控制的评估。

[0221] 实施例 5：评估 DNA 甲基化的变型

[0222] 本发明的另一应用是确定不可接近的染色质区的 DNA 甲基化状态。DNA 甲基化是指在 5- 位处胞嘧啶的天然甲基化，产生 5- 甲基胞嘧啶。DNA 甲基化参与调节基因活性和染色质结构；DNA 甲基化的异常谱型与人类病理学和特定疾病状态有关。在此变型中，遵照以前所述的方法，但不包括 PCR 扩增的步骤。然后使用标准方法通过亚硫酸氢盐修饰纯化的基因组 DNA，使得未甲基化的胞嘧啶转变为尿嘧啶，并且 5- 甲基胞嘧啶没有变化。然后靶向特定 DNA 区的引物组用来扩增从已经用核酸酶处理的细胞纯化的亚硫酸氢盐 - 修饰的

PARSE EX.1007.325

DNA。因此，仅最初处于不可接近的染色质构型并且因此不被 DNA 修饰剂切割的 DNA 可以被扩增。然后通过标准方法诸如熔解曲线分析、高分辨率熔解分析、亚硫酸氢盐 -DNA 测序、数字 PCR、甲基化特异的 PCR、焦磷酸测序等对扩增的 DNA 分析 DNA 甲基化。这样的分析可以确定不可接近的染色质区的 DNA 甲基化状态，并且可以提供关于与表观遗传基因失活相关的 DNA 甲基化的特定谱型的有价值信息。

[0223] 要理解本文所述的实施例和实施方案仅用于说明的目的，并且提议本领域技术人员据此作出多种修改或变化，并且这些修改或变化包括在本申请的精神和界限以及附带的权利要求的范围内。本文引用的所有出版物、专利和专利申请的全部内容通过引用合并于此用于所有目的。

PARSE EX.1007.326

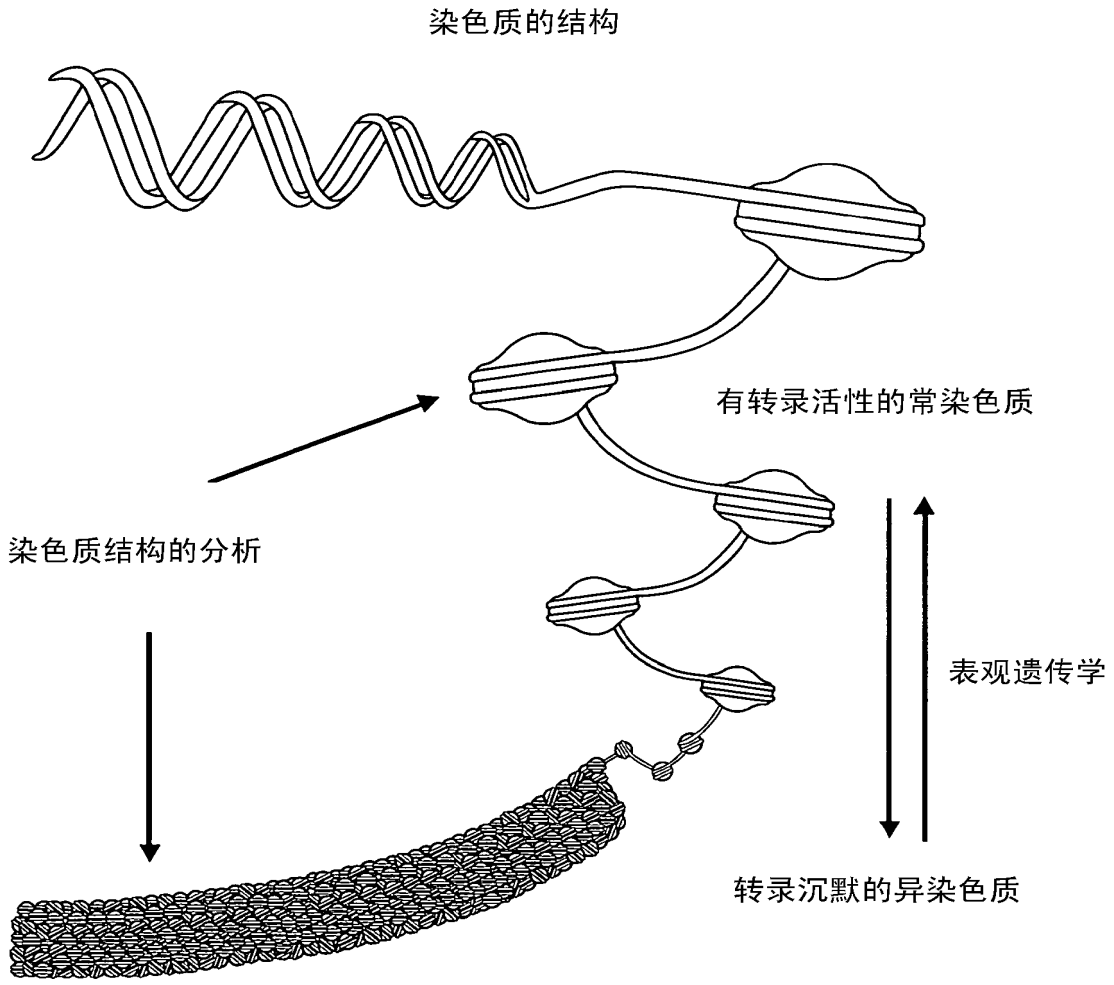


图 1

PARSE EX.1007.327

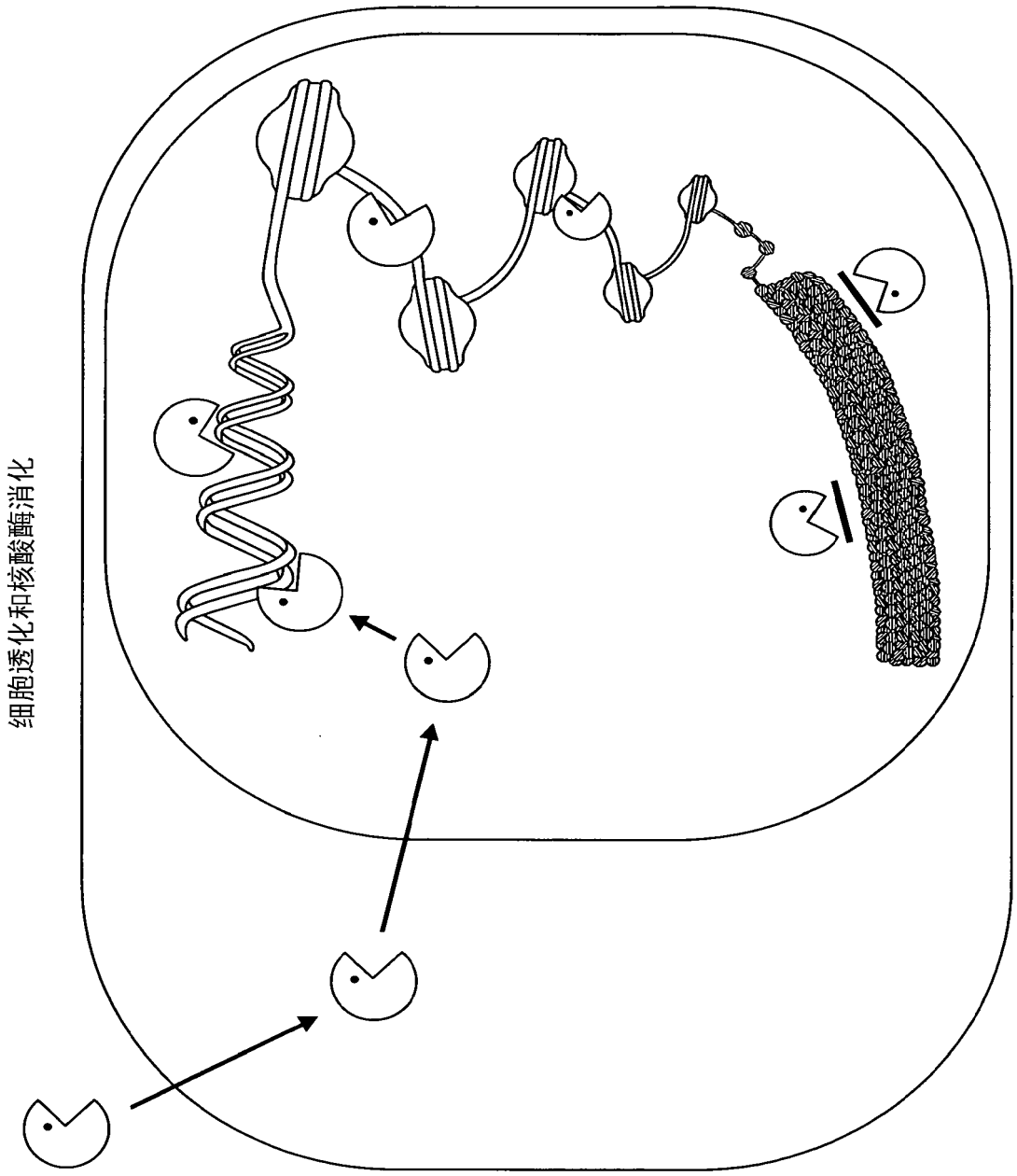


图 2

PARSE EX.1007.328

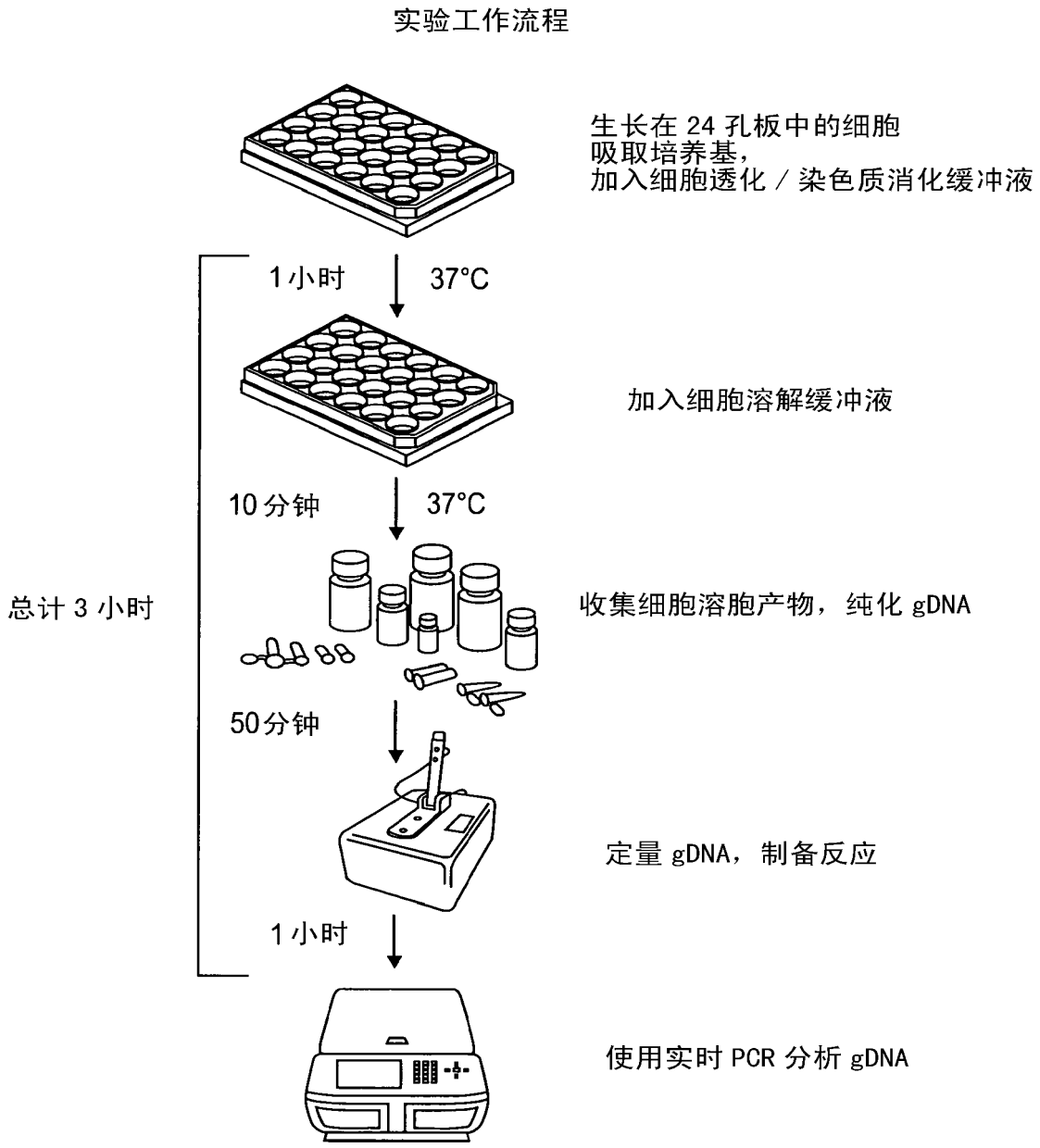


图 3

PARSE EX.1007.329

使用 MnI 1 作为探针分析靶基因

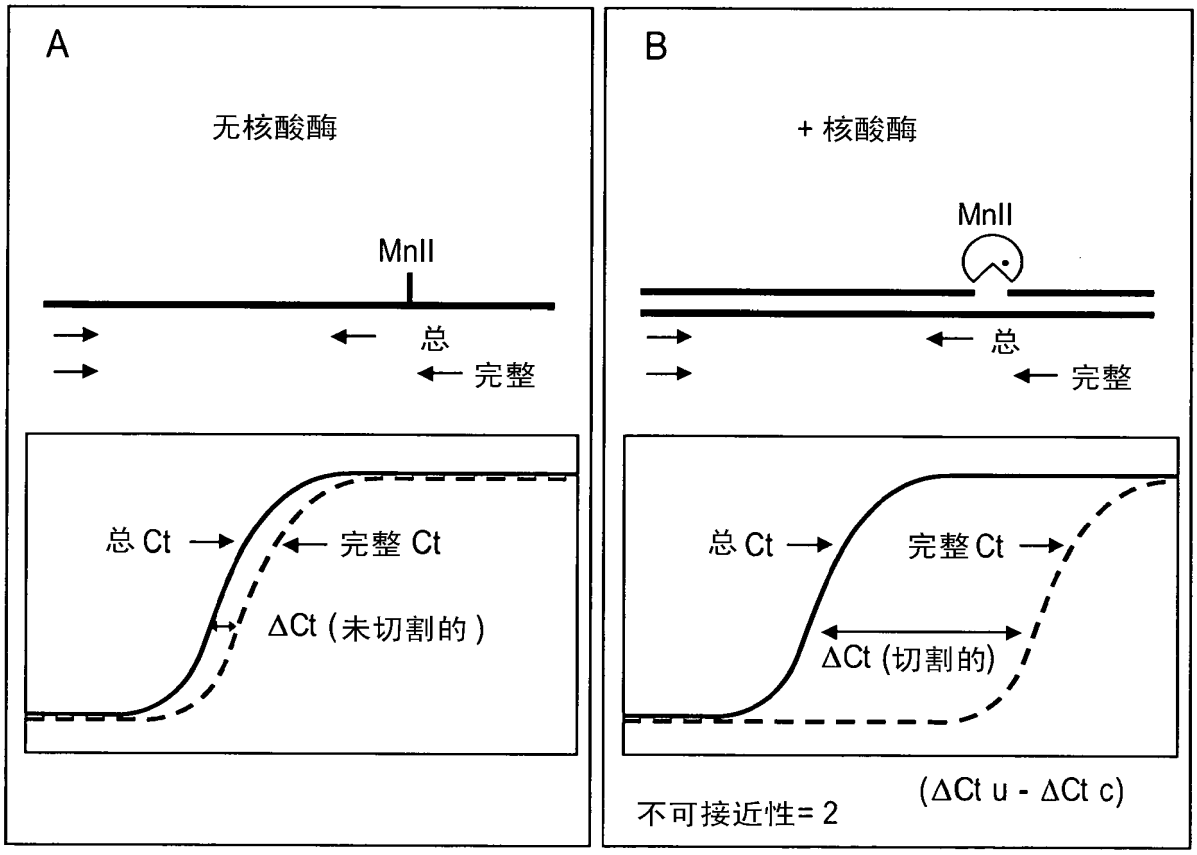


图 4

PARSE EX.1007.330

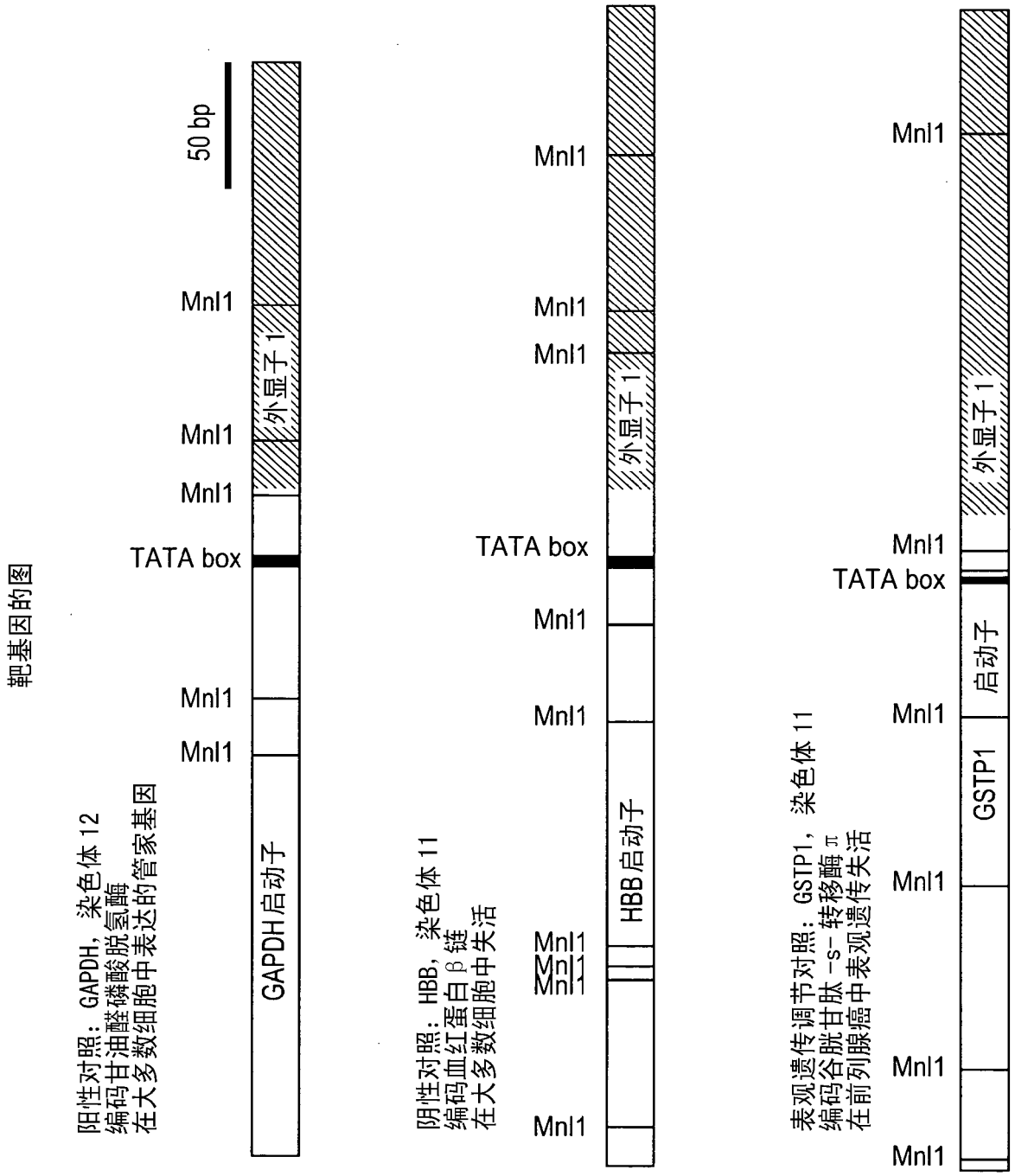


图 5

PARSE EX.1007.331

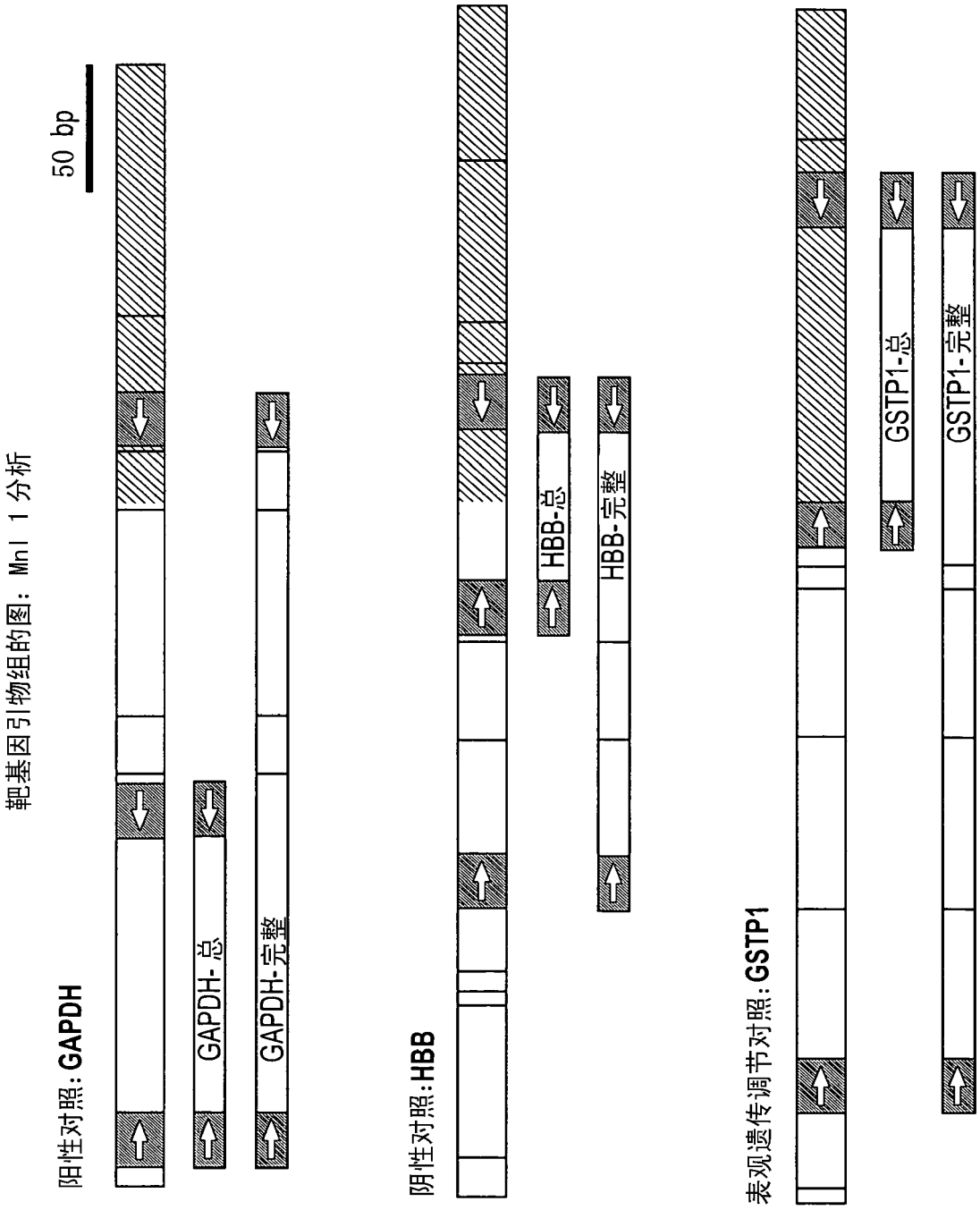


图 6

PARSE EX.1007.332

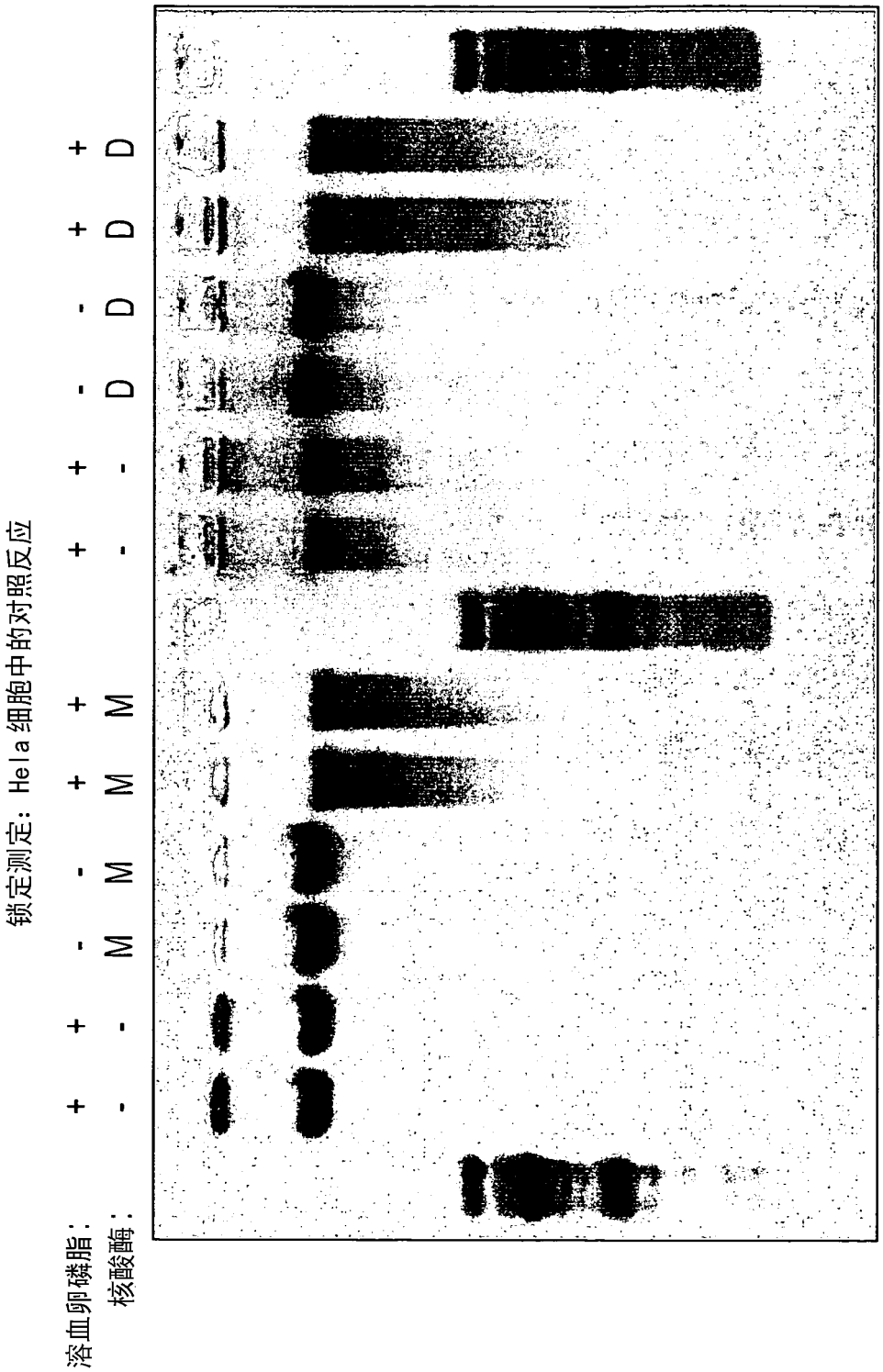


图 7

PARSE EX.1007.333

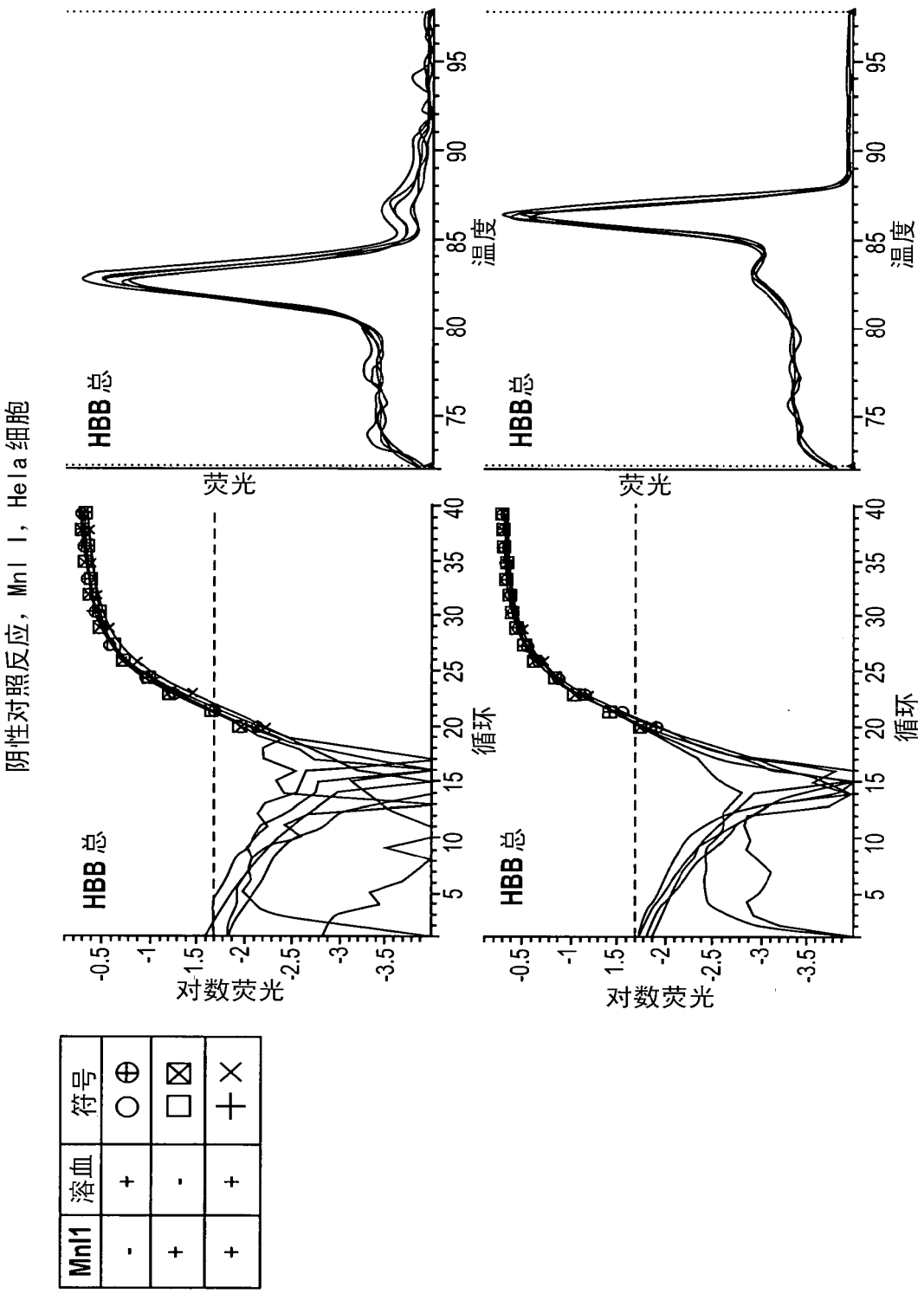


图 8

PARSE EX.1007.334

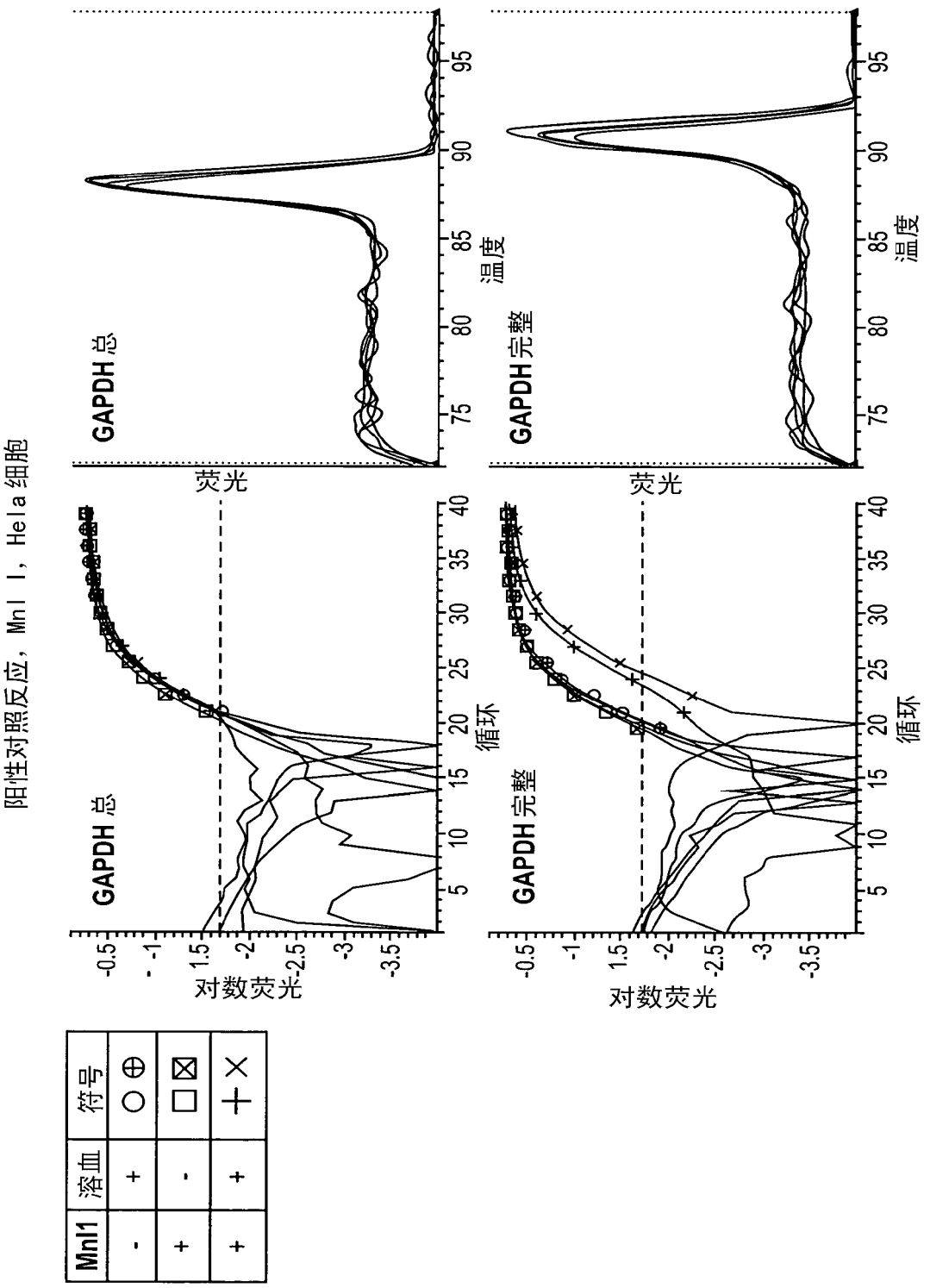


图 9

PARSE EX.1007.335

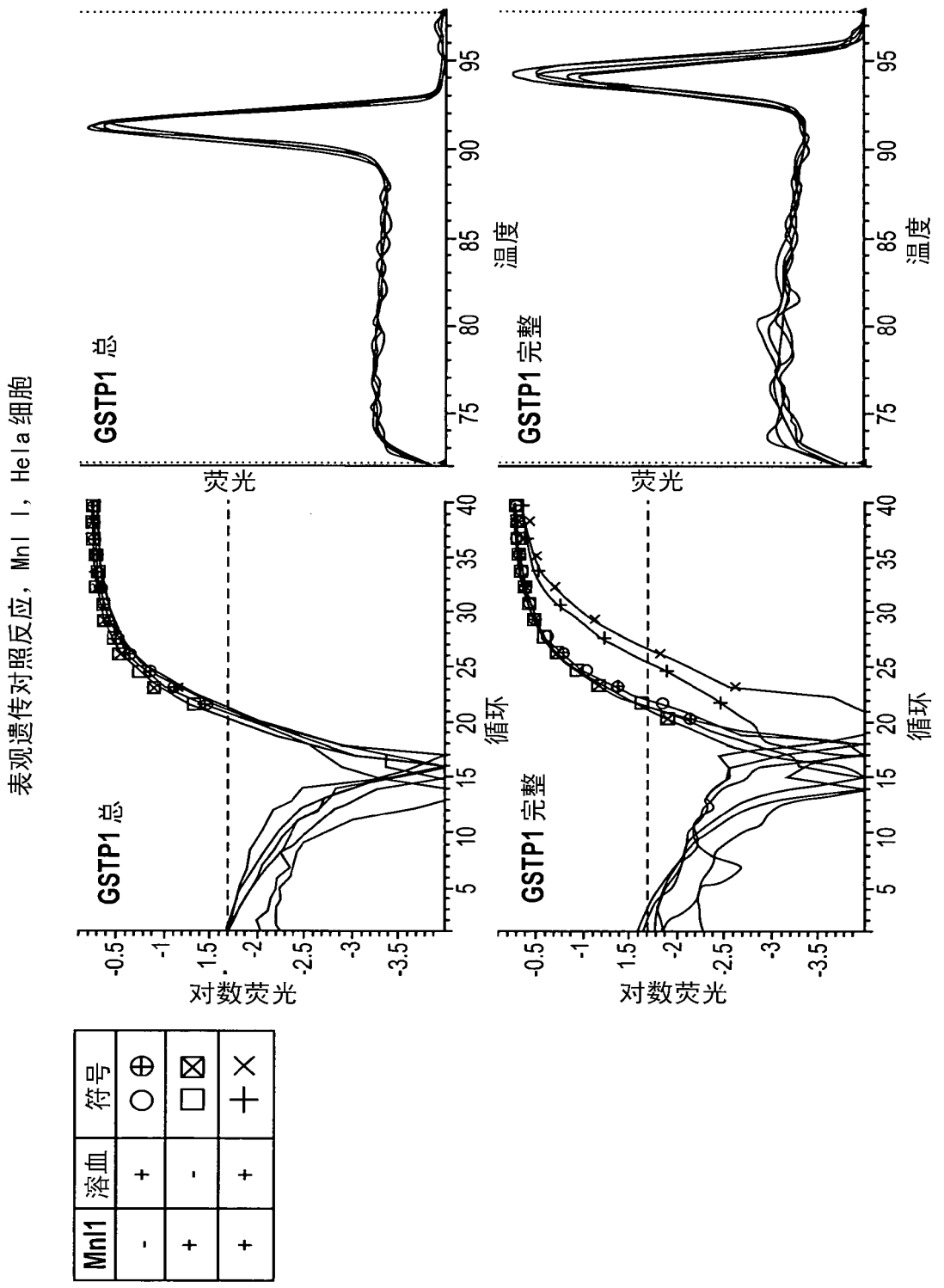


图 10

PARSE EX.1007.336

对照反应计算，Hela 细胞

酶	溶血	HBB 锁住的	GAPDH 锁住的	GSTP1 锁住的
-	+	STD	STD	STD
MnlI	-	99%	109%	101%
MnlI	+	131%	6.3%	5.7%

图 11

PARSE EX.1007.337

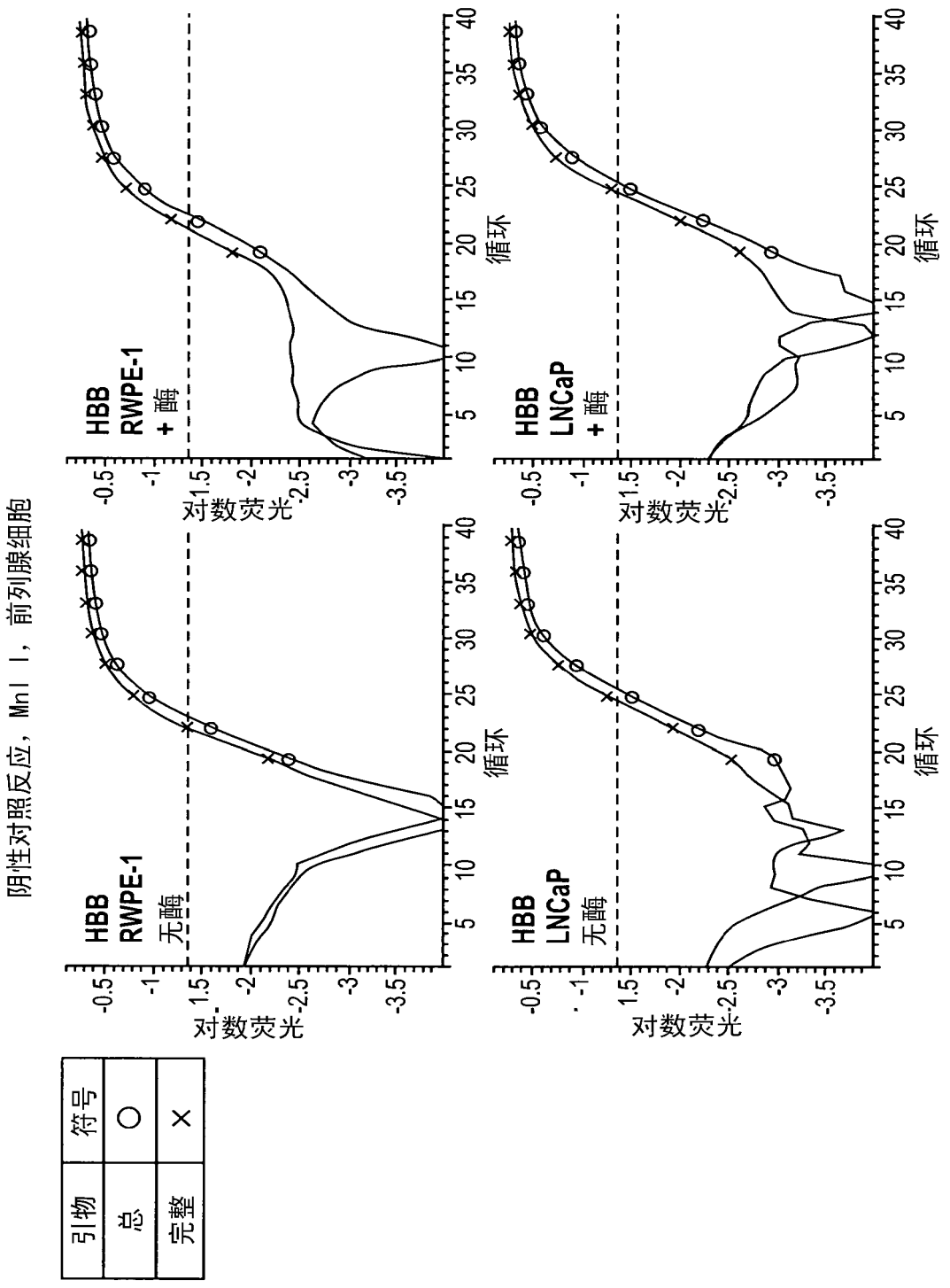


图 12

PARSE EX.1007.338

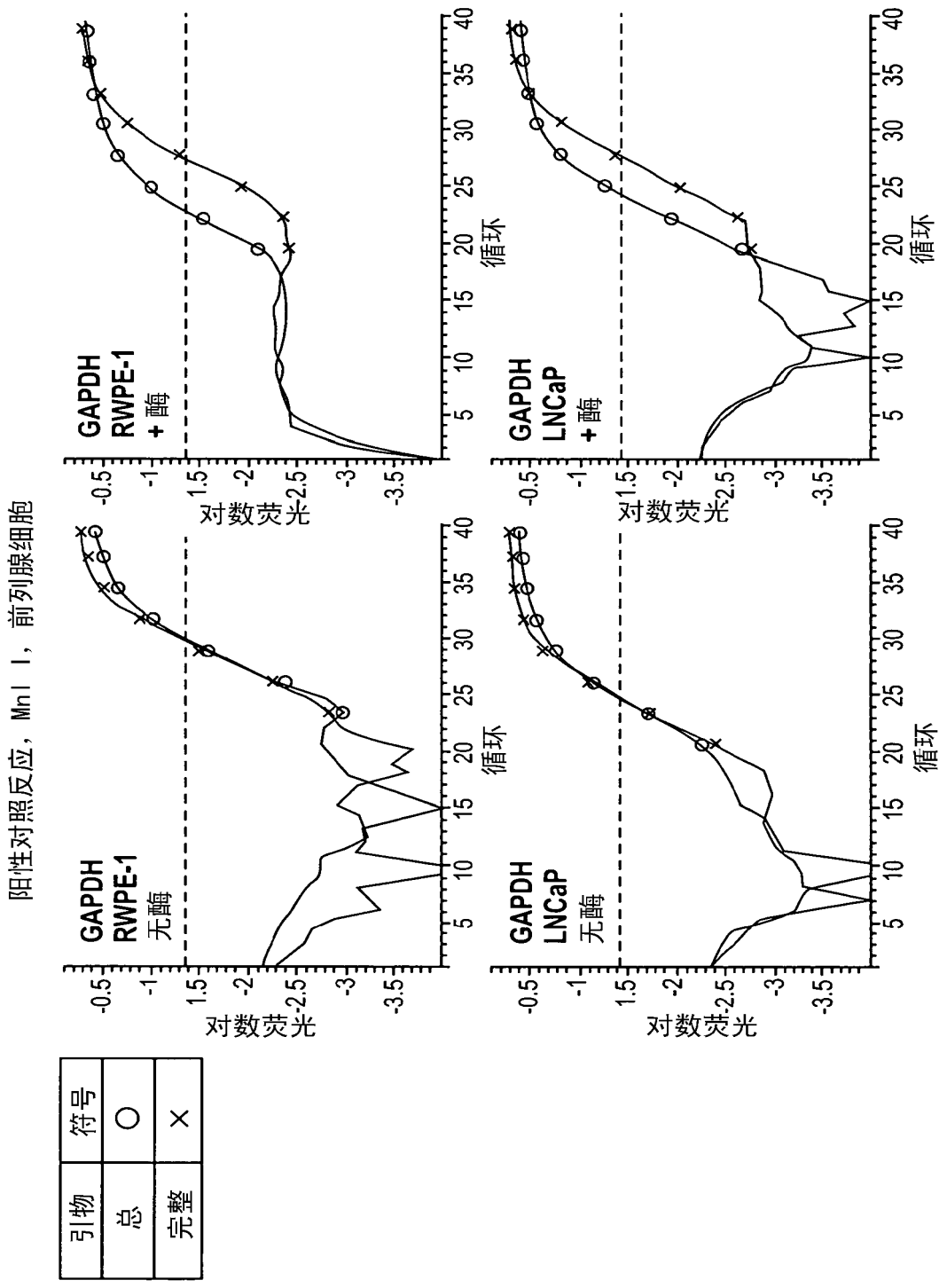


图 13

PARSE EX.1007.339

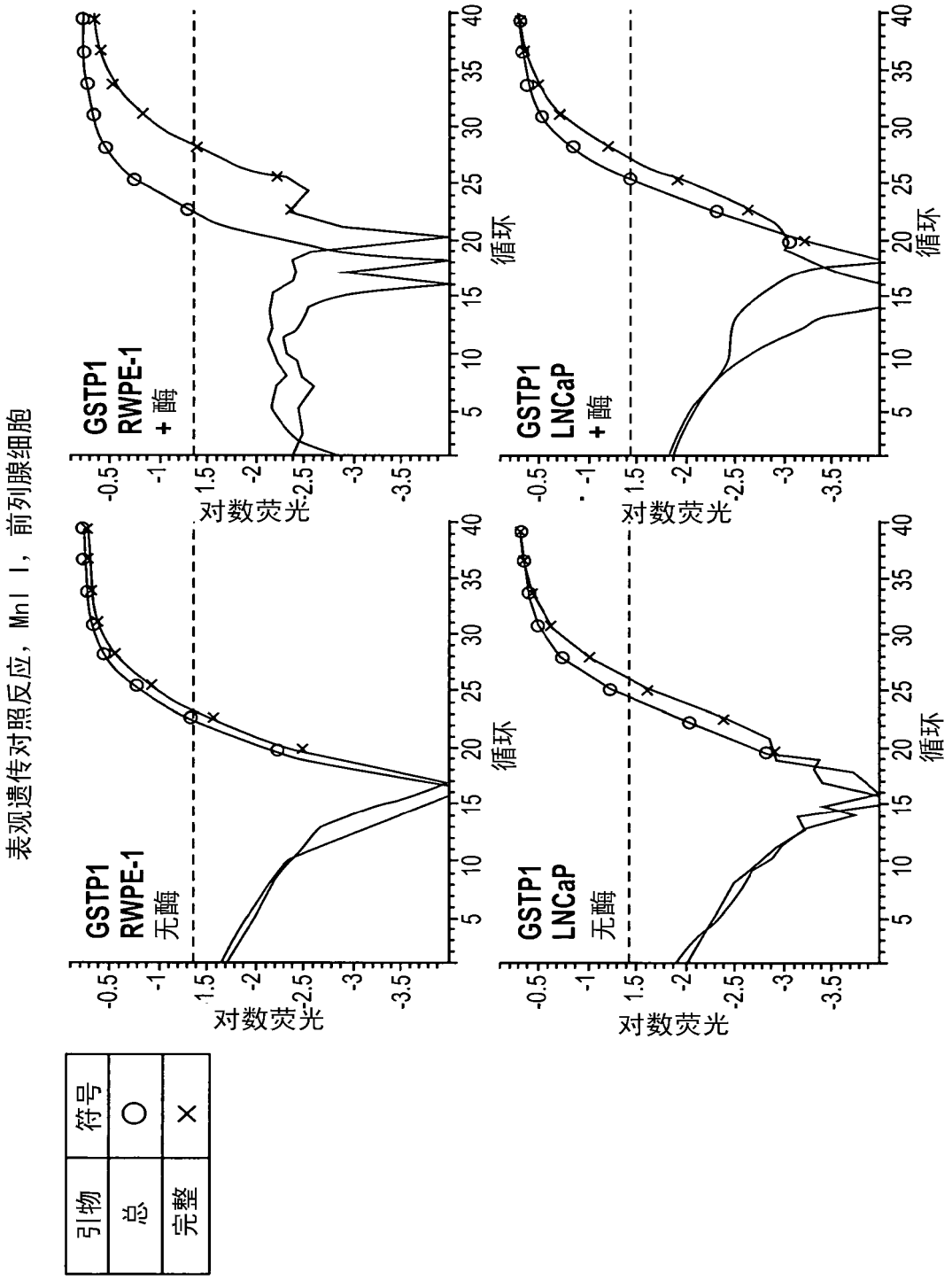


图 14

PARSE EX.1007.340

计算，前列腺细胞

细胞	核酸酶	HBB	GAPDH	GSTP1
		锁住的	锁住的	锁住的
RWPE-1	MnlI	97%	4.4%	6.3%
LNCaP	MnlI	92%	11%	96%

图 15

PARSE EX.1007.341

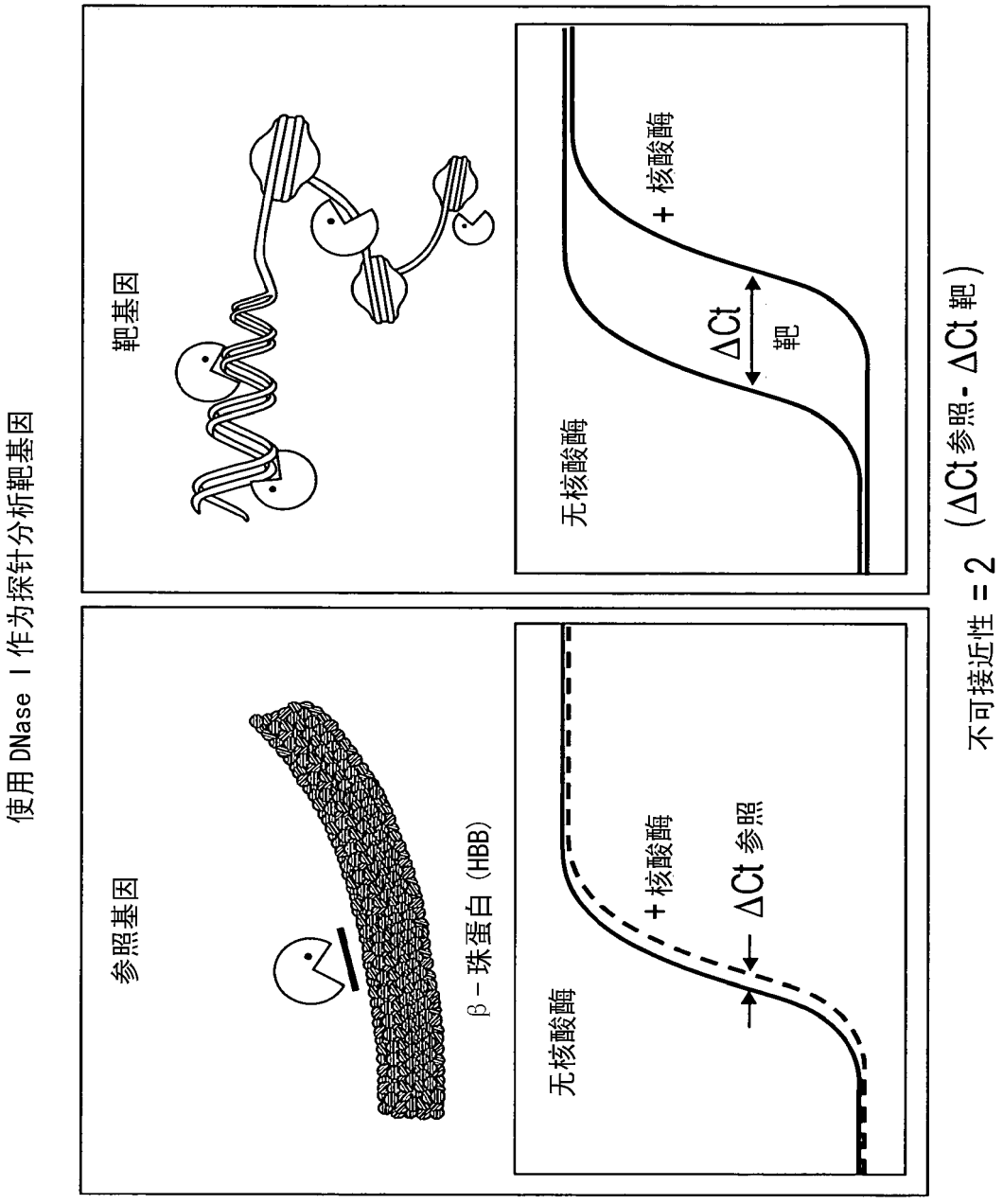


图 16

PARSE EX.1007.342

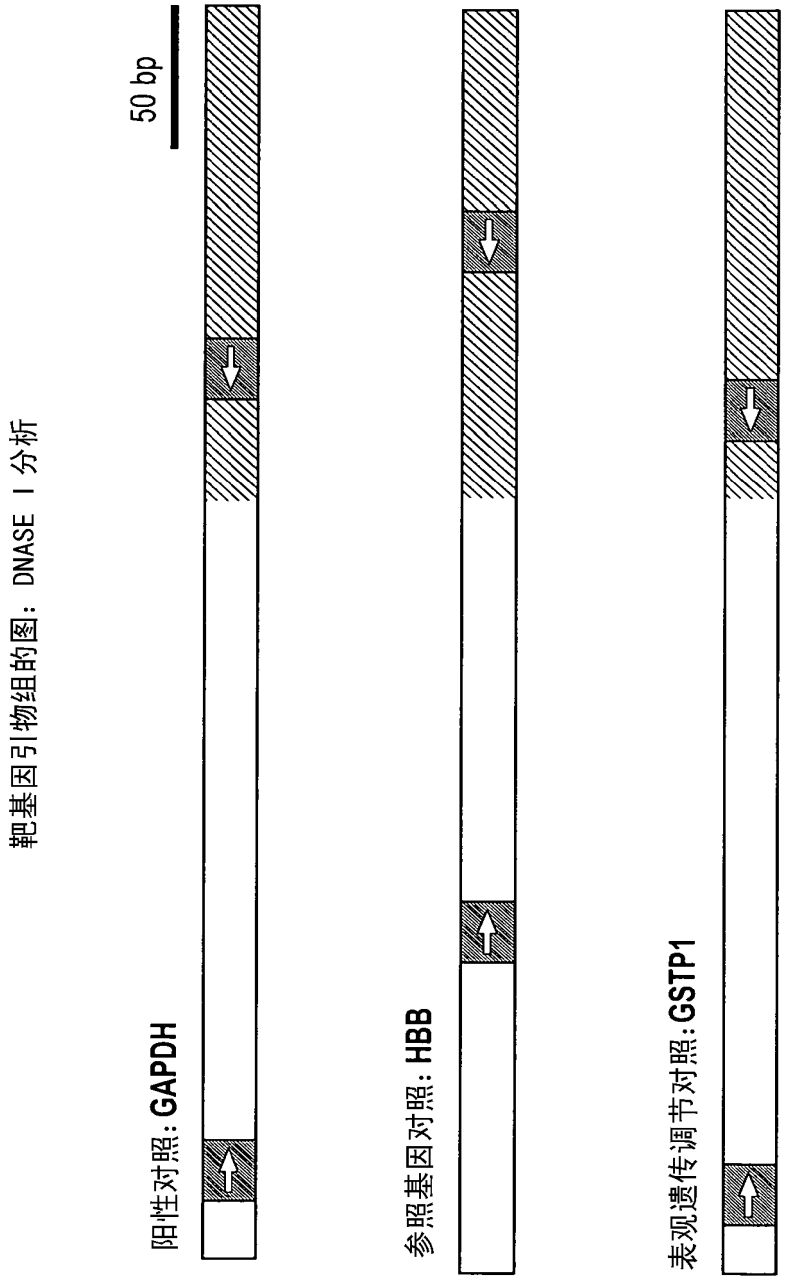


图 17

PARSE EX.1007.343

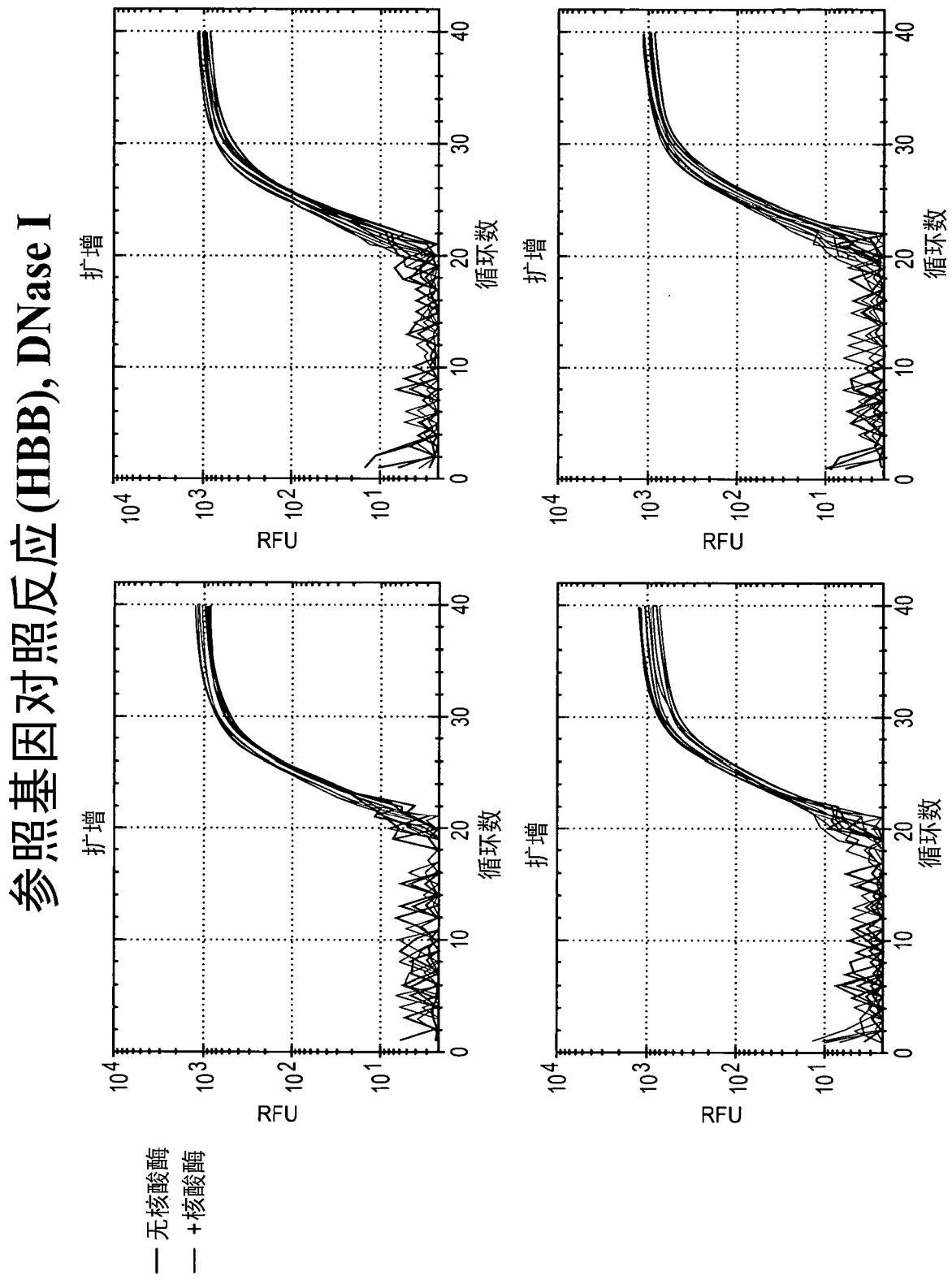


图 18

PARSE EX.1007.344

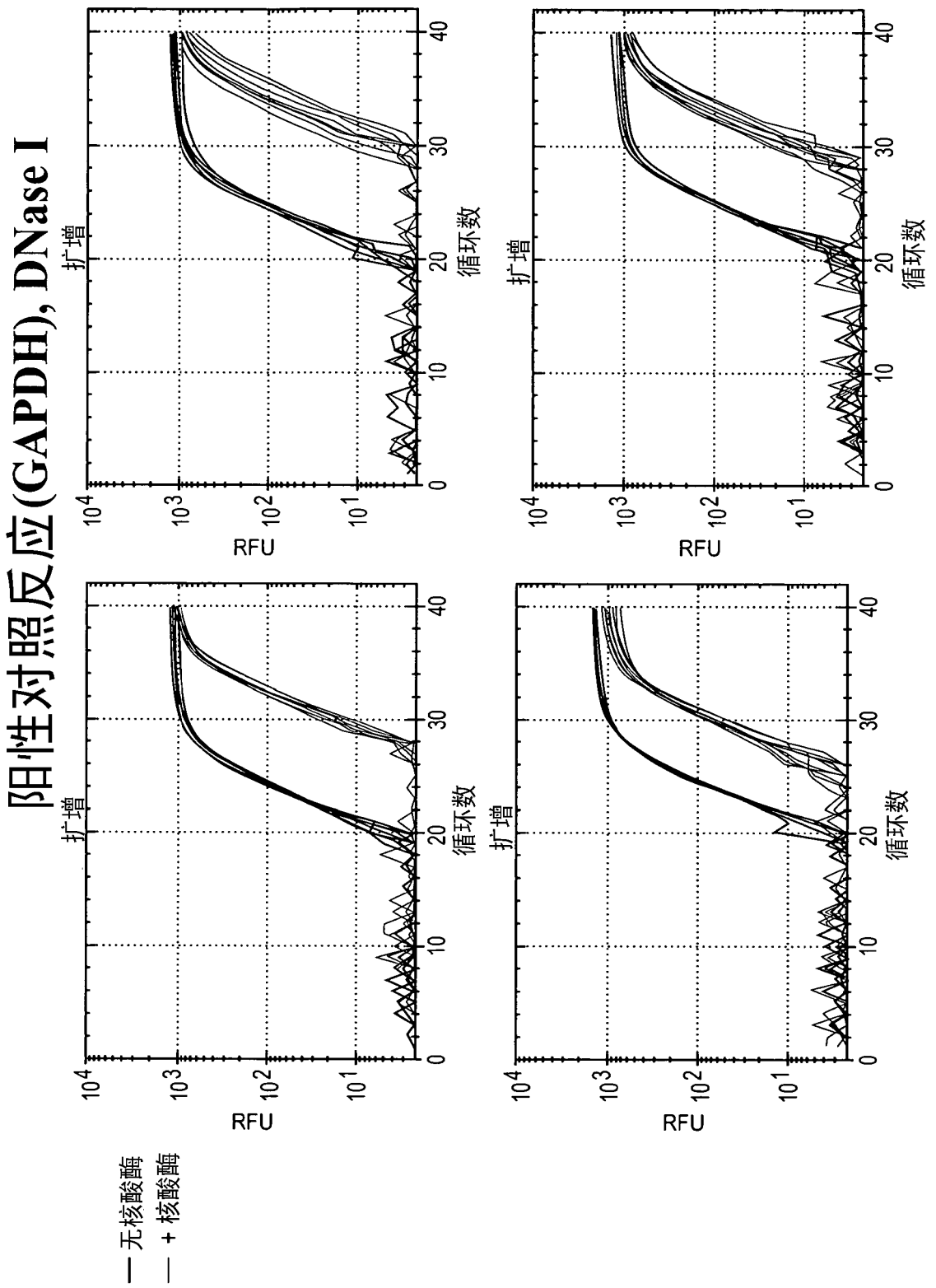


图 19

PARSE EX.1007.345

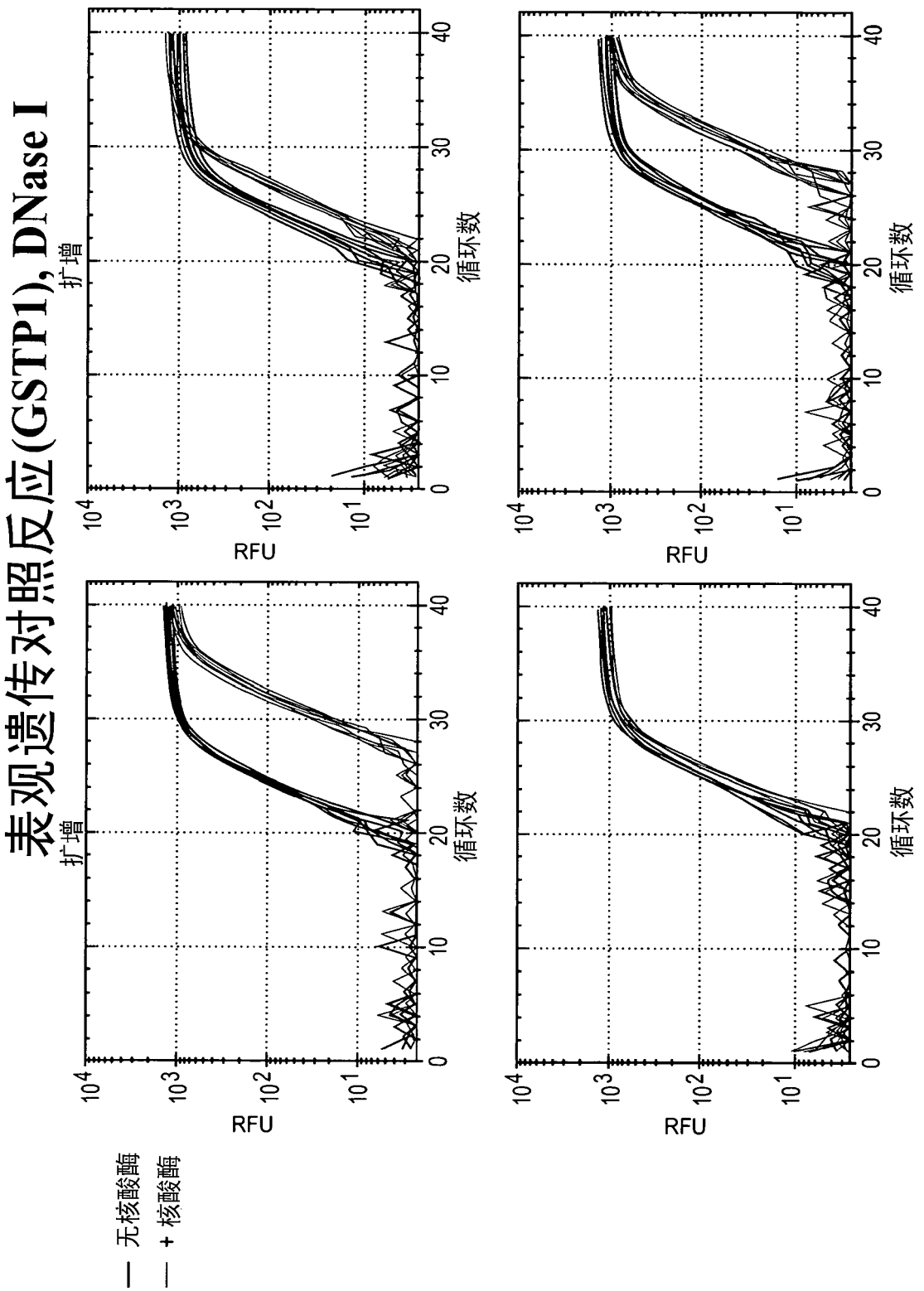


图 20

PARSE EX.1007.346

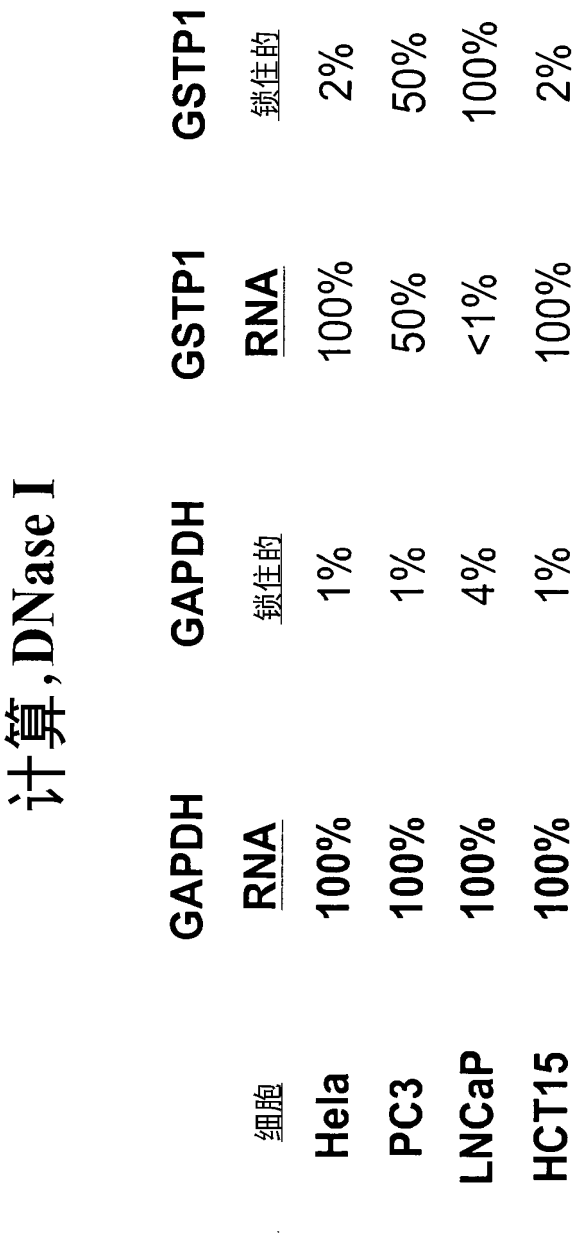


图 21

PARSE EX.1007.347

Electronic Acknowledgement Receipt

EFS ID:	33945744
Application Number:	16043874
International Application Number:	
Confirmation Number:	2837
Title of Invention:	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS
First Named Inventor/Applicant Name:	Paul Giresi
Customer Number:	77974
Filer:	James Samuel Keddie/Sherry Long
Filer Authorized By:	James Samuel Keddie
Attorney Docket Number:	STAN-1111CON
Receipt Date:	08-OCT-2018
Filing Date:	24-JUL-2018
Time Stamp:	18:34:45
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Information Disclosure Statement (IDS) Form (SB08)	STAN-1111CON_IDS.pdf	1035316 9832e787fc3d9714d58ce4495869dda675b3736b	no	4

Warnings:

Information:					
2	Foreign Reference	WO2012106546A2.pdf	6343226 9eada39cb88b245e05105fbfcf29eda21d2c a6d9	no	127
Warnings:					
Information:					
3	Foreign Reference	CN102292455A.pdf	2772338 d30cab48928c09360c6112e16651c8f3306 382cc	no	50
Warnings:					
Information:					
4	Non Patent Literature	NPL.pdf	1002146 ebbe6bae2c0aed06dc6283b77fe28498b31 6cb7a	no	17
Warnings:					
Information:					
			Total Files Size (in bytes):	11153026	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON	2837

77974 7590 10/18/2018

Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

EXAMINER	
WILDER, CYNTHIA B	

ART UNIT	PAPER NUMBER
1637	

NOTIFICATION DATE	DELIVERY MODE
10/18/2018	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket@bozpat.com

PARSE EX.1007.350

Corrected ***Notice of Allowability***	**Application No.** 16/043,874	**Applicant(s)** Giresi et al.	
	Examiner CYNTHIA B WILDER	**Art Unit** 1637	**AIA Status** Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☑ This communication is responsive to 10/2/2018.
   ☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2. ☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

3. ☑ The allowed claim(s) is/are 78-100 . As a result of the allowed claim(s), you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see **http://www.uspto.gov/patents/init_events/pph/index.jsp** or send an inquiry to **PPHfeedback@uspto.gov.**

4. ☑ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

   Certified copies:

   a) ☑ All b) ☐ Some *c) ☐ None of the:

   1. ☐ Certified copies of the priority documents have been received.
   2. ☑ Certified copies of the priority documents have been received in Application No. 14784250 .
   3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

   * Certified copies not received: _____ .

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file areply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.
THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
   ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date _____ .

   Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).

6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☑ Information Disclosure Statements (PTO/SB/08), Paper No./Mail Date 10/8/2018.
3. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material _____.
4. ☐ Interview Summary (PTO-413), Paper No./Mail Date. _____.
5. ☐ Examiner's Amendment/Comment
6. ☐ Examiner's Statement of Reasons for Allowance
7. ☐ Other _____.

/CYNTHIA B WILDER/
Primary Examiner, Art Unit 1637

PARSE EX.1007.351

Doc code: IDS
Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/08a (03-15)
Approved for use through 07/31/2016. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16043874
Filing Date	2018-07-24
First Named Inventor	Giresi, Paul
Art Unit	1637
Examiner Name	WILDER, CYNTHIA B.
Attorney Docket Number	STAN-1111CON

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1					

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
/C.B.W/	1	20140274740		2014-09-18	Srinivasan et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
/C.B.W/	1	WO2012106546	WO	A2	2012-08-09	SHENDURE et al.		
/C.B.W/	2	CN102292455	CN	A	2011-12-21	Bio-Rad Laboratories, Inc.		

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

/CYNTHIA B WILDER/ (10/15/2018) 10/15/2018

PARSE EX.1007.352

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	16043874
Filing Date	2018-07-24
First Named Inventor	Giresi, Paul
Art Unit	1637
Examiner Name	WILDER, CYNTHIA B.
Attorney Docket Number	STAN-1111CON

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
/C.B.W/	1	Epicentre., "EZ-Tn5TM Custom Transposome Construction Kits", www.epicentre.com, pages 1-17, 2012.	

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature	/CYNTHIA B WILDER/ (10/15/2018)	Date Considered	10/15/2018

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

PART B - FEE(S) TRANSMITTAL

Complete and send this form, together with applicable fee(s), to: Mail **Mail Stop ISSUE FEE**
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
or Fax (571)-273-2885

INSTRUCTIONS: This form should be used for transmitting the ISSUE FEE and PUBLICATION FEE (if required). Blocks 1 through 5 should be completed where appropriate. All further correspondence including the Patent, advance orders and notification of maintenance fees will be mailed to the current correspondence address as indicated unless corrected below or directed otherwise in Block 1, by (a) specifying a new correspondence address; and/or (b) indicating a separate "FEE ADDRESS" for maintenance fee notifications.

CURRENT CORRESPONDENCE ADDRESS (Note: Use Block 1 for any change of address)

77974 7590 10/02/2018
Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

Note: A certificate of mailing can only be used for domestic mailings of the Fee(s) Transmittal. This certificate cannot be used for any other accompanying papers. Each additional paper, such as an assignment or formal drawing, must have its own certificate of mailing or transmission.

Certificate of Mailing or Transmission
I hereby certify that this Fee(s) Transmittal is being deposited with the United States Postal Service with sufficient postage for first class mail in an envelope addressed to the Mail Stop ISSUE FEE address above, or being facsimile transmitted to the USPTO (571) 273-2885, on the date indicated below.

(Depositor's name)
(Signature)
(Date)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON	2837

TITLE OF INVENTION: TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

APPLN. TYPE	ENTITY STATUS	ISSUE FEE DUE	PUBLICATION FEE DUE	PREV. PAID ISSUE FEE	TOTAL FEE(S) DUE	DATE DUE
nonprovisional	SMALL	$500	$0.00	$0.00	$500	01/02/2019

EXAMINER	ART UNIT	CLASS-SUBCLASS
WILDER, CYNTHIA B	1637	435-006000

1. Change of correspondence address or indication of "Fee Address" (37 CFR 1.363).

☐ Change of correspondence address (or Change of Correspondence Address form PTO/SB/122) attached.

☐ "Fee Address" indication (or "Fee Address" Indication form PTO/SB/47; Rev 03-02 or more recent) attached. **Use of a Customer Number is required.**

2. For printing on the patent front page, list
(1) The names of up to 3 registered patent attorneys or agents OR, alternatively,
(2) The name of a single firm (having as a member a registered attorney or agent) and the names of up to 2 registered patent attorneys or agents. If no name is listed, no name will be printed.

1 James S. Keddie
2 Bozicevic, Field & Francis LLP
3

3. ASSIGNEE NAME AND RESIDENCE DATA TO BE PRINTED ON THE PATENT (print or type)

PLEASE NOTE: Unless an assignee is identified below, no assignee data will appear on the patent. If an assignee is identified below, the document has been filed for recordation as set forth in 37 CFR 3.11. Completion of this form is NOT a substitute for filing an assignment.

(A) NAME OF ASSIGNEE: The Board of Trustees of the Leland Stanford Junior University

(B) RESIDENCE: (CITY and STATE OR COUNTRY): Stanford, California

Please check the appropriate assignee category or categories (will not be printed on the patent) : ☐ Individual ☒ Corporation or other private group entity ☐ Government

4a. The following fee(s) are submitted:

☒ Issue Fee

☒ Publication Fee (No small entity discount permitted)

☐ Advance Order - # of Copies _____

4b. Payment of Fee(s): **(Please first reapply any previously paid issue fee shown above)**

☐ A check is enclosed.

☒ Payment by credit card. ~~Form PTO-2038 is attached.~~

☒ The director is hereby authorized to charge the required fee(s), any deficiency, or credits any overpayment, to Deposit Account Number 50-0815 (enclose an extra copy of this form).

5. **Change in Entity Status** (from status indicated above)

☐ Applicant certifying micro entity status. See 37 CFR 1.29

☐ Applicant asserting small entity status. See 37 CFR 1.27

☐ Applicant changing to regular undiscounted fee status.

NOTE: Absent a valid certification of Micro Entity Status (see forms PTO/SB/15A and 15B), issue fee payment in the micro entity amount will not be accepted at the risk of application abandonment.
NOTE: If the application was previously under micro entity status, checking this box will be taken to be a notification of loss of entitlement to micro entity status.
NOTE: Checking this box will be taken to be a notification of loss of entitlement to small or micro entity status, as applicable.

NOTE: This form must be signed in accordance with 37 CFR 1.31 and 1.33. See 37 CFR 1.4 for signature requirements and certifications.

Authorized Signature /James S. Keddie/ Date October 26, 2018

Typed or printed name James S. Keddie Registration No. 48920

PTOL-85 Part B (10-13) Approved for use through 10/31/2013. OMB 0651-0033 U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PARSE EX.1007.354

Electronic Patent Application Fee Transmittal

Application Number:	16043874
Filing Date:	24-Jul-2018
Title of Invention:	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS
First Named Inventor/Applicant Name:	Paul Giresi
Filer:	James Samuel Keddie/Sherry Long
Attorney Docket Number:	STAN-1111CON

Filed as Small Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
PUBL. FEE- EARLY, VOLUNTARY, OR NORMAL	1504	1	0	0
UTILITY APPL ISSUE FEE	2501	1	500	500
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**500**

Electronic Acknowledgement Receipt

EFS ID:	34127889
Application Number:	16043874
International Application Number:	
Confirmation Number:	2837
Title of Invention:	TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS
First Named Inventor/Applicant Name:	Paul Giresi
Customer Number:	77974
Filer:	James Samuel Keddie/Sherry Long
Filer Authorized By:	James Samuel Keddie
Attorney Docket Number:	STAN-1111CON
Receipt Date:	26-OCT-2018
Filing Date:	24-JUL-2018
Time Stamp:	15:27:49
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	CARD
Payment was successfully received in RAM	$500
RAM confirmation Number	102918INTEFSW15405900
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Issue Fee Payment (PTO-85B)	STAN-1111CON_IssueFeeTransmittal.pdf	1001069 31dfe19d22f73ead02a9a6e243af911ef407fa0a	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	31848 674c10aa0468b7f02c16a822e84b21b1c13731d6	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	1032917		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	ISSUE DATE	PATENT NO.	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/043,874	12/11/2018	10150995	STAN-1111CON	2837

77974 7590 11/20/2018

Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

ISSUE NOTIFICATION

The projected patent number and issue date are specified above.

Determination of Patent Term Adjustment under 35 U.S.C. 154 (b)
(application filed on or after May 29, 2000)

The Patent Term Adjustment is 0 day(s). Any patent to issue from the above-identified application will include an indication of the adjustment on the front page.

If a Continued Prosecution Application (CPA) was filed in the above-identified application, the filing date that determines Patent Term Adjustment is the filing date of the most recent CPA.

Applicant will be able to obtain more detailed information by accessing the Patent Application Information Retrieval (PAIR) WEB site (http://pair.uspto.gov).

Any questions regarding the Patent Term Extension or Adjustment determination should be directed to the Office of Patent Legal Administration at (571)-272-7702. Questions relating to issue and publication fee payments should be directed to the Application Assistance Unit (AAU) of the Office of Data Management (ODM) at (571)-272-4200.

APPLICANT(s) (Please see PAIR WEB site http://pair.uspto.gov for additional applicants):

Paul Giresi, Palo Alto, CA;
The Board of Trustees of the Leland Stanford Junior University, Stanford, CA
Jason D. Buenrostro, Redwood City, CA;
Howard Y. Chang, Stanford, CA;
William J. Greenleaf, Menlo Park, CA;

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The USA offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to encourage and facilitate business investment. To learn more about why the USA is the best country in the world to develop technology, manufacture products, and grow your business, visit SelectUSA.gov.

PARSE EX.1007.359

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/043,874	07/24/2018	Paul Giresi	STAN-1111CON

CONFIRMATION NO. 2837

77974
Stanford University Office of Technology Licensing
Bozicevic, Field & Francis LLP
201 REDWOOD SHORES PARKWAY
SUITE 200
REDWOOD CITY, CA 94065

PUBLICATION NOTICE

OC000000104464882

Title:TRANSPOSITION OF NATIVE CHROMATIN FOR PERSONAL EPIGENOMICS

Publication No.US-2018-0355424-A1
Publication Date:12/13/2018

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Public Records Division. The Public Records Division can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Public Records Division, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently https://portal.uspto.gov/pair/PublicPair. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

PARSE EX.1007.360

AO 120 (Rev. 08/10)

TO:	**Mail Stop 8** **Director of the U.S. Patent and Trademark Office** **P.O. Box 1450** **Alexandria, VA 22313-1450**	**REPORT ON THE FILING OR DETERMINATION OF AN ACTION REGARDING A PATENT OR TRADEMARK**

In Compliance with 35 U.S.C. § 290 and/or 15 U.S.C. § 1116 you are hereby advised that a court action has been filed in the U.S. District Court for the District of Delaware on the following

☐ Trademarks or ☑ Patents. (☐ the patent action involves 35 U.S.C. § 292.):

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
	8/24/2022	for the District of Delaware
PLAINTIFF		DEFENDANT
10X GENOMICS, INC.		PARSE BIOSCIENCES, INC. and THE BOARD OF TRUSTEES OF LELAND STANFORD JUNIOR UNIVERSITY

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1	10,155,981	12/18/2018	10x Genomics, Inc.
2	10,697,013	6/30/2020	10x Genomics, Inc.
3	10,240,197	3/26/2019	10x Genomics, Inc.
4	10,150,995	12/11/2018	The Board of Trustees of the Leland Stanford Junior University
5	10,619,207	4/14/2020	The Board of Trustees of the Leland Stanford Junior University

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading		
	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
1			
2			
3			
4			
5			

In the above—entitled case, the following decision has been rendered or judgement issued:

DECISION/JUDGEMENT

CLERK	(BY) DEPUTY CLERK	DATE

Copy 1—Upon initiation of action, mail this copy to Director **Copy 3—Upon termination of action, mail this copy to Director**
Copy 2—Upon filing document adding patent(s), mail this copy to Director **Copy 4—Case file copy**

ADDENDUM TO AO 120 (ADDITIONAL PATENTS)

DOCKET NO.	DATE FILED	U.S. DISTRICT COURT
	8/24/2022	for the District of Delaware

PLAINTIFF	DEFENDANT
10X GENOMICS, INC.	PARSE BIOSCIENCES, INC. and THE BOARD OF TRUSTEES OF LELAND STANFORD JUNIOR UNIVERSITY

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
6	10,738,357	August 11, 2020	The Board of Trustees of the Leland Stanford Junior University
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

In the above—entitled case, the following patent(s)/ trademark(s) have been included:

DATE INCLUDED	INCLUDED BY ☐ Amendment ☐ Answer ☐ Cross Bill ☐ Other Pleading

	PATENT OR TRADEMARK NO.	DATE OF PATENT OR TRADEMARK	HOLDER OF PATENT OR TRADEMARK
21			
22			
23			
24			
25			