

IMPROVING CELL-FREE DNA QUALITY

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Application No. 62/748,767, 5 entitled "IMPROVING CELL-FREE DNA QUALITY," filed on October 22, 2018, the disclosure of which is incorporated by reference in its entirety for all purposes.

BACKGROUND

[0002] The detection of circulating cell-free fetal DNA (cfDNA) in maternal plasma has been increasingly adopted in noninvasive prenatal testing (NIPT) since its discovery in 1997.¹ cfDNA 10 molecules of different origins have characteristic size distributions, with the most abundant 166-bp peak for maternal cfDNA molecules and a relatively prominent 143-bp peak for the fetal cfDNA molecules.² Size differences between fetal and maternal cfDNA have been used for the development of size-based NIPT.²⁻⁴ Fetal DNA fraction is a major contributor for successful NIPT. Estimation of the fetal DNA fraction has been shown to be achievable, even with shallow 15 sequencing depths.⁵ Despite the international adoption of NIPT, a number of issues need further refinements, including false-negative and no-call results.⁶ To these ends, a number of approaches have been explored to enrich fetal DNA from maternal plasma.^{3,7,8} However, all of these methods have their disadvantages.

[0003] DNA damage, for example, single-strand nicks, exists and distributes throughout 20 cfDNA molecules.^{9,10} Apoptosis, which is believed to be one of the major mechanisms of cfDNA production, involves DNA fragmentation during karyorrhexis, followed by the formation and production of apoptotic bodies.¹¹⁻¹³ Commonly used sequencing library preparation methods employ the repair of DNA ends only, leading to the loss of damaged cfDNA fragments such as those with single-strand nicks.^{10,14,15} Therefore, the detection of these damaged cfDNA 25 molecules by the current paired-end massively parallel sequencing (MPS) platforms is challenging.

[0004] Systems and methods described below address DNA damage and other DNA issues. The quality of cell-free DNA for sequencing library preparation and/or other analysis may improve as a result of the systems and methods.

BRIEF SUMMARY

- [0005] Embodiments of the present invention provide for improving the quality of cell-free DNA for analysis. Cell-free DNA may include DNA with defects that do not allow for analysis of those DNA with techniques such as sequencing and targeted capture enrichment. These defects may be defects within the strands of the DNA and not present at the ends of the DNA. Embodiments of the present invention repair these intrastrand defects in cell-free DNA. The repair of the defects in cell-free DNA may then allow for these repaired cell-free DNA to be analyzed by techniques, including sequencing and targeted capture enrichment.
- [0006] Repairing cell-free DNA may increase the amount of DNA available for analysis and/or may allow for longer DNA fragments to be analyzed. Because more DNA without defects would be available after repair, analysis of DNA would be possible in some cases where the amount of DNA without defects was not sufficient for analysis or sufficient to allow for a statistically significant conclusion from the analysis. Additionally, longer DNA fragments may be more prone to defects repairable by techniques described herein. Repairing the defects may then allow for more longer DNA fragments to be analyzed. Long DNA fragments allow for analysis of disorders and conditions with sequence patterns are repeated over a minimum number of loci.
- [0007] The pattern or severity of defects in the cell-free DNA may be an indication of a disorder or a condition. A certain disorder, for example, may have more defects in cell-free DNA than defects in cell-free DNA from a subject without the disorder. The number or pattern of defects can then be used to provide a classification of whether an individual has a disorder.
- [0008] A better understanding of the nature and advantages of embodiments of the present invention may be gained with reference to the following detailed description and the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0009] FIGS. 1A, 1B, and 1C show different defect modes affecting DNA according to embodiments of the present invention.
- [0010] FIGS. 2A and 2B show repair changes in the size profile of cfDNA in maternal plasma according to embodiments of the present invention.

[0011] FIGS. 3A-3D show changes in size profile as a result of repair on a logarithmic scale according to embodiments of the present invention.

[0012] FIGS. 4A and 4B show treatment repairs cfDNA molecules of both fetal and maternal origins according to embodiments of the present invention.

5 [0013] FIG. 5 shows fetal DNA fraction characterizations of repaired cfDNA according to embodiments of the present invention. Fetal DNA fraction of sham or repaired cfDNA from first and third trimester maternal plasma

[0014] FIGS. 6A and 6B show size fractionated fetal DNA fractions of first and third trimester maternal plasma cfDNA after sham or repair treatment according to embodiments of the present
10 invention.

[0015] FIG. 7 shows an error rate comparison in between sham and repaired cfDNA from first and third trimester maternal plasma according to embodiments of the present invention.

[0016] FIG. 8 shows an enrichment of long cfDNA molecules after repair treatment in target captured libraries according to embodiments of the present invention.

15 [0017] FIG. 9A shows increased sequencing depth of fetal cfDNA molecules by chromosome Y targeted capture according to embodiments of the present invention.

[0018] FIG. 9B shows quantification of short and long cfDNA chromosome Y molecules from first and third trimester maternal plasma after sham or repair treatment according to embodiments of the present invention.

20 [0019] FIG. 10 shows that repair treatment enriches cfDNA molecules longer than 250 bp according to embodiments of the present invention.

[0020] FIG. 11 shows size profiles comparison of libraries prepared by different protocols from M12490 third trimester maternal plasma DNA according to embodiments of the present invention.

25 [0021] FIG. 12 shows size profiles of cfDNA from mice with and without repair according to embodiments of the present invention.

[0022] FIG. 13 shows a table of size distributions of short (0-250 bp) and long (>250 bp) cfDNA fragments from maternal plasma of pregnant women carrying a fetus with trisomy 21 according to embodiments of the present invention.

5 [0023] FIGS. 14A and 14B show averaged size profiles with and without repair from five women each carrying a trisomy 21 fetus according to embodiments of the present invention. FIG. 14A is a graph with a log scale, and FIG. 14B is a graph with a linear scale.

[0024] FIG. 15 shows the quantification of short (0-250 bp) and long (251-600 bp) cfDNA molecules from five women each carrying a trisomy 21 fetus after sham-repaired or PreCR-repaired treatment according to embodiments of the present invention.

10 [0025] FIG. 16 shows improved noninvasive prenatal testing (NIPT) performance of trisomy 21 after PreCR repair treatment according to embodiments of the present invention.

[0026] FIG. 17 shows a receiver operating characteristic curve for the improvement in determining trisomy after repair according to embodiments of the present invention.

15 [0027] FIG. 18 shows a table of the SLE patients from whom cfDNA was extracted according to embodiments of the present invention.

[0028] FIG. 19 shows the size profile on a log scale of DNA fragments with and without repair from four SLE patients according to embodiments of the present invention.

[0029] FIG. 20 shows the size profile on a linear scale of DNA fragments with and without repair from four SLE patients according to embodiments of the present invention.

20 [0030] FIGS. 21A and 21B show averaged size profiles comparison of sham-repaired and PreCR-repaired cfDNA from 4 SLE patients (2 inactive SLE; 2 active SLE) according to embodiments of the present invention. FIG. 21A is a graph with a log scale, and FIG. 21B is a graph with a linear scale.

25 [0031] FIG. 22 shows a table of the percentage changes and percentages in different size ranges of DNA fragments in active and inactive SLE patients after cfDNA repair according to embodiments of the present invention.

[0032] FIG. 23 shows a table of changes in different size ranges of DNA fragments in active and inactive SLE patients after cfDNA repair according to embodiments of the present invention.

[0033] FIG. 24A and FIG. 24B plot the molecular diagnostic potential of SLE patients using sham and repair DNA fragments according to embodiments of the present invention.

5 [0034] FIG. 25 shows the DNA packing scheme in normal cells according to embodiments of the present invention.

[0035] FIG. 26 shows the presence of CGG triplet repeat cfDNA fragments in patients with Fragile-X Syndrome (FXS) according to embodiments of the present invention.

10 [0036] FIG. 27 shows shows allele haplotyping of the HLA gene according to embodiments of the present invention.

[0037] FIG. 28 shows linkage analysis for Congenital adrenal hyperplasia (CAH), according to embodiments of the present invention.

[0038] FIG. 29 shows a method of preparing a sequencing library of a biological sample according to embodiments of the present invention.

15 [0039] FIG. 30 shows a method of determining a classification of whether an individual has a disorder according to embodiments of the present invention.

[0040] FIG. 31 illustrates a system according to embodiments of the present invention.

[0041] FIG. 32 shows a block diagram of an example computer system usable with system and methods according to embodiments of the present invention.

20

TERMS

[0042] As used herein, the term “locus” or its plural form “loci” is a location or address of any length of nucleotides (or base pairs) which has a variation across genomes. A “sequence read” refers to a string of nucleotides sequenced from any part or all of a nucleic acid molecule. For
25 example, a sequence read may be a short string of nucleotides (e.g., 20-150) sequenced from a nucleic acid fragment, a short string of nucleotides at one or both ends of a nucleic acid fragment, or the sequencing of the entire nucleic acid fragment that exists in the biological sample. A

sequence read may be obtained in a variety of ways, e.g., using sequencing techniques or using probes, e.g., in hybridization arrays or capture probes, or amplification techniques, such as the polymerase chain reaction (PCR) or linear amplification using a single primer or isothermal amplification.

- 5 [0043] A “tissue” corresponds to a group of cells that group together as a functional unit. More than one type of cells can be found in a single tissue. Different types of tissue may consist of different types of cells (e.g., hepatocytes, alveolar cells or blood cells), but also may correspond to tissue from different organisms (mother vs. fetus) or to healthy cells vs. cells showing a disorder.
- 10 [0044] A “*biological sample*” refers to any sample that is taken from a subject (e.g., a human, such as a pregnant woman, a person with a disorder, or a person suspected of having a disorder, an organ transplant recipient or a subject suspected of having a disease process involving an organ (e.g., the heart in myocardial infarction, or the brain in stroke, or the hematopoietic system in anemia) and contains one or more nucleic acid molecule(s) of interest. The biological sample
- 15 can be a bodily fluid, such as blood, plasma, serum, urine, vaginal fluid, fluid from a hydrocele (e.g. of the testis), vaginal flushing fluids, pleural fluid, ascitic fluid, cerebrospinal fluid, saliva, sweat, tears, sputum, bronchoalveolar lavage fluid, discharge fluid from the nipple, aspiration fluid from different parts of the body (e.g. thyroid, breast), etc. Stool samples can also be used. In various embodiments, the majority of DNA in a biological sample that has been enriched for
- 20 cell-free DNA (e.g., a plasma sample obtained via a centrifugation protocol) can be cell-free, e.g., greater than 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the DNA can be cell-free. The centrifugation protocol can include, for example, 3,000 g x 10 minutes, obtaining the fluid part, and re-centrifuging at for example, 30,000 g for another 10 minutes to remove residual cells. The cell-free DNA in a sample can be derived from cells of various tissues, and thus the sample
- 25 may include a mixture of cell-free DNA.
- [0045] “*Nucleic acid*” may refer to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term may encompass nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding
- 30 properties as the reference nucleic acid, and which are metabolized in a manner similar to the

reference nucleotides. Examples of such analogs may include, without limitation, phosphorothioates, phosphoramidites, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

[0046] Unless otherwise indicated, a particular nucleic acid sequence also implicitly
5 encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608
10 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

[0047] The term “*nucleotide*,” in addition to referring to the naturally occurring ribonucleotide or deoxyribonucleotide monomers, may be understood to refer to related structural variants thereof, including derivatives and analogs, that are functionally equivalent with respect to the
15 particular context in which the nucleotide is being used (*e.g.*, hybridization to a complementary base), unless the context clearly indicates otherwise.

[0048] A “*sequence read*” refers to a string of nucleotides sequenced from any part or all of a nucleic acid molecule. For example, a sequence read may be the entire nucleic acid fragment that exists in the biological sample. A sequence read may be obtained from a single-molecule
20 sequencing. “*Single-molecule sequencing*” refers to sequencing of a single template DNA molecule to obtain a sequence read without the need to interpret base sequence information from clonal copies of a template DNA molecule. The single-molecule sequencing may sequence the entire molecule or only part of the DNA molecule. A majority of the DNA molecule may be sequenced, *e.g.*, greater than 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

[0049] A “*classification*” refers to any number(s) or other characters(s) that are associated with a particular property of a sample. For example, a “+” symbol (or the word “positive”) could signify that a sample is classified as having deletions or amplifications. The classification can be binary (*e.g.*, positive or negative) or have more levels of classification (*e.g.*, a scale from 1 to 10 or 0 to 1). The term “*cutoff*” and “*threshold*” refer to a predetermined number used in an
30 operation. For example, a cutoff size can refer to a size above which fragments are excluded. A

threshold value may be a value above or below which a particular classification applies. Either of these terms can be used in either of these contexts.

[0050] The term “*size profile*” generally relates to the sizes of DNA fragments in a biological sample. A size profile may be a histogram that provides a distribution of an amount of DNA fragments at a variety of sizes. Various statistical parameters (also referred to as size parameters or just parameter) can be used to distinguish one size profile to another. One parameter is the percentage of DNA fragment of a particular size or range of sizes relative to all DNA fragments or relative to DNA fragments of another size or range.

[0051] The term “*cutoff value*” or amount as used in this disclosure means a numerical value or amount that is used to arbitrate between two or more states of classification for a biological sample — for example, the presence or absence of a genetic sequence that is associated or linked with a particular phenotypic condition or disease, or a susceptibility to a phenotypic condition or disease. For example, if a parameter is greater than the cutoff value, a first classification of the quantitative data is made, or if the parameter is less than the cutoff value, a different classification of the quantitative data is made.

DETAILED DESCRIPTION

[0052] Cell-free DNA in maternal plasma and in other biological samples can be used for diagnostic and other purposes. However, cell-free DNA molecules may be damaged within the strands of a double-stranded nucleic acid molecule. The damage may result from causes to the biological sample (e.g., heat, agitation, ultraviolet light) or internal to the biological sample (e.g., errors in replication). Conventional assays, including sequencing and using targeted probes, may be unable to determine reads from these double-stranded nucleic acid molecules with intrastrand defects. As a result, the damaged cell-free DNA may not provide any information that could be used for analysis. Embodiments of the present invention may allow for more information and more accurate information to be produced from cell-free DNA in a biological sample.

[0053] **FIG. 1A** shows intact double-stranded DNA and intrastrand defects. Intact dsDNA shows that there are no defects in the sugar phosphate backbone or in the base pairs. Nicked DNA shows that the backbone between two adjacent nucleotides is missing. Specifically, no phosphodiester bond is present between adjacent nucleotides. Gapped DNA is when one or more

nucleotides of a strand are missing. Abasic DNA is when one or more bases are missing, with the backbone unaffected.

[0054] FIG. 1B shows other defects that may affect the bases within the DNA. Thymidine dimers may form when adjacent thymine bases bond to each other through carbon-carbon double bonds, often in the presence of ultraviolet radiation. As a result, the thymine base pair bonding with the adenine is broken. Oxidized pyrimidines include when guanine is oxidized to form 8-oxoguanine. Deaminated cytosine includes when cytosine loses an amine group and forms uracil.

[0055] FIG. 1C shows a DNA-protein crosslinking defect. The double-stranded DNA crosslinks with the protein. DNA-protein crosslinking defects may be fixed through reverse crosslinking, which may be achieved by adding high concentrations of sodium chloride.

[0056] To salvage the damaged DNA molecules, we hypothesized that a repair mixture using Taq DNA polymerase, Bst DNA polymerase, Taq DNA ligase, endonuclease VIII, endonuclease IV, T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), or uracil-DNA glycosylase (UDG), could be used to repair cfDNA in maternal plasma. An example of a repair mixture is PreCR Repair Mix. Repaired maternal plasma DNA with higher DNA integrity might enhance the success for downstream analysis. Due to the fact that fetal-derived cfDNA molecules are shorter and more fragmented than the maternal ones,^{2,16} we proposed that fetal cfDNA molecules possess more DNA damages. By repairing these damaged fetal cfDNA molecules, the fetal DNA fraction may be enriched.

[0057] In this study, we applied repair treatment on cfDNA from maternal plasma of first- and third-trimester pregnancies. We studied and compared the size profiles and fetal DNA fractions between repaired cfDNA samples and their sham controls. We would also discuss the potential impact of repair treatment on the performance of NIPT and on the performance of assessing whether an individual has a disorder.

I. DNA REPAIR STUDY WITH MATERNAL PLASMA

[0058] DNA fragments from maternal plasma were treated to repair intrastrand defects such as those in FIG. 1A. The repaired DNA fragments were analyzed for their effect on short and long fragments and fetal and maternal DNA.

A. Materials and Methods

1. Subjects, sample collection and DNA extraction

[0059] This study was approved by the institutional research ethics committee. Pregnant women with singleton male fetuses were recruited from the Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong with informed consent. Plasma was isolated from 20 mL of EDTA-anticoagulated maternal peripheral blood as previously described.² For third trimester cases, 1 cm³ of placental tissue was dissected freshly after delivery from a region 2 cm deep and 5 cm away from the umbilical cord insertion. For first trimester cases, chorionic villus samples (CVS) were obtained. Plasma and buffy coat genomic DNA were extracted using the QIAamp DSP DNA Blood Mini Kit (Qiagen) and QIAamp DNA Blood Mini Kit (Qiagen), respectively, and quantified by Qubit 3.0 (Invitrogen). Placental DNA and chorionic villus DNA were extracted using the QIAamp DNA Mini Kit (Qiagen) according to manufacturer's protocols.

2. Plasma cfDNA repair and sequencing library preparation

[0060] 10 ng of extracted plasma DNA was subjected for DNA repair, using the PreCR Repair Mix (New England Biolabs). A sham control was done in parallel with the same amount of DNA input. The sham control includes the reaction buffer but without the active enzymes. In this case, the sham control does not include the active enzymes of the PreCR Repair Mix. A sham control is preferred to a normal control (i.e., no reaction buffers) by removing the potential bias caused from the reaction buffers. Briefly, 10 ng extracted DNA is mixed with 1X ThermoPol Buffer, 100 µM dNTPs, 1X NAD⁺, and H₂O in a 49 µl reaction, followed by addition of either 1 µl of H₂O (sham) or PreCR Repair Mix (repair). Sham- or repaired-treated plasma DNA were purified by a MinElute Reaction Cleanup Kit (Qiagen) to remove residual enzymes and reagents, followed by paired-end sequencing library preparations. Double-stranded DNA (dsDNA) libraries were generated as previously described.¹⁷ Adaptor-ligation single-stranded DNA (LIG-ssDNA) libraries were generated by the Accel-NGS 1S Plus DNA Library Kit (SWIFT Biosciences). Libraries were quantified by Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and real-time quantitative PCR (KAPA Library Quantification kit) on a LightCycler 96 System (Roche). The size profiles of the sequencing libraries were examined by Agilent 4200

TapeStation System with High Sensitivity D1000 ScreenTape (Agilent) to confirm successful library preparation before sequencing.

3. Targeted capture enrichment and massively parallel sequencing

[0061] Targeted capture of sequencing libraries was performed as described.¹⁸ The capture
5 probes (Roche Nimblegen) predominantly targeted sequences on Chr6 and ChrY. An average
sequencing depth of 152X per base (ranging from 82X to 210X) was obtained. Target capture
enrichment efficiency was validated by quantitative real-time PCR. All libraries were sequenced
with a paired-end (PE) format of 75 bp x 2 on a NextSeq 500 system (Illumina). Adaptor
sequences and low quality bases (i.e., quality score < 5) were removed and sequencing reads
10 were aligned to the non-repeat-masked human reference genome (hg19) using the Short
Oligonucleotide Alignment Program 2 (SOAP2).¹⁹ Up to two nucleotide mismatches, but not
indels, were allowed for each member of the pair-end reads.

4. Single Molecule Real-Time (SMRT) sequencing

[0062] The pre-target captured sequencing libraries with Illumina format adapters were pooled
15 and subjected to SMRT sequencing template construction using a SMRTbell Template Prep Kit
1.0 - SPv3 (Pacific Biosciences). The post-target captured Illumina sequencing libraries were
similarly pooled and further subjected to SMRT sequencing library construction. The amplicon
template preparation and sequencing protocol was used, with minor modifications: DNA was
purified with 1.8x AMPure PB beads, and library size was estimated using a TapeStation
20 instrument (Agilent). Sequencing primer annealing and polymerase binding conditions were
calculated with the SMRT Link v5.1.0 software (Pacific Biosciences). Briefly, sequencing
primer v3 was annealed to the sequencing template, then polymerase was bound to templates
using a Sequel Binding and Internal Control Kit 2.1 (Pacific Biosciences). Sequencing was
performed on a Sequel SMRT Cell 1M v2. Sequencing movies were collected on the Sequel
25 system for 10 hours with a Sequel sequencing kit 2.1 (Pacific Biosciences).

5. Microarray genotyping and single nucleotide polymorphism (SNP) identification

[0063] Fetal and maternal genomic DNA samples were genotyped with the Infinium Omni2.5-8 V1.3 Kit and the iScan system (Illumina). SNPs were called by the Birdseed v2 algorithm.²⁰

5 The genotypes of the CVS and placentas were compared with those of the mothers to identify the fetal-specific and maternal specific SNP alleles. A SNP was considered as fetal-specific if it was homozygous in the mother and heterozygous in the fetus, and the reverse for maternal-specific SNPs. The fetal DNA fractions were deduced as described.² In the size-fractionated fetal fraction analysis, fetal DNA fragments were divided into 10-bp bins as previously described,¹⁷ with
10 modified size range of analysis. Error rates were deduced by the following equation:

$$\text{Err \%} = \frac{\# \text{ non-iSCAN genotype allele}}{\# \text{ total SNP covering allele}}$$

[0064] For mouse plasma cfDNA repair, total blood samples from two wildtype (WT) C57BL/6 mice were obtained via cardiac puncture at 10-12 weeks of age. The blood was collected into EDTA-containing collection tubes and plasma was isolated as previously
15 described.²¹ 10ng of cfDNA input was subjected to sham- or repair treatment, before dsDNA sequencing library preparation.

B. Results

[0065] Results of repair treatment of cfDNA from maternal plasma show that repair treatment can recover a subset of long (> 250 bp) cfDNA molecules of both fetal and maternal origins. The
20 repair treatment gives a small but consistent fetal fraction increase. The results from analyzing repaired DNA show a relative higher enrichment of fetal long cfDNA molecules (> 250 bp) after repair when compared to their maternal counterparts. The repair treatment was not found to alter DNA fidelity for downstream assays of molecular diagnostics. Repair of defects of double-stranded DNA was found to be more effective than repair of defects on single-stranded DNA.

1. Size profile characteristics of sham or repaired cfDNA from first- and third-trimester maternal plasma samples

[0066] We first studied if there was any size profile change of cfDNA in maternal plasma after sham or PreCR DNA repair treatment. Grossly, the size profiles between sham and repair groups did not show major differences within the range of 0-250 bp from sequencing. Both groups have a major peak size at around 166 bp and a series of peaks with 10-bp periodicity from 70-150 bp. However, we observed there was an average of 2.4% decrease (96.64% to 94.33%) of short (0-250 bp) cfDNA fragments and an average of 68.8% increase (3.36% to 5.67%) of long (251-600 bp) cfDNA fragments after repair treatment compared with their sham counterparts. Statistical analyses showed that both changes were significant (both $p < 0.01$; Student's paired t -test). Student's paired t -tests and Wilcoxon Signed rank test were performed using IBM SPSS Statistics (IBM). A p -value of less than 0.05 was considered as statistically significant.

[0067] FIG. 2A shows the size profile comparison of sham and repaired cfDNA from a representative third trimester maternal plasma (M12778). Lower panels represent the magnified size profiles of long cfDNA molecules (black boxes). The figure shows an increase in the percentage fraction of fragments from 251-600 bp.

[0068] This trend was consistent in all subjects tested, including all maternal plasma samples from first and third trimesters (FIG. 2B). FIG. 2B shows quantification of cfDNA molecules with different sizes in first and third trimester maternal plasma after sham or repair treatment. There is a consistent decrease in the fraction of short (0-250 bp) and increase in the fraction of long (251-600 bp) cfDNA molecules after repair. Student's paired t -tests, $p < 0.01$. The decrease in the percentage fraction of short molecules is observed to be a result of an increase in the number of long molecules after repair.

[0069] FIGS. 3A, 3B, 3C, and 3D present data from samples in FIGS. 2A and 2B on a logarithmic scale to better illustrate the increases in frequency of longer fragments. FIG. 3A shows a size profile comparison of sham and repaired cfDNA in log scale representation from a representative third trimester maternal plasma (M12778). FIGS. 3B-3D show averaged size profile comparison of sham and repaired cfDNA. FIG. 3B shows results from all cases. FIG. 3C

shows the first trimester only. FIG. 3D shows the third trimester only. Enrichments in long (>250 bp) cfDNA molecules in repaired sequencing library is prominent in these figures.

[0070] We next asked if these size changes after repair treatment showed any relationship to the fetal or maternal origin of the cfDNA molecule. Using SNP-based genotypes, we separated the cfDNA fragments of all subjects into their respective maternal and fetal origins. cfDNA fragments carrying fetal or maternal-specific SNP alleles were separated and considered as fetal or maternal specific cfDNA molecules, respectively.¹⁷ We observed a similar trend in repaired cfDNA molecules from both fetal and maternal origins (**FIG. 4A**). FIG. 4A shows the averaged size profile comparison of sham and repaired cfDNA from fetal origin on the left side and from maternal origin on the right side. Lower panels represent the magnified size profiles of long cfDNA molecules (black box).

[0071] There was an average increase of 82.7% (2.26% to 4.13%) and 66.7% (3.42% to 5.7%) of long fetal and maternal cfDNA molecules after repair treatment, respectively (**FIG. 4B**). FIG. 4B shows quantification of short (0-250 bp) and long (251-600 bp) cfDNA molecules in fetal cfDNA on the left side and maternal cfDNA molecules on the right side. Data shown are from first and third trimester maternal plasma after sham or repair treatment. Statistical analyses showed that both increments were significant (both $p < 0.01$; Student's paired t -test). These results suggested that there was a relative higher enrichment of fetal long cfDNA molecules after repair, when compared to their maternal counterparts.

2. Fetal DNA fraction characteristics of sham or repaired cfDNA libraries from first- and third-trimester maternal plasma samples

[0072] As the fetal DNA fraction is an important parameter for successful NIPT, an increase in this metric may benefit the sensitivity of NIPT, especially in cases with insufficient fetal contribution.⁶ Therefore, we characterized and compared the overall fetal DNA fractions of maternal plasma from first and third trimester samples after sham and repair treatments. When pooling the data from all 8 samples together, we observed a consistent 4.31% increase of fetal DNA fraction (21.85% to 22.79%) after the repair treatment compared with their sham controls (**FIG. 5**). Repair treatment gives a small but consistent fetal fraction increase. Statistical analysis showed that such increment was significant ($p = 0.012$; Wilcoxon Signed rank test). The source of

such mild fetal DNA fraction increment may be from the newly repaired fetal long cfDNA molecules. To prove this, we applied size-fractionated fetal fraction analysis on both sham and repaired subjects. Fetal DNA fragments were divided into 10-bp bins, ranging from 0 bp to 600 bp. Fetal DNA fractions within each bin were then deduced separately.¹⁷ We observed that such fetal DNA fraction increment was mainly contributed by the increase of fetal cfDNA molecules with the size of around 300 bp (**FIGS. 6A and 6B**). FIG. 4A shows first trimester maternal plasma. FIG. 6B shows third trimester maternal plasma. The fetal DNA fractions in long (approx. 300 bp) cfDNA molecules have a mild increase in repaired groups. Taken together, these results suggested that a higher enrichment of long fetal cfDNA molecules over their maternal counterparts in maternal plasma.

[0073] DNA damage may be a pervasive cause of sequencing errors.²² Novel errors can be introduced into cfDNA through multiple steps of the cfDNA repair protocol, such as strand displacement and correction of oxidized nucleotides. Therefore, we next checked if there was significant change of error rate in repaired libraries. No significant changes in error rate were observed in both first- and third-trimester cases (**FIG. 7**). This suggests that repair treatment does not alter DNA fidelity for downstream assays of molecular diagnostics.

[0074] As fetal cfDNA molecules are the minority species in the maternal plasma, most of the data obtained from total maternal plasma cfDNA sequencing is dominated by maternally-derived cfDNA molecules. To obtain increased sequencing depth of fetal cfDNA molecules specifically, we used a target capture approach with probes hybridizing to sequences on Chr6 and ChrY.¹⁸ We achieved an average sequencing depth of 152X per base for all sham and repaired samples. The observation from the targeted sequencing results was concordant with the non-target captured data, with a 13.6% enrichment (10.53% to 11.96%) of long cfDNA molecules between 250 bp and 600 bp in size, after repair treatment (**FIG. 8**). Targeted capture probes were hybridizing on sequences of Chr6 and ChrY. Averaged size profile comparison of sham and repaired cfDNA from pooled first and third trimester maternal plasma. Lower panels are the magnified size profiles of long cfDNA molecules (black boxes).

[0075] When counting only the captured fetal cfDNA molecules, we observed an average of 14.8% enrichment (6.36% to 7.3%) of long chromosome Y cfDNA molecules after repair treatment over their sham counterparts (**FIG. 9A**). The top graphs show the averaged size profile

comparison of sham and repaired cfDNA molecules from chromosome Y in the first trimester (left side) and third trimester (right side) maternal plasma. Lower panels represent the magnified size profiles of long cfDNA molecules (black box). **FIG. 9B** shows quantification of short (0-250 bp) and long (251-600 bp) cfDNA chromosome Y cfDNA molecules from first and third trimester maternal plasma after sham or repair treatment. Statistical analyses showed that such increment was significant ($p < 0.01$; Student's paired t -test). Hence, we have demonstrated that repair treatment can recover a subset of long (> 250 bp) cfDNA molecules in maternal plasma, both of fetal and maternal origins. The small but consistent increase of total fetal DNA fraction after repair was contributed by a higher enrichment of long fetal cfDNA molecules over their maternal counterparts.

3. Enrichment of cfDNA molecules longer than 250 bp from maternal plasma after repair treatment

[0076] As the major size differences between sham and repair cfDNA resided on the cfDNA molecules longer than 250 bp, detection of such molecules by conventional NGS platforms, such as the Illumina platform, may not be preferred due to their short read lengths. A third generation sequencing platform from Pacific Biosciences, which employed the SMRT technology, can achieve sequencing read lengths exceeding 20 kb.²³ We hypothesized that using such a technology, we might improve our ability to observe the cfDNA molecules longer than 250 bp in the maternal plasma. Due to the limitation of minimum cfDNA input, we pooled all Illumina formatted pre- and post-target captured libraries and re-sequenced them on SMRT sequencing platform rather than directly sequencing DNA template. We expected an enrichment of long cfDNA molecules after repair treatment. Consistent with our sequencing data from the Illumina platform, we observed a respective 31.3% and 28.9% enrichment of cfDNA molecules longer than 250 bp in pooled libraries with or without target capture after repair treatment (**FIG. 10**). Pooled samples of pre- (left) and post- (right) target captured libraries showed an enrichment of cfDNA molecules longer than 250 bp after repair treatment, with more prominent enrichment for molecules longer than 400 bp. These enrichments were even more remarkable for cfDNA molecules longer than 600 bp (52% and 50.4% in pre- and post-target captured libraries, respectively). These data not only further confirmed our previous observations of long cfDNA molecule enrichment after repair treatment in paired-end, shorted-read sequencing platform, but

also shed light on a population of long cfDNA molecules that had not been analyzed by massively parallel sequencing in previous studies.

4. Performance of cfDNA repair in different library preparation protocols

5 [0077] Since we have shown previously that extra-short fragments were enriched by the single-stranded DNA (ssDNA) sequencing library preparation method,¹⁷ it would be worthy to evaluate the cfDNA repair performance in such protocol versus conventional double-stranded DNA (dsDNA) method. We assumed that some fractions of extra-short cfDNA fragments obtained by Adaptor-ligation ssDNA (LIG-ssDNA) protocol are actually released from damaged
10 DNA structures, where nicks and/or gaps are present in the single strands of the double stranded DNA structures. We speculated that cfDNA repair could diminish such extra-short cfDNA fragments (<100 bp) enriched by LIG-ssDNA protocol through filling gaps and nicks on DNA strands. To validate such assumption, we performed in parallel dsDNA and LIG-ssDNA library preparation methods following either sham or PreCR treatment. We observed a decrease of
15 extra-short cfDNA fragments (<100 bp) in repaired LIG-ssDNA library compared with its sham counterpart. For cfDNA fragments longer than 250 bp, increment of cfDNA fragments was very limited (**FIG. 11**). We concluded that although some portion of extra-short cfDNA fragments were diminished after repair, there was no evidence suggested such portion contributed to an increase of longer cfDNA fragments. Therefore, the ability of long cfDNA fragment recoveries
20 in LIG-ssDNA libraries was inferior than the dsDNA libraries. Repair of defects with PreCR is less effective on single-stranded DNA library preparation protocol than with double-stranded DNA library preparation protocol.

5. cfDNA repair performance in cfDNA from mouse plasma

[0078] To validate if the repair of cfDNA damages is also present in non-human mammals, we
25 evaluated the performance of cfDNA repair in cfDNA from a mouse. Plasma was isolated from two C57BL/6 wide type (WT) mice, DNA was extracted from plasma followed by either sham or repair treatment. Sequencing dsDNA libraries were prepared in the exact human scenario. We observed there was a 70% increase of long (200-1000 bp) cfDNA fragments after repair treatment (**FIG. 12**). There was a 1.41 times enrichment of long (>200 bp) cfDNA fragments in

repaired than the counterpart without repair treatment. cfDNA fragment frequencies are presented in log scale. This result is concordant with the human cohort, suggesting similar cfDNA damage patterns are shared by high order organisms, such as mammals.

II. DNA REPAIR STUDY WITH FETAL TRISOMY AND SLE PATIENTS

- 5 [0079] DNA fragments from additional subjects were treated to repair intrastrand defects. DNA fragments from maternal plasma of pregnant women, each carrying a singleton trisomy 21 fetus, were repaired. Additionally, DNA fragments from patients with active or inactive systemic lupus erythematosus (SLE) were repaired. The repaired DNA fragments were analyzed for their effect on short and long fragments and for their impact in detecting a condition of the sample.
- 10 The references in Section II are numbered separately from the references in the other sections, and the references are provided in a list at the end of the section.

A. Materials and Methods

- [0080] This study was approved by the institutional ethics committee. Ten first trimester and ten third trimester pregnant women with singleton male fetuses were recruited. Five pregnant
- 15 women with singleton trisomy 21 fetuses were also recruited. Four patients with systemic lupus erythematosus (SLE) were recruited. All SLE patients fulfilled the American College of Rheumatology diagnostic criteria and their lupus disease activities were assessed by the SLE disease activity index (SLEDAI), which is a clinical measurement of the disease activity.¹ Two patients had inactive SLE (SLEDAI median: 1; range 0-2) and the other two had active SLE
- 20 (SLEDAI both equal to 8). Anti-dsDNA antibody levels of all patients were also measured by the time of blood sampling. All pregnancy samples were recruited from the Department of Obstetrics and Gynaecology, Prince of Wales Hospital, Hong Kong. SLE samples were recruited from the rheumatology clinic at the Department of Medicine and Therapeutics, Prince of Wales Hospital, Hong Kong. All subjects had informed consent.

25 1. Library preparation

[0081] Plasma was isolated from 20 mL of EDTA-anticoagulated maternal peripheral blood as previously described.² For third trimester cases, 1 cm³ placental tissue was dissected freshly after delivery from a region 2 cm deep and 5 cm away from the umbilical cord. For first trimester cases, chorionic villus samples (CVS) were obtained. Plasma and buffy coat genomic DNA were

5 extracted using the QIAamp DSP DNA Blood Mini Kit (Qiagen) and QIAamp DNA Blood Mini Kit (Qiagen) respectively, and quantified by Qubit 3.0 (Invitrogen). Placental DNA and chorionic villus DNA were extracted using the QIAamp DNA Mini Kit (Qiagen) according to manufacturer's protocols. For cell-free DNA (cfDNA) repair, 10 ng of extracted plasma DNA input for each subject was repaired by using PreCR Repair Mix (New England Biolabs).

[0082] A sham control was done in parallel with the same amount of DNA input and reaction buffer only. All DNA repair or sham treatments were done according to manufacturer's instructions. Sham- or repaired-treated plasma DNA were purified by MinElute Reaction Cleanup Kit (Qiagen) to remove residual enzymes and reagents, followed by sequencing library preparation. Double-stranded DNA (dsDNA) libraries were generated by the KAPA HTP Library Preparation Kit for Illumina (KAPA Biosystems) according to manufacturer's instructions. Adaptor-ligation single stranded DNA (LIG-ssDNA) libraries were generated by the Accel-NGS 1S Plus DNA Library Kit (SWIFT Biosciences). Libraries were quantified by Qubit and real-time quantitative PCR on a LightCycler 96 System (Roche). The size profiles of the sequencing libraries were examined by the Agilent High Sensitivity D1000 ScreenTape System (Agilent). All libraries were sequenced with a paired-end (PE) format of 75 bp x 2 on a NextSeq 500 system (Illumina). Adaptor sequences and low quality bases (i.e., quality score, <5) were removed and libraries were aligned to the non-repeat-masked human reference genome (hg19) using the Short Oligonucleotide Alignment Program 2 (SOAP2).³ Up to two nucleotide mismatches, but not indels, were allowed for each member of the pair-end reads.

[0083] For mouse plasma cfDNA repair, total blood samples from two wildtype (WT) C57BL/6 mice were obtained via cardiac puncture at 10-12 weeks of age. The blood was collected into EDTA-containing collection tubes and plasma was isolated as previously described.⁴ 10ng of cfDNA input was subjected to sham- or repair treatment, before dsDNA sequencing library preparation. For SLE plasma samples, all extracted cfDNA of each sample were evenly separated into sham- or repair treatment before dsDNA sequencing library preparation.

2. Microarray genotyping and single nucleotide polymorphism (SNP) identification

[0084] Fetal and maternal genomic DNA samples were genotyped with the Infinium Omni2.5-8 V1.3 Kit and the iScan system (Illumina). SNPs were called by the Birdseed v2 algorithm with a confidence score cutoff of 0.15.⁵ The genotypes of the CVS and placentas were compared with those of the mothers to identify the fetal-specific and maternal-specific SNP alleles. A SNP was considered as fetal-specific if it was homozygous in the mother and heterozygous in the fetus, and the reverse for maternal-specific SNPs. The fetal DNA fractions were deduced as described.² Briefly, fetal DNA fraction (F) was deduced by the allelic ratio between a fetal specific SNP allele (p) and a common SNP allele (q) shared by the mother and the fetus using the following formula:⁶

$$F = \frac{2p}{p + q} \times 100\%$$

[0085] In the size-fractionated fetal fraction analysis, fetal DNA fragments were divided into 10-bp bins as previously described,⁷ with modified size range of analysis. For detection of trisomy 21, overrepresentation of Chr21 of each sample was quantified by the z-score using the following formula:

$$z - score = \frac{GRchr21_{test\ sample} - mean\ GRchr21_{euploid}}{SD_{euploid}}$$

where GRchr21 is the genomic representation (GR) of chromosome 21.

B. Results

[0086] Repairing DNA fragments was found to enrich longer fragments in both biological samples from pregnant women carrying singleton trisomy fetuses and in biological samples from active and inactive SLE patients. Repairing DNA fragments from maternal plasma was found to improve sensitivity and specificity for determining trisomy in a fetus. PreCR repair treatment enriched longer (>250 bp) cfDNA molecules without significantly losing shorter (<166 bp) ones. Repairing DNA fragments allows for more DNA fragments from a biological sample to be used for analysis of trisomy or SLE or other conditions than a biological sample without repair. Repairing DNA fragments gave a larger absolute separation difference between inactive and

active SLE samples. This translated into a relative 33.7% improvement for potential molecular diagnostics development.

1. Size profile characteristics of sham- or PreCR-repaired cfDNA from trisomy 21 maternal plasma samples

5 [0087] FIG. 13 shows a table of size distributions of short (0-250 bp) and long (>250 bp) cfDNA fragments from maternal plasma of pregnant women carrying a fetus with trisomy 21. Similar to the results from first and third trimester maternal plasma samples, we observed that there was an average of 3.3% absolute decrease (95.5%-92.2%) of short (0-250 bp) cfDNA
10 fragments and an average of 3.7% absolute increase (4.1%-7.8%) of long (>250 bp) cfDNA fragments in all trisomy 21 samples after PreCR repair treatment.

[0088] FIGS. 14A and 14B show the size profiles of DNA fragments in trisomy 21 samples before and after repair. FIG. 14A and FIG. 14B are graphs of the percentage fraction of the total frequency versus DNA fragment size. FIG. 14A shows the percentage fraction on a log scale,
15 and FIG. 14B shows the percentage fraction on a linear scale for the same data. Sham and PreCR-repaired treated fragments are plotted on each graph. The PreCR-repaired fragments show a larger fraction of longer fragments than the sham fragments.

[0089] FIG. 15 shows the quantification of short (0-250 bp) and long (251-600 bp) cfDNA molecules from five women each carrying a trisomy 21 fetus after sham-repaired or PreCR-repaired treatment. The graphs in FIG. 15 show the percentage fraction of the total frequency for
20 both sham and repair treatment fragments. The short fragments show a drop in the percentage fraction. The long fragments show an increase in the percentage fraction. Statistical analyses showed that both changes were significant (both $p < 0.01$; Student paired t test).

25 2. Improved NIPT performance of trisomy 21 after PreCR repair treatment

[0090] As the repaired cfDNA sample has more intact long cfDNA molecules that are analyzable from the fetus, we aimed to demonstrate its potential application for prenatal

diagnosis. We recruited five pregnancies with trisomy 21 fetuses and compared the chromosomal representation of Chr21 with the 10 first trimester euploid cases, which would be the most clinically relevant samples for NIPT. We observed that after PreCR repair treatment, Chr21 z-scores for the trisomy 21 cases showed an average increase in 17.3% (ranging from 14.4% to 19.8%) and exhibited better separation from the z-scores of the euploid samples (**FIG. 16**).

[0091] In **FIG. 16**, the chromosome 21 z-scores of five trisomy 21 samples are shown on the left, and the chromosome 21 z-scores of 10 euploid samples are shown on the right. The repaired trisomy 21 samples showed increased z-scores and greater separation from the euploid samples. These results suggested that PreCR repair treatment may improve NIPT performance of trisomy 21 pregnancies. The repair treatment may allow a lower fetal fraction to be used for aneuploidy analysis. The repair may also allow for a classification of an aneuploidy in situations where non-repaired DNA for an aneuploidy would not have shown a high enough z-score to classify as an aneuploidy. The repair treatment may result in a method with increased sensitivity, specificity, and/or accuracy over a method without a repair treatment.

15 a) Simulation details

[0092] In order to demonstrate the plasma DNA repair effect quantitatively, we performed a simulation analysis followed by receiver operating characteristic (ROC) analysis to determine whether using plasma DNA reads with repair provides an advantage over plasma DNA reads without repair using DNA reads mapped to chr21 in pregnant women carrying trisomy 21 fetuses. In the simulation, the fetal fraction was assumed to be 1% and the total sequenced reads were assumed to be 3 million. The relative increase of fetal fraction after PreCR repair was set to be 4.8%.

[0093] In the computer simulation, the number of sequenced reads of plasma DNA derived from a particular chromosome is assumed to follow the binomial distribution. For a pregnant woman carrying an euploid fetus, the proportion of sequenced reads of plasma DNA originating from chromosome 21 (chr21) is denoted as GR_{21} . Among a total sequenced reads (n), the number of reads derived from chr21 (E) would follow the distribution below:

$$E \sim \text{Binom}(n, GR_{21}), (1)$$

where 'Binom' represents the binomial distribution.

[0094] For a pregnant woman carrying a trisomy fetus, the proportion of sequenced reads of plasma DNA originating from chr21 is denoted as GR'_{21} :

$$GR'_{21} = \left(1 + \frac{f}{2}\right) \times GR_{21}, (2)$$

where f is the fetal DNA fraction in the maternal plasma DNA of a pregnant woman. Among
5 total sequenced reads (n), the number of reads derived from chr21 (T) would follow the distribution below:

$$T \sim \text{Binom}(n, GR'_{21}), (3)$$

which could be rewritten as:

$$T \sim \text{Binom}(n, \left(1 + \frac{f}{2}\right) \times GR_{21}), (4)$$

10 [0095] After the repair of plasma DNA prior to sequencing, f would increase to f' , which is governed by the below formula:

$$f' = (1 + \alpha) \times f, (5)$$

where α is the relative increase of the fetal DNA fraction after plasma DNA repair. In this
scenario, for a pregnant woman carrying a trisomy fetus, among total sequenced reads (n), the
15 number of reads derived from chr21 (G) would follow the distribution below:

$$G \sim \text{Binom}(n, \left(1 + \frac{(1 + \alpha) \times f}{2}\right) \times GR_{21}), (6)$$

[0096] According to (1), (4), and (6), we simulated sequenced reads originating from chr21 for
100 plasma DNA samples from pregnant women carrying euploid fetuses and those carrying
trisomy 21 fetuses, respectively. We also simulated 100 plasma DNA followed by DNA repair
20 from those carrying trisomy 21 fetuses. For each sample, the fetal fraction was assumed to be 1%
and the total sequenced reads (n) were assumed to be 3 million. Compared with pregnant women
carrying euploid fetuses, the receiver operating characteristic (ROC) analysis was used to
determine the diagnostic differences between the percentage of plasma DNA reads with and
without DNA repair mapped to chr21 in pregnant women carrying trisomy 21 fetuses.

25 [0097] The binomial distributions were generated with the R function *rbinom*. ROC curves for
the groups with and without DNA repair were plotted with R package pROC (version 1.15.3).

The DeLong's test was adopted to compare ROC curves to determine if the improvement of trisomy 21 detection using the method with DNA repair protocol is statistically significant compared with a method without DNA repair.

b) Simulation results

5 [0098] FIG. 17 shows the ROC curve for the results of the simulation comparing detecting trisomy 21 with DNA repair and without DNA repair. The results showed an improvement of area under the curve (AUC) from 0.729 to 0.822 after repair. Statistical analysis using DeLong's test showed the improvement was significant (sham vs. repair, DeLong's test: $p=0.006013$). Hence, repairing DNA can improve sensitivity and specificity of determining trisomy in a fetus
10 using DNA fragments from maternal plasma.

3. Increased longer cfDNA molecules in plasma from SLE patients after PreCR repair treatment

15 [0099] The cfDNA size profiles of cfDNA molecules from SLE patients is different compared to normal subjects.⁸ Prominent enrichment of cfDNA fragments shorter than 115 bp has been previously reported.⁸ Moreover, the severity of the disease corresponds to the amount of such extra-short cfDNA fragments, i.e., active SLE patients have more extra-short fragments in their plasma. Based on this, we aimed to explore the impact of PreCR repair treatment on the size
20 profiles of cfDNA molecules from SLE patients. We recruited 2 inactive (SLEDAI=0-2) and 2 active (SLEDAI=8) SLE patients, extracted cfDNA from their plasma and subjected to sham- or repair treatment.

[0100] FIG. 18 includes the sample number, SLE disease activity index (SLEDAI) level, anti-ds DNA antibody level of the SLE patients, and the total DNA amount from plasma. SLEDAI
25 and anti-ds DNA antibody level are clinical measurements of the disease activity. A higher SLEDAI and a higher anti-ds DNA antibody level correspond to a more active SLE status.

[0101] We observed an increase in longer cfDNA fragments after repair, similar to the results from the maternal plasma samples.

[0102] FIG. 19 and FIG. 20 show the size profiles of the cfDNA with and without repair from each of the SLE patients. FIG. 19 shows the log of the fraction versus the size of the DNA fragments. FIG. 19 shows the same data on a linear scale. Samples from four patients (S113, S117, S312, S494) were analyzed, with each patient having a sham and a repair sample. The sham data are shown with dashed lines, and the repaired data are shown with solid lines. Two patients had inactive SLE, and two patients had active SLE. The graphs show an increase in longer DNA fragments (>250 bp) with repair. Shorter DNA fragments (<200 bp) show a slight decrease with repair.

[0103] FIG. 21A and FIG. 21B show the averaged results from the inactive and active SLE patients displayed in FIG. 19 and FIG. 20. FIG. 21A shows the log fraction versus DNA fragment size. FIG. 21B shows the same data on a linear scale. The graphs show an increase in longer DNA fragments (>250 bp) with repair. Shorter DNA fragments (<200 bp) show a slight decrease with repair.

[0104] FIG. 22 shows a table of the percentage changes and percentages in different size ranges of DNA fragments in active and inactive SLE patients after cfDNA repair. FIG. 22 involves the same data displayed in FIG. 19 and FIG. 20. The table has four different size ranges: short (0-250 bp), long (>250 bp), 0-115 bp, and >200 bp. The size ranges of 0-115 bp and >200 bp are ranges that were used in Chan et al., PNAS (2014).⁸ The percentage of each size range is reported for inactive SLE patients and active SLE patients. The first two percentages in each cell of the table are the percentage of the size range before repair and after repair. The percentage in parentheses indicates the relative percentage increase or decrease after repair.

[0105] FIG. 23 shows the average percentage changes of inactive SLE and active SLE patients after repair. FIG. 23 shows the percentage changes after averaging data from the patients, similar to the data plotted in FIG. 21A and FIG. 21B. For inactive SLE patients, there was an average of 2.3% decrease (97.4%-95.2%) of short (0-250 bp) cfDNA fragments and an average of an 88.0% (2.6%-4.8%) increase of long (251-600) cfDNA fragments after PreCR repair treatment. For active SLE patients, there was an average of 1.4% decrease (97.2%-95.8%) of short cfDNA fragments and an average of 49.3% (2.8%-4.2%) increase of long cfDNA fragments after PreCR repair treatment. However, we did not observe any significant change in amount of cfDNA molecules with the size shorter or equal to 166 bp.

[0106] When revising the size ranges of short and long DNA fragments into 0-115 bp and >200 bp, respectively, the changes were similar to the 0-250 bp and >250 bp changes. For inactive SLE samples, there was an average of relative 14.5% decrease (4.3%-3.7%) and 46.8% (5.4%-7.6%) increase of short (0-115 bp) and long (>250 bp) cfDNA fragments after PreCR repair treatment, respectively. For active SLE patients, there was an average of relative 4.3% decrease (11.8%-11.3%) and 21.7% (7.8%-9.4%) increase of short (0-115 bp) and long (>250 bp) cfDNA fragments after PreCR repair treatment, respectively. The PreCR repair treatment increases the number of DNA fragments available for diagnosis of SLE. For evaluating the diagnostic potential of SLE patients using PreCR repair, we size-binned (limited to 120-130 bp) and separately plotted the absolute percentages of inactive and active SLE samples after either sham or PreCR repair treatment.

[0107] FIG. 24A and FIG. 24B plot the molecular diagnostic potential of SLE patients using sham and repair DNA fragments. FIG. 24A is a graphical representation of data for 120-130 bp in FIG. 22. FIG. 24A plots the absolute percentages of inactive and active SLE samples after sham or PreCR repair treatment. Size binning was set to 120-130 bp. D_1 and D_2 represent the absolute separation differences between inactive and active SLE samples after sham or PreCR repair treatment, respectively. FIG. 24B plots a bar chart showing a relative 33.7% increase in the absolute separation difference in repaired SLE samples. This better separation between inactive and active SLE samples after PreCR repair is an advantage in determining the level of SLE in a patient compared to non-repaired DNA fragments. Multiple size ranges were tested, with the 120-130 bp size range showing the best separation between inactive and active SLE samples after repair treatment. However, other size ranges may also be used, including 100-110 bp, 110-120 bp, 130-140 bp, or 140-150 bp.

4. References for Sections II, III.B, and III.C

1. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. *Arthritis Rheum.* 1992;35(6):630-640.

2. Lo YM, Chan KC, Sun H, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med*. 2010;2(61):61ra91.
3. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25(15):1966-1967.
- 5 4. Cheng THT, Lui KO, Peng XL, et al. DNase1 Does Not Appear to Play a Major Role in the Fragmentation of Plasma DNA in a Knockout Mouse Model. *Clin Chem*. 2018;64(2):406-408.
5. Hong H, Xu L, Tong W. Assessing consistency between versions of genotype-calling algorithm Birdseed for the Genome-Wide Human SNP Array 6.0 using HapMap samples.
10 *Adv Exp Med Biol*. 2010;680:355-360.
6. Jiang P, Peng X, Su X, et al. FetalQuant. *NPJ Genom Med*. 2016;1:16013.
7. Vong JSL, Tsang JCH, Jiang P, et al. Single-Stranded DNA Library Preparation Preferentially Enriches Short Maternal DNA in Maternal Plasma. *Clin Chem*. 2017;63(5):1031-1037.
- 15 8. Chan RW, Jiang P, Peng X, et al. Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing. *Proc Natl Acad Sci U S A*. 2014;111(49):E5302-5311.
9. Chen EZ, Chiu RW, Sun H, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One*. 2011;6(7):e21791.
- 20 10. Chiu RW, Akolekar R, Zheng YW, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ*. 2011;342:c7401.
11. Palomaki GE, Deciu C, Kloza EM, et al. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international
25 collaborative study. *Genet Med*. 2012;14(3):296-305.
12. Nicolaides KH, Syngelaki A, Ashoor G, Birdir C, Touzet G. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. *Am J Obstet Gynecol*. 2012;207(5):374.e371-376.
13. Pergament E, Cuckle H, Zimmermann B, et al. Single-nucleotide polymorphism-based
30 noninvasive prenatal screening in a high-risk and low-risk cohort. *Obstet Gynecol*. 2014;124(2 Pt 1):210-218.

14. Cervera R, Khamashta MA, Font J, et al. Morbidity and mortality in systemic lupus erythematosus during a 10-year period: a comparison of early and late manifestations in a cohort of 1,000 patients. *Medicine (Baltimore)*. 2003;82(5):299-308.
15. Raptis L, Menard HA. Quantitation and characterization of plasma DNA in normals and
5 patients with systemic lupus erythematosus. *J Clin Invest*. 1980;66(6):1391-1399.
16. Chen JA, Meister S, Urbonaviciute V, et al. Sensitive detection of plasma/serum DNA in patients with systemic lupus erythematosus. *Autoimmunity*. 2007;40(4):307-310.

III. DISCUSSION

- 10 [0108] Our findings demonstrated that cfDNA repair using a repair mixture (e.g., PreCR Repair Mix) could recover a subset of long cfDNA molecules in maternal plasma and from plasma from SLE patients. Relative enrichment of long cfDNA molecules after repair over short ones suggests that long cfDNA molecules contain more reparable DNA damages than their short counterparts. Given that long cfDNA molecules (e.g. >250 bp) existed in the maternal circulation
15 in trace amounts, any enrichment by repair might result in a significant increment in relative terms. Indeed, our data from the Illumina and Pacific Biosciences SMRT sequencing platforms illustrated enrichments of 68.8% and 31.3%, respectively, of cfDNA molecules longer than 250 bp. The enrichment became even more notable when one focused on even longer cfDNA molecules. For example, for cfDNA molecules between 450 bp and 600 bp in size, the
20 enrichment amounts were respectively 109.4% and 44.3%, as measured using the Illumina and the Pacific Biosciences SMRT platforms. For cfDNA molecules longer than 600 bp as measured by the Pacific Biosciences SMRT platform, the enrichment following DNA repair was 50.4%. In short, repair treatment could successfully recover long cfDNA molecules from maternal plasma. It appeared that the longer the cfDNA molecule was, the higher was the enrichment.
- 25 [0109] Repairing intrastrand defects in cell-free DNA was not previously performed for several reasons. The concentration of cell-free DNA in a biological sample is relatively low, and the concentration of cell-free DNA with defects would be even lower. Repairing the intrastrand defects was therefore thought to provide little benefit to analysis. Additionally, the fraction of clinically-relevant DNA may not increase significantly after repair. For example, FIG. 5 shows
30 that the fetal DNA fraction measured from sham and repair samples was often unchanged. Hence,

repairing defects in cell-free DNA may not enhance the signal of clinically-relevant DNA over non-clinically-relevant DNA. However, repairing cell-free DNA was found to have different effects on short and long cell-free DNA.

[0110] Different mechanisms may be at play for the generation of short and long cfDNA molecules in plasma. For example, cfDNA molecules shorter than 200 bp would likely be produced by enzymatic digestion during apoptosis. On the other hand, long (>250 bp) cfDNA molecules would likely be generated by other cell death mechanisms, such as necrosis.^{11,24-27} As discussed earlier, the enrichment of cfDNA molecules by repair was more prominent between 250 bp and 600 bp. The latter lengths are reminiscent of di- and tri-nucleosomal patterns (circa 330 bp and 500 bp, respectively).

[0111] FIG. 25 shows the DNA packing scheme in normal cells. DNA is wrapped around a nucleosome, which is composed of an octamer of 4 core histone proteins (forming a “nucleosome core” wrapped by 147 bp of DNA with a ~10 bp helical repeat), linker histones and linker DNA (mean size around 20 bp). It has been demonstrated that plasma DNA is not randomly fragmented. High resolution plasma DNA size profiling revealed a predominant peak at 166 bp and a 10-bp periodicity below 150 bp.² This size profile has been proposed to be closely related to the nucleosomal structure. The peak at 166 bp corresponds to the length of DNA wrapped around one nucleosome core. The peak between 250 and 450 bp (e.g., FIG. 8) may correspond to the length of DNA wrapped around two nucleosome cores. The peak between 450 bp and 600 bp may correspond to the length of DNA wrapped around three nucleosome cores. Repair of cell-free DNA may allow for analysis of more di-nucleosomal and tri-nucleosomal DNA lengths.

[0112] From our data, the enrichment of long, but not short, cfDNA molecules after repair suggested that most of the reparable DNA damages might originate from cell death mechanisms other than apoptosis (e.g. necrosis). In contrast, an absence of short cfDNA molecule enrichment after repair treatment suggested that such short cfDNA molecules might not contain reparable DNA damages.

[0113] Breaking down total cfDNA into fetal and maternal fractions, our data suggested that DNA damages are not randomly distributed throughout cfDNA molecules of different origins. Although our data revealed that long cfDNA enrichments were occurring in both fetal and

maternal cfDNA molecules, the magnitude of enrichment of long fetal cfDNA molecules (82.3%) was higher than that for cfDNA molecules of maternal origin (66.7%) (FIG. 2). In line with this, the small but consistent overall fetal fraction increment after repair suggested that fetal cfDNA molecules might bear more DNA damages (FIG. 5). Possible reasons for this observation could be due to higher accessibility of fetal cfDNA molecules by DNA cutting enzymes.²⁸ The relatively lower enrichment (14.7%) of long fetal cfDNA molecules following DNA repair in target captured libraries comparing to non-captured enrichment (82.3%) could be due to competition effect between DNA templates with different lengths during target probes hybridization.

10 [0114] The ability of the repair treatment to restore damaged DNA opens up many potential applications. In prenatal diagnosis, repair of cfDNA may improve the success rate of NIPT. In particular, scenarios with non-reportable NIPT results caused by insufficient fetal DNA fraction²⁹⁻³³ or poor cfDNA quality might be improved by DNA repair. Examples of such scenarios might include NIPT in very early pregnancies, samples obtained from pregnant women with high body mass indices, etc.³⁴⁻³⁷ Furthermore, the ability of cfDNA repair to reveal more long analyzable cfDNA molecules in plasma might open up the possibility of using NIPT for long genomic targets, e.g., sequences involved in triplet repeat disorders such as the Fragile-X Syndrome (FXS). In FXS, the length of the CGG tandem repeats in patients with fully mutated *FMR1* alleles can be longer than 600 bp.^{38,39} Repair of cell-free DNA for use in determining FXS is described below.

25 [0115] Application of cfDNA repair might also be extended to the development of linkage-based NIPT for monogenic diseases. With an increase in analyzable cfDNA length, for selected genomic regions, mutated alleles of single genes from parental carriers flanked by polymorphic markers could be detected in the maternal plasma. This approach might be feasible for genomic regions with high concentrations of SNP clusters occurring in tandem, such as with the human leukocyte antigen (HLA) region. Using repaired cell-free DNA with analysis of the HLA region is described below.

30 [0116] In conclusion, this study has revealed a preferential recovery of long cfDNA molecules in maternal plasma and in plasma from SLE patients after repair treatment. Small but consistent increment of overall fetal DNA fraction is contributed by higher fetal-derived cfDNA molecules

enrichment. We hope that the data presented here might catalyze further research to translate these observations into clinical enhancements in NIPT.

A. Advantages of cell-free DNA repair in molecular diagnostics

[0117] DNA damage is introduced to cells during multiple cellular processes, such as metabolic activities and DNA replication during cell proliferation. In most occasions, cells are able to identify types of DNA damages and tackle them by different types of DNA damage repair processes. For irreparable DNA damages, cells are fated to undergo apoptosis to prevent from accumulation of carcinogenic mutations. It is believed that apoptosis and necrosis are two of the major sources of cell-free DNA (cfDNA) present in the circulation.^{10,27,40} Similar to the genomic DNA, there are multiple types of cfDNA damages, namely DNA physical breakages such as DNA nicks, gaps and double strand breaks formation, modifications of DNA molecules such as oxidation, hydrolysis such as deamination, loss of DNA bases and pyrimidine-dimer formation. Such DNA damages may decrease cfDNA integrity and may cause suboptimal effects to downstream analyses in molecular diagnostics, such as Next-Generation Sequencing (NGS), quantitative PCR (qPCR), size profile characterization, etc.

[0118] Most of cfDNA fragments have a peak size of 166 bp, resembling a 146-bp mononucleosomal structure plus 20-bp linker regions. Beside these major cfDNA species, there are minor cfDNA species with longer length, ranging from 166 bp to over 600 bp. Such rare cfDNA species may carry valuable information as important as those shorter ones. However, due the limitation of size range detection of current Next-Generation Sequencing (NGS) platforms, the detection of such long cfDNA species requires extra high throughput to achieve reasonable sequencing coverages, which is impractical and cost ineffective. Despite the technical obstacles, the read length limitation has somehow been partially compromised by recent third-generation Single Molecule Real-Time (SMRT) sequencing, in which super high read length (>20kb) can be produced from a single circulated cfDNA molecule. But still, the exploration of extra-long cfDNA species in the circulation remains to be a difficult task.

[0119] In contrast, cfDNA repair is a relatively cost-effective, quick and user-friendly method to achieve long cfDNA enrichment without altering sequencing throughput and sequencing library structure. Currently, one DNA repair kit is PreCR Repair Mix by New England Biolabs.

PreCR Repair Mix is capable to repair a variety of naturally-occurred DNA damages, such as abasic hydrolysis, nicks, basic oxidation and thymidine dimerization.⁴¹ However, it cannot repair DNA fragmentation and DNA-protein crosslinks.

B. Cell-free DNA repair for detecting fetal trisomy

5 [0120] The ability of the PreCR repair mix to restore damaged DNA opens up many potential applications. In prenatal diagnosis, our data (e.g., FIG. 16) demonstrated that PreCR repair treatment may contribute to improved NIPT performance for trisomy 21 screening. Z-scores of repaired trisomy 21 samples were better separated from those of the repaired euploid samples. This is especially important for a case where the Chr21 z-score is borderline and close to the
10 euploid - aneuploid cutoff without repair treatment. Combinational simulation and ROC analyses further confirmed such NIPT improvement could significantly increase the sensitivity as well as the specificity of the test. Beyond enhancing the statistical confidence of discrimination of euploidy and aneuploidy, repair of cfDNA may improve the call rate of NIPT. In particular, scenarios with non-reportable NIPT results caused by insufficient fetal DNA
15 fraction⁹⁻¹³ or poor cfDNA quality might be improved by DNA repair.

C. Cell-free DNA repair to assess SLE

[0121] SLE is a systematic autoimmune disease which can cause chronic inflammations at multiple sites, resulting to tissue damage and organ failure. Examples of such are renal complications, infections, and myocardial infarctions.¹⁴ SLE has also been reported to be
20 associated with the elevations of cfDNA in plasma.^{15,16} It has also been reported that SLE patients have a higher proportion of cfDNA with size shorter than 115 bp.⁸ The amount of such cfDNA molecules in the plasma of active SLE patients was threefold higher than that of healthy individuals.⁸ The elevation of these shorter (<115 bp) cfDNA molecules was more pronounced in active SLE patients compared to that in inactive SLE patients.⁸ Therefore, the ability of cfDNA
25 quality improvement by PreCR Repair Mix might cause an impact in size profiles of these cfDNA molecules in SLE patients. The observation that PreCR repair treatment enriched longer (>250 bp) cfDNA molecules without significantly losing shorter (<166 bp) ones suggests that most reparable cfDNA molecules are longer than 250 bp. The enriched extra short cfDNA (<115 bp), which is a feature of severe SLE patients,⁸ are generally irreparable by PreCR Repair Mix.

PreCR treatment repairs DNA fragments, allowing for more fragments to be analyzed for SLE or other conditions. Further investigations needed to reveal the nature of DNA damages carried by such cfDNA molecules. For diagnostic purposes, a better separation between inactive and active SLE samples after PreCR repair treatment improves SLE assessment.

5 D. Potential applications of cfDNA repair

[0122] DNA repair can be applied before a wide range of DNA assays and molecular tests, such as DNA fingerprinting in forensics, DNA archaeological samples retrieval for taxonomy and species classification, DNA recovery from old, archival FFPE samples, etc. In the cfDNA field, repair-pretreated cfDNA may increase NIPT accuracy and sensitivity. In particular,
10 situations with uninformative NIPT test results contributed by low cfDNA input or poor cfDNA quality could be improved by increasing intact, assayable fetal DNA in the maternal plasma. Examples of such conditions would be NIPT in first trimester pregnancies and obese pregnancies, where the fetal cfDNA are sparse and difficult to be assayed. Furthermore, by applying cfDNA repair into present haplotype-based NIPT protocols using Single Nucleotide Polymorphism
15 (SNP),^{18,42,43} detection and identification of special pregnancy conditions such as monogenic diseases, vanishing twin and triploid pregnancy would be more efficient and accurate.

E. Extended application of cfDNA repair beyond present detection limits

[0123] The ability of cfDNA repair to visualize long cfDNA fragments in the circulation opens up a potential extended application of molecular diagnostics on various diseases. In one example,
20 detection of long tandem repeats and repeated microsatellites species in the circulation of patients carrying suspected triplet repeat disorders was difficult and tedious, as those species are extremely sparse and heavily damaged. This results in difficulty of molecular diagnostics of such diseases by NIPT. However, with the advantages of cfDNA repair, the detection of such long tandem repeats and microsatellite species could be plausible.

25 [0124] **FIG. 26** shows the presence of CGG triplet repeat cfDNA fragments in patients with Fragile-X Syndrome (FXS), where the length of the fully mutated *FMRI* alleles can reach up to 600 bp or more (Garber, Kathryn B., Jeannie Visootsak, and Stephen Warren T., "Fragile X Syndrome," *Nature.com*, Nature Publishing Group, 9 Apr. 2008. Web. 09 Sept. 2016.). The CGG repeat number may exceed 200 in fully mutated (pathogenic) *FMRI* allele. Traditionally,

prenatal diagnosis of FXS remains invasive; expanded CGG tandem repeats of *FMRI* is detected by Triplet repeat Primed-PCR (TP-PCR) coupled with confirmation by Southern blot hybridization.^{44,45}

[0125] Diagnosis of FXS using NIPT is restricted by the cfDNA length detection limit. Indeed, NIPT diagnosis of FXS requires considerable integrity of the expanded *FMRI* alleles to be sequenced, the longer the CGG tandem repeat is from a fully mutated *FMRI* allele, the heavier it is damaged. A screening approach focusing on the relative methylation status of *FMRI* promoter region⁴⁶ allows NIPT of FXS. Researchers of this work discovered a distinct DNA-methylation boundary at locus upstream of *FMRI*, which protects the spreading of DNA methylation from the upstream terminus to the gene promoter and gene body of *FMRI*. They also discovered that this boundary is disrupted and methylation is penetrated into the *FMRI* promoter, which caused *FMRI* silencing. With cfDNA repair, enriched cfDNA fragments from hyper- or hypomethylated *FMRI* promoter with its proximal upstream terminus would accurately detect the DNA-methylation boundary and increase NIPT successful rate of FXS.

[0126] In another example, cfDNA repair is also applicable into assays where sufficient amount of cfDNA fragments with considerable length is necessary. One example would be allele haplotyping of *HLA* and *HBB* genes, both which are highly polymorphic with concentration of tandem SNP clusters. The relative short inter-SNP distances in the *HLA* and *HBB* loci allow potential haplotyping possibility, using cfDNA with linked and proximal tandem SNPs present in the circulation. *HLA* typing is a critical step before matching donors and recipients in cord blood and bone marrow transplants. Mismatched *HLA* pairing may result in post-transplant complications called graft-versus-host (GVHD) disease. Importantly, the successful detection of tandem SNPs within a target gene locus depends on existence of intact, assayable cfDNA fragments.

[0127] FIG. 27 shows allele haplotyping of the *HLA* gene. The *HLA* region covers the *HLA* region of the short arm of chromosome 6. Haplotyping the *HLA* gene uses a long length of DNA in order to identify the specific *HLA* genome sequence.⁴⁷ Longer DNA fragments covering multiple *HLA* genes aids accurate haplotyping with correct linkage.

[0128] The feasibility of such haplotyping method can also be further applied to haplotyping of other monogenic diseases, such as Congenital adrenal hyperplasia (CAH). FIG. 28 shows

linkage analysis for CAH. CAH is an autosomal recessive disorder arising from mutations of the *CYP21A2* gene. Linkage analysis can be achieved by locating and connecting 2 proximal SNPs flanking the mutated site.¹⁸ This linkage analysis is facilitated by using longer lengths of DNA. The distance between the mutation site and a nearby SNP varies, depending on the particular
5 SNP density of the genomic region of the mutation. The closer the distance between the mutation site and nearby SNP likely will result in a better outcome, as cfDNA is highly fragmented in the maternal circulation. In the case of linkage analysis, a long cfDNA covering the mutation site and both flanking SNPs provides linkage haplotype information.

[0129] In another example, cfDNA repair can facilitate the real-time monitoring of various
10 disease conditions in the course of treatments. Importantly, tissue mapping by differential cfDNA methylation signatures measurements provides a powerful tool for the detection of cfDNA origins.^{48,49} In pregnancy-related conditions such as pre-eclampsia and trophoblastic dysplasia, the real-time detection of placental-derived cfDNA in maternal plasma becomes a powerful monitoring tool. Any relapse of a disease or disorder might be immediately detected
15 and tackled for optimal treatment. For post-transplantation recipients, similar real-time tissue mapping can be applied for more effective monitoring for possible organ rejections. However, in the course of DNA bisulfite modification of sequencing library, a critical step of the bisulfite sequencing protocol, generates extrinsic DNA damages on sequencing library. Such damages might decrease the sensitivity of tissue mapping by converting sequencible fragments into
20 uninformative species. A possible extra repair treatment after bisulfite treatment manages to rescue those damaged bisulfite-converted fragments into intact, sequencible species.

[0130] In another example, detection of reparability of cfDNA may assist with the detection, assessment, and monitoring of disease progression of conditions associated with cfDNA damage, for example autoimmune disorders and inflammatory conditions. Particularly, the amount of
25 cfDNA damage can be deduced by differences between repaired samples and their non-repair controls; i.e. a higher sham-repair difference may reflect a higher content of cfDNA damage. In autoimmune and inflammatory conditions, such as systemic lupus erythematosus (SLE), the proportion of extra-short cfDNA molecules (≤ 115 bp) is much higher than that of healthy individuals. Progression of the disease (from inactive to active form) is correlated with the

increase of such extra-short cfDNA molecules.⁵⁰ We suspect that the extent and the nature of DNA damage in these diseases, especially in the severe or active form would be different also.

[0131] The detection of the extent of cfDNA damage, or conversely, the extent of reparability would allow the detection and assessment of such diseases. The higher the extent of DNA damage or the higher degree of reparability would reflect the existence of such diseases or high activity or higher severity of the disease. The measurement of the extent of reparability could be based on comparing the amount of various sized DNA molecules with or without DNA repair. In addition, assessment of the extent of DNA damage could be based on comparing the extent of DNA repair (the difference between the addition and the omission of the repair step) between the test sample and a control sample collection from a person without the disease or with a mild form of the disease. Such parameters can be generalized for the characterization of cfDNA damage. Similarly, the quality of any cfDNA / gDNA / ctDNA sample can also be assessed by the measurement of DNA reparability, prior to any downstream analysis or assay. For example, samples with high DNA reparability indicate poor DNA quality, and an additional repair treatment step may be recommended prior to downstream protocols, such as sequencing library preparation.

[0132] DNA damages can be also acquired and accumulated in a time-dependent manner. The longer a DNA fragment is released and stay in the circulation, the more chance it acquires and accumulates DNA damages. Therefore, real-time monitoring of the degree of DNA damages could be a potential approach of therapeutic efficiency.

F. cfDNA repair as a valuable tool for molecular diagnostics

[0133] As different cell types in the micro-environment of a certain tissue have different metabolic turnover rates, different tissues may have their unique “damage signatures.” Different cfDNA may be released from different damaged tissues. By tissue mapping (Sun et al., PNAS 2015), the origins of cfDNA can be traced and identified. The tissue mapping may be done in parallel with the quantification of cfDNA damage in circulation (e.g., using labelled dNTPs). As a result, a dynamic “damage signature” from a certain organ in a period of time may then be deduced and detected.

[0134] Another major issue of cfDNA repair development in the future would be the variety of repairable DNA. Currently, there are a number of DNA damages that cannot be reversed, namely DNA fragmentation and DNA-protein crosslinking. This is because the repair of DNA fragmentation, which is a type of double-strand break (DSB), requires a series of repair complex proteins to undergo non-homologous end joining (NHEJ) in the biological condition. The current commercial repair reaction, however, does not contain such repair complex. Therefore, repair of DNA fragmentation remains a current challenge. In resolving this, it could be plausible of spiking-in NHEJ proteins needed for DSB repair, such as Mre11-Rad50-Nbs1 (MRN) complex and Ku-DNA-PKcs complex. The DNA phasing issues after DSB repair could be resolved by bioinformatic means through a molecular barcoding system designed in between conventional DNA repair and DSB repair steps.

[0135] In contrast to DNA fragmentation, DNA crosslinkages such as inter-strand crosslinking and DNA-protein crosslinking are much more difficult to repair. The strong adherent force of covalent bonds in between DNA strand and proteins makes it very difficult to break. Due to the nature that a vast diversity of DNA binding proteins is present in the vicinity of circulation, it is possible that a portion of them can form DNA-protein crosslinks very easily. If this happens to the major plasma proteins, such as human serum albumin (HSA), that would result a vast amount of cfDNA being crosslinked to HSA which are hardly accessible without the procedure of reverse crosslinking. Therefore, it would be an intrinsic need to break the cfDNA-protein crosslinks and release those cfDNA molecules free to be downstream assayed. To encounter this, reverse crosslink steps should be designed and included in the DNA repair procedures.

[0136] Beside DNA damage type characterization, damage location on the cfDNA molecule might also be a piece of valuable information in future molecular diagnostics. It is important that the DNA damages patterns are traceable; knowing the exact positions of DNA damages on cfDNA molecules would be useful to deduce the source of such damages. However, there are a number of challenges to be overcome in the development of such "repair-based liquid biopsy" method. Firstly, without the known nicking patterns and cutting overhangs of multiple nucleases, it is impossible to deduce damage-initiating machineries retrospectively by surveying the entire pool of cfDNA damage types present in the circulation. For example, typical nucleases responsible for cfDNA fragmentation, such as DNaseI and DNaseI-like-3, do not contain

consensus cutting sites. These enzymes tend to cut DNA molecules in a randomized manner. Conversely, these enzymes have their unique attacking patterns on DNA molecules; DNase1 tends to attack and cut naked DNA molecules where DNase1-like-3 tends to generate multi-nucleosomal patterns.⁵¹

5 [0137] Secondly, the buildup of an atlas of all damage-containing cfDNA fragments with their exact damage locations will be necessary for molecular diagnostics development. With the advantage of cfDNA repair to recover long cfDNA fragments, it is possible to detect the previously undiscovered cfDNA species, with large gaps in the middle of one strand and extra-long overhangs on the other strand. Likewise, cfDNA molecules with combinations of multiple
10 damage types can also be systematically revealed, provided the damages are repairable.

[0138] Thirdly, tracing repaired cfDNA molecules in the mist of out-numbered intact cfDNA molecules could be resolved by spiking in labeled and non-toxic dNTPs. Labeled dNTPs may be used to repair the defects. The degree of damage could be calculated by the ratio of unlabeled and labeled dNTPs in a single cfDNA molecule. This is particularly useful in the continuous
15 monitoring of disease progress and therapeutic effectiveness by the measurement of the repair dynamics. In conclusion, the ultimate goal for cfDNA in molecular diagnostics is, almost surely, the “hunt” of novel class of undiscovered, damaged cfDNA molecules in the circulation will be possible with the improving of DNA repair technologies.

IV. EXAMPLE METHODS

20 A. Improving cell-free nucleic acid quality

[0139] FIG. 29 shows a method 2900 of improving analysis of a first biological sample. The first biological sample may include cell-free nucleic acid molecules. The first biological sample may include maternal plasma from a female subject pregnant with a fetus or plasma from a subject having a disorder or suspected to have a disorder. Plasma may be extracted from a blood
25 sample. The first biological sample may be a sample before any component not derived from the subject or the fetus of the subject is added.

[0140] At block 2902, the double-stranded nucleic acid molecules may be obtained from the cell-free nucleic acid molecules to produce a second biological sample. One or more double-stranded nucleic acid molecules of the plurality of double-stranded nucleic acid molecules each

has one or more defects. For each of the one or more double-stranded nucleic acid molecules having a defect of the one or more defects, the defect is present in the respective double-stranded nucleic acid molecules at a location at least one nucleotide away from the closest end of the respective double-stranded nucleic acid molecule. In other words, the defect may be an
5 intrastrand defect and not a defect at either end of either strand of the double-stranded nucleic acid molecule. In some embodiments, the defect may be at a location 2, 3, 4, 5, 10, 15, 20, 25, or 30 or more nucleotides away from the closest end of the respective double-stranded nucleic acid molecule.

[0141] The one or more defects may include a defect or defects selected from the group
10 consisting of a nick, gap, abasic site, thymidine dimer, oxidized pyrimidine, deaminated cytosine, blocked 3' end defect, or a combination thereof.

[0142] The one or more double-stranded nucleic acid molecules may have lengths in a range from 251 to 600 bp, from 251 to 450 bp, or from 451 bp to 600 bp. The lengths may be associated with di-nucleosomal or tri-nucleosomal nucleic acid molecules.

15 [0143] At block 2904, a mixture including an enzyme may be added to the second biological sample. The enzyme may include at least one of a polymerase, a ligase, an endonuclease, or a glycosylase. The mixture may include at least one enzyme of *Taq* DNA polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV, T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), or uracil-DNA glycosylase (UDG). In some embodiments, the
20 mixture may include *Taq* DNA polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV, T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), and uracil-DNA glycosylase (UDG). The mixture may also include any combination of NAD⁺, ThermoPol Reaction Buffer, and deoxynucleoside triphosphates (dNTPs).

[0144] At block 2906, one or more defects in each of the one or more double-stranded nucleic
25 acid molecules may be repaired using the enzyme to produce a repaired set of double-stranded nucleic acid molecules. The repaired set of double-stranded nucleic acid molecules may be free of the one or more defects, including any defect described herein. Repairing the one or more defects may include repairing a greater number of nucleic acid molecules having lengths from 251 to 600 bp than nucleic acid molecules having lengths from 0 to 250 bp. In other words, more

defects may be repaired from dinucleosomal and/or trinucleosomal nucleic acid molecules than from mononucleosomal nucleic acid molecules.

[0145] For a biological sample containing clinically-relevant DNA (such as fetal DNA), repairing defects in the one or more double-stranded nucleic acid molecules may raise the fraction of the clinically-relevant DNA to allow for further analysis. For example, if the first biological sample is obtained from a female subject pregnant with a fetus, the one or more double-stranded nucleic acid molecules with defects may include a plurality of fetal-derived nucleic acid molecules. After repairing the one or more defects, the second biological sample may be characterized by a fetal fraction calculated from reads obtained from analyzing the repaired set of double-stranded nucleic acid molecules. The fetal fraction may be greater than a threshold fraction preferred or needed for further analysis. For example, the fetal fraction after repair, may be greater than or equal to 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, or 0.10. The fetal fraction after repair may be calculated from reads to be greater than a fetal fraction calculated from reads before repair. The actual fraction of DNA from the fetus may not change, however. Instead, the calculated fetal fraction may increase after repair because reads of fetal-derived DNA may be obtained only after repair.

[0146] Methods may include performing blunt-end ligation of the plurality of double-stranded nucleic acid molecules or of the repaired set of double-stranded nucleic acid molecules. Blunt-end ligation is not the repair of the one or more defects in block 2906 because blunt-end ligation addresses nucleotides at the end of the double-stranded nucleic acid molecule, not intrastrand defects.

[0147] At block 2908, a sequencing library may be produced using the repaired set of double-stranded nucleic acid molecules. The repaired set of double-stranded nucleic acid molecules is subjected to sequencing library preparation procedures, including end repairing, A-tailing, sequencing adaptor ligation, and indexed PCR amplification.

[0148] At block 2912, the sequencing library may be analyzed to determine a set of reads. In some embodiments, at least 1,000 cell-free DNA molecules are analyzed. In other embodiments, at least 10,000 or 50,000 or 100,000 or 500,000 or 1,000,000 or 5,000,000 cell-free DNA molecules or more can be analyzed. The set of reads may be determined by sequencing the sequencing library. Sequencing techniques may include paired-end sequencing, massively

parallel sequencing, single molecule real-time (SMRT) sequencing, PCR, reverse transcription PCR (RT-PCR), digital PCR, or sequencing. In some embodiments, the set of reads may be determined by performing targeted capture enrichment. The set of reads may not include reads from double-stranded nucleic acid molecules having one or more defects. For example, the set of reads may not include reads from either single-stranded nucleic acid molecules from the double-stranded nucleic acid molecules having one or more defects. In some embodiments, the set of reads may be from only one strand of the double-stranded nucleic acid molecules. In other embodiments, the set of reads may be from both strands of the double-stranded nucleic acid molecules.

10 [0149] Methods may include using the sequencing library prepared using the repaired set of double-stranded nucleic acid molecules. Methods may include detecting an aneuploidy in a fetus from a maternal plasma sample. In some embodiments, methods may include detecting a sequence imbalance. Methods may include detecting a microamplification or microdeletion in a fetus. In addition, methods may include detecting any disorder described herein.

15 [0150] For example, a method may include determining whether a sequence imbalance exists in the second biological sample. A plurality of sequence reads from the repaired set of double-stranded nucleic acid molecules may be obtained. The plurality of sequence reads may be analyzed by a computer system. Analyzing a sequence read may include identifying a location of the sequence read in a reference genome by aligning the sequence read to the reference genome.

20 The amount of sequence reads from a genomic region using the identified locations may be determined by a computer system. The value of a normalized parameter for the amount of sequence reads from the genomic region may be obtained. The value of the normalized parameter may be compared to a cutoff value. Whether the sequence imbalance exists may be determined based on the comparison. The genomic region may be a chromosome, and the
25 sequence imbalance may be an aneuploidy.

[0151] The repaired set of double-stranded nucleic acid molecules may be used for methods where having longer nucleic acid molecules is preferred. For example, the repaired set of nucleic acid molecules may be used to determine a classification whether Fragile-X Syndrome (FXS) exists using a first biological sample from a female subject pregnant with a fetus. The method
30 may include enriching the repaired set of double-stranded nucleic acid molecules from the FMR1

promoter region and the proximal upstream terminus of the FMR1 promoter to form an enriched set of double-stranded nucleic acid molecules. A methylation-aware assay may be performed on the enriched set of double-stranded nucleic acid molecules. In some embodiments, the methylation-aware assay may include bisulfite sequencing. A plurality of sequence reads from the enriched set of double-stranded nucleic acid molecules may be obtained. The plurality of sequence reads may be aligned to a reference genome. The presence or absence of a DNA-methylation boundary indicative of FXS at a location upstream of the FMR1 promoter is detected using the aligned sequence reads. The classification may be that FXS exists when the DNA-methylation boundary is absent.

- 10 [0152] Methods may include haplotyping the repaired set of double-stranded nucleic acid molecules. The HLA gene may be haplotyped. A monogenic disease may be haplotyped by obtaining a plurality of sequence reads from the enriched set of double-stranded nucleic acid molecules. The plurality of sequence reads may be aligned to a reference genome. A mutation may be identified at a locus between two proximal SNP using the aligned plurality of sequence reads. The monogenic disease may be congenital adrenal hyperplasia.
- 15

B. Classifying a disorder

- [0153] FIG. 30 shows a method 3000 of determining a classification of whether an individual has a disorder. The disorder may include systemic lupus erythematosus (SLE), congenital adrenal hyperplasia (CAH), Fragile-X Syndrome, graft-versus-host (GVHD) disease, pregnancy-related disorder (e.g., pre-eclampsia and trophoblastic dysplasia), aneuploidy, sequence imbalance, or any disorder described herein.
- 20

- [0154] At block 3002, a first sample may be received. The first sample may include a first set of double-stranded nucleic acid molecules derived from cell-free nucleic acid molecules in a biological sample. The biological sample may be any biological sample described herein. The one or more double-stranded nucleic acid molecules of the first set of double-stranded nucleic acid molecules may each have one or more defects. The one or more defects may be present in the respective double-stranded nucleic acid molecule at a location at least one nucleotide away from a closest end of the respective double-stranded nucleic acid molecule. The one or more defects may be any defects described herein.
- 25

[0155] At block 3004, a second sample may be received. The second sample may include a second set of double-stranded nucleic acid molecules derived from the cell-free nucleic acid molecules in the biological sample. As an example, the first sample and the second sample may be prepared by dividing a biological sample into two samples. In some embodiments, the first sample and the second sample may be equal in volume, where equal can include volumes that differ within experimental error. The first sample may have a volume within 5%, 10%, or 15% of the volume of the second sample.

[0156] At block 3006, a first mixture including an enzyme may be added to the first sample. The first mixture and the enzyme may be any mixture and enzyme described herein. For example, the mixture may include a plurality of enzymes described herein.

[0157] At block 3008, one or more defects in each of the one or more double-stranded nucleic acid molecules of the first set of double-stranded nucleic acid molecules may be repaired using the enzyme. A repaired first set of double-stranded nucleic acid molecules may be produced from the repair.

[0158] The second sample may be formed from a third sample comprising a third set of double-stranded nucleic acid molecules. A second mixture may be added to the third sample to produce the second set of double-stranded nucleic acid molecules. The second mixture may exclude the enzyme. For example, the second mixture may be a sham mixture that includes components of the first mixture except for the enzyme or enzymes that repair intrastrand defects. The second set of double-stranded nucleic acid molecules may be the same or different from the third set of double-stranded nucleic acid molecules. In other embodiments, the second sample does not have the sham mixture added to it.

[0159] At block 3010, a value of a parameter characterizing a difference in defects between the repaired first set of double-stranded nucleic acid molecules and the second set of double-stranded nucleic acid molecules may be determined.

[0160] The value of the parameter may be determined using the reads determined from the repaired first set of double-stranded nucleic acid molecules and the second set of double-stranded nucleic acid molecules. For example, methods may include sequencing or performing targeted capture enrichment of the repaired first set of double-stranded nucleic acid molecules to

determine a first set of reads. Methods may also include sequencing or performing targeted capture enrichment of the second set of double-stranded nucleic acid molecules to determine a second set of reads. The first set of reads and the second set of reads may not include reads from double-stranded nucleic acid molecules having one or more defects. Damaged DNA fragments that survive the end repairing process may be subject to cluster formation (one of the steps of Illumina-based sequencing). Damaged DNA may cause failed cluster formation, resulting in PCR failure or DNA polymerase stalling. Mismatches may also occur when a nucleotide substitution occurs, which results in a sequencing error. The first set of reads may have a first amount of reads. The second set of reads may have a second amount of reads.

- 10 [0161] The value of the parameter may be determined using the first amount of reads and the second amount of reads. For example, the value of the parameter may be determined using a difference between the first amount of reads and the second amount of reads. In other embodiments, the value may be determined using a ratio of the first amount of reads and the second amount of reads. The ratio may be calculated using the first amount divided by the second amount, the second amount divided by the first amount, the first amount divided by the sum of the first amount and the second amount, or the second amount divided by the sum of the first amount and the second amount. In some embodiments, the amount of reads may be limited to reads for a certain size range or a certain genomic location. The amounts of reads may be a normalized amount of reads (e.g., a percentage, a fraction, or a concentration).
- 20 [0162] The value may be determined from statistical values describing the size of the sets of reads. Methods may include calculating a first statistical value characterizing the sizes of the first set of reads and calculating a second statistical value characterizing the sizes of the second set of reads. The statistical value may be a median, mode, mean, or a percentile describing the sizes of the nucleic acid molecules of the sets of reads. The value may be determined using the first statistical value and the second statistical value. For example, the value may be a difference between the statistical values or a ratio of the statistical values.

- [0163] The parameter may be multidimensional. A first dimension of the parameter may be based on a size of a nucleic acid molecule. The first dimension may be a range of sizes. In other embodiments, the first dimension may be a range of ratios of the size of the nucleic acid molecule to a reference size. A second dimension of the parameter may include an amount of the
- 30

nucleic acid molecule. For example, the parameter may be a matrix of the amount of nucleic acid molecules for certain sizes of nucleic acid molecules. As another example, the parameter may be a graph (e.g., histogram) of the amounts of nucleic acid molecules for certain sizes of nucleic acid molecules.

- 5 [0164] In some embodiments, the first dimension of the parameter may include a location in a reference genome of a nucleic acid molecule. The second dimension may include an amount of the nucleic acid molecule. The parameter may then describe the amounts of nucleic acid molecules at certain locations in the reference genome.

- 10 [0165] The reference value may also be multidimensional. The reference value may specify an amount of nucleic acid molecules at different sizes of the nucleic acid molecules. The reference value may be threshold amounts at different sizes or a certain pattern of amounts among different sizes (e.g, amounts should peak at a certain size). In some embodiments, the reference value may specify an amount of nucleic acid molecules at certain locations in the reference genome. These genomic locations may be may correspond to locations between nucleosomes. The reference value may be determined using one or more subjects identified to have or not have the disorder.

[0166] At block 3012, the value of the parameter may be compared to a reference value. Comparing the value of the parameter may include determining whether the value of the parameter exceeds the reference value (e.g., the value of the parameter being more positive or more negative than the reference value). The reference value may be a threshold value.

- 20 [0167] At block 3014, a classification of whether the individual has the disorder may be determined based on the comparison of the value of the parameter to the reference value. A level or severity of the disorder may be determined by the comparison. The severity may depend on the difference between the value of the parameter and the reference value. A greater difference may mean a more severe disorder. The disorder may be classified as more severe when the difference is greater than a cutoff value. Several cutoff values may be used, with each cutoff value associated with a different severity.

25 [0168] In some embodiments, methods may include treating the specific disorder upon classifying that the individual has the disorder or the individual has a high likelihood of having the disorder. Embodiments may include treating the disease or condition in the patient after

determining the level of the disease or condition in the patient. Treatment may include any suitable therapy, drug, or surgery, including any treatment described in a reference mentioned herein. Information on treatments in the references are incorporated herein by reference.

V. EXAMPLE SYSTEMS

5 [0169] FIG. 31 illustrates a system 3100 according to an embodiment of the present invention. The system as shown includes a sample 3105, such as cell-free DNA molecules within a sample holder 3110, where sample 3105 can be contacted with an assay 3108 to provide a signal of a physical characteristic 3115. In some embodiments, sample 3105 may be a single cell with nucleic acid material. An example of a sample holder can be a flow cell that includes probes
10 and/or primers of an assay or a tube through which a droplet moves (with the droplet including the assay). Physical characteristic 3115, such as a fluorescence intensity value, from the sample is detected by detector 3120. Detector can take a measurement at intervals (e.g., periodic intervals) to obtain data points that make up a data signal. In one embodiment, an analog to digital converter converts an analog signal from the detector into digital form at a plurality of
15 times. A data signal 3125 is sent from detector 3120 to logic system 3130. Data signal 3125 may be stored in a local memory 3135, an external memory 3140, or a storage device 3145.

[0170] Logic system 3130 may be, or may include, a computer system, ASIC, microprocessor, etc. It may also include or be coupled with a display (e.g., monitor, LED display, etc.) and a user input device (e.g., mouse, keyboard, buttons, etc.). Logic system 3130 and the other components
20 may be part of a stand-alone or network connected computer system, or they may be directly attached to or incorporated in a thermal cycler device. Logic system 3130 may also include optimization software that executes in a processor 3150.

[0171] Any of the computer systems mentioned herein may utilize any suitable number of subsystems. Examples of such subsystems are shown in FIG. 32 in computer system 10. In some
25 embodiments, a computer system includes a single computer apparatus, where the subsystems can be the components of the computer apparatus. In other embodiments, a computer system can include multiple computer apparatuses, each being a subsystem, with internal components. A computer system can include desktop and laptop computers, tablets, mobile phones, and other mobile devices.

[0172] The subsystems shown in FIG. 32 are interconnected via a system bus 75. Additional subsystems such as a printer 74, keyboard 78, storage device(s) 79, monitor 76, which is coupled to display adapter 82, and others are shown. Peripherals and input/output (I/O) devices, which couple to I/O controller 71, can be connected to the computer system by any number of connections known in the art such as input/output (I/O) port 77 (e.g., USB, FireWire®). For example, I/O port 77 or external interface 81 (e.g. Ethernet, Wi-Fi, etc.) can be used to connect computer system 10 to a wide area network such as the Internet, a mouse input device, or a scanner. The interconnection via system bus 75 allows the central processor 73 to communicate with each subsystem and to control the execution of a plurality of instructions from system memory 72 or the storage device(s) 79 (e.g., a fixed disk, such as a hard drive, or optical disk), as well as the exchange of information between subsystems. The system memory 72 and/or the storage device(s) 79 may embody a computer readable medium. Another subsystem is a data collection device 85, such as a camera, microphone, accelerometer, and the like. Any of the data mentioned herein can be output from one component to another component and can be output to the user.

[0173] A computer system can include a plurality of the same components or subsystems, e.g., connected together by external interface 81 or by an internal interface. In some embodiments, computer systems, subsystem, or apparatuses can communicate over a network. In such instances, one computer can be considered a client and another computer a server, where each can be part of a same computer system. A client and a server can each include multiple systems, subsystems, or components.

[0174] Aspects of embodiments can be implemented in the form of control logic using hardware (e.g. an application specific integrated circuit or field programmable gate array) and/or using computer software with a generally programmable processor in a modular or integrated manner. As used herein, a processor includes a single-core processor, multi-core processor on a same integrated chip, or multiple processing units on a single circuit board or networked. Based on the disclosure and teachings provided herein, a person of ordinary skill in the art will know and appreciate other ways and/or methods to implement embodiments of the present invention using hardware and a combination of hardware and software.

[0175] Any of the software components or functions described in this application may be implemented as software code to be executed by a processor using any suitable computer language such as, for example, Java, C, C++, C#, Objective-C, Swift, or scripting language such as Perl or Python using, for example, conventional or object-oriented techniques. The software code may be stored as a series of instructions or commands on a computer readable medium for storage and/or transmission. A suitable non-transitory computer readable medium can include random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard-drive or a floppy disk, or an optical medium such as a compact disk (CD) or DVD (digital versatile disk), flash memory, and the like. The computer readable medium may be any combination of such storage or transmission devices.

[0176] Such programs may also be encoded and transmitted using carrier signals adapted for transmission via wired, optical, and/or wireless networks conforming to a variety of protocols, including the Internet. As such, a computer readable medium may be created using a data signal encoded with such programs. Computer readable media encoded with the program code may be packaged with a compatible device or provided separately from other devices (e.g., via Internet download). Any such computer readable medium may reside on or within a single computer product (e.g. a hard drive, a CD, or an entire computer system), and may be present on or within different computer products within a system or network. A computer system may include a monitor, printer, or other suitable display for providing any of the results mentioned herein to a user.

[0177] Any of the methods described herein may be totally or partially performed with a computer system including one or more processors, which can be configured to perform the operations. Thus, embodiments can be directed to computer systems configured to perform the operations of any of the methods described herein, potentially with different components performing a respective operations or a respective group of operations. Although presented as numbered operations, operations of methods herein can be performed at a same time or in a different order. Additionally, portions of these operations may be used with portions of other operations from other methods. Also, all or portions of an operation may be optional. Additionally, any of the operations of any of the methods can be performed with modules, units, circuits, or other approaches for performing these operations.

[0178] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0179] It is to be understood that the methods described herein are not limited to the particular methodology, protocols, subjects, and sequencing techniques described herein and as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the methods and compositions described herein, which will be limited only by the appended claims. While some embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only.

10 Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein may be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

15 [0180] Several aspects are described with reference to example applications for illustration. Unless otherwise indicated, any embodiment can be combined with any other embodiment. It should be understood that numerous specific details, relationships, and methods are set forth to provide a full understanding of the features described herein. A skilled artisan, however, will readily recognize that the features described herein can be practiced without one or more of the

20 specific details or with other methods. The features described herein are not limited by the illustrated ordering of acts or events, as some acts can occur in different orders and/or concurrently with other acts or events. Furthermore, not all illustrated acts or events are required to implement a methodology in accordance with the features described herein.

[0181] While some embodiments of the present invention have been shown and described

25 herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions

30 will now occur to those skilled in the art without departing from the invention.

[0182] Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

[0183] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither, or both limits are included in the smaller ranges is also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included.

[0184] As used herein and in the appended claims, the singular forms “a”, “an”, and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a method” includes a plurality of such methods and reference to “the particle” includes reference to one or more particles and equivalents thereof known to those skilled in the art, and so forth. The invention has now been described in detail for the purposes of clarity and understanding. However, it will be appreciated that certain changes and modifications may be practice within the scope of the appended claims.

VI. REFERENCES EXCLUDING THOSE IN SECTIONS II, III.B, AND III.C

1. Lo YM, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet*. 1997;350(9076):485-487.
2. Lo YM, Chan KC, Sun H, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med*. 2010;2(61):61ra91.

3. Lun FM, Tsui NB, Chan KC, et al. Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc Natl Acad Sci U S A*. 2008;105(50):19920-19925.
4. Yu SC, Chan KC, Zheng YW, et al. Size-based molecular diagnostics using plasma DNA
5 for noninvasive prenatal testing. *Proc Natl Acad Sci U S A*. 2014;111(23):8583-8588.
5. Jiang P, Peng X, Su X, et al. FetalQuant. *NPJ Genom Med*. 2016;1:16013.
6. Chan KC, Ding C, Gerovassili A, et al. Hypermethylated RASSF1A in maternal plasma: A universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. *Clin Chem*. 2006;52(12):2211-2218.
- 10 7. Dhallan R, Au WC, Mattagajasingh S, et al. Methods to increase the percentage of free fetal DNA recovered from the maternal circulation. *JAMA*. 2004;291(9):1114-1119.
8. Li Y, Di Naro E, Vitucci A, Zimmermann B, Holzgreve W, Hahn S. Detection of paternally inherited fetal point mutations for beta-thalassemia using size-fractionated cell-free DNA in maternal plasma. *JAMA*. 2005;293(7):843-849.
- 15 9. Burnham P, Kim MS, Agbor-Enoh S, et al. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci Rep*. 2016;6:27859.
10. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016;164(1-2):57-
20 68.
11. Nagata S. Apoptotic DNA fragmentation. *Exp Cell Res*. 2000;256(1):12-18.
12. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin Chim Acta*. 2001;313(1-2):139-142.
- 25 13. van der Vaart M, Pretorius PJ. The origin of circulating free DNA. *Clin Chem*. 2007;53(12):2215.
14. Mouliere F, El Messaoudi S, Pang D, Dritschilo A, Thierry AR. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol*. 2014;8(5):927-941.

15. Chandrananda D, Thorne NP, Bahlo M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med Genomics*. 2015;8:29.
16. Chan KC, Zhang J, Hui AB, et al. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem*. 2004;50(1):88-92.
17. Vong JSL, Tsang JCH, Jiang P, et al. Single-Stranded DNA Library Preparation Preferentially Enriches Short Maternal DNA in Maternal Plasma. *Clin Chem*. 2017;63(5):1031-1037.
18. New MI, Tong YK, Yuen T, et al. Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. *J Clin Endocrinol Metab*. 2014;99(6):E1022-1030.
19. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25(15):1966-1967.
20. Hong H, Xu L, Tong W. Assessing consistency between versions of genotype-calling algorithm Birdseed for the Genome-Wide Human SNP Array 6.0 using HapMap samples. *Adv Exp Med Biol*. 2010;680:355-360.
21. Cheng THT, Lui KO, Peng XL, et al. DNase1 Does Not Appear to Play a Major Role in the Fragmentation of Plasma DNA in a Knockout Mouse Model. *Clin Chem*. 2018;64(2):406-408.
22. Chen L, Liu P, Evans TC, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science*. 2017;355(6326):752-756.
23. Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res*. 2018;46(5):2159-2168.
24. Giacona MB, Ruben GC, Iczkowski KA, Roos TB, Porter DM, Sorenson GD. Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas*. 1998;17(1):89-97.
25. Holdenrieder S, Stieber P, Förg T, et al. Apoptosis in serum of patients with solid tumours. *Anticancer Res*. 1999;19(4A):2721-2724.

26. Jahr S, Hentze H, Englisch S, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* 2001;61(4):1659-1665.
27. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* 2016;35(3):347-376.
28. Sun K, Jiang P, Wong AIC, et al. Size-tagged preferred ends in maternal plasma DNA shed light on the production mechanism and show utility in noninvasive prenatal testing. *Proc Natl Acad Sci U S A.* 2018;115(22):E5106-E5114.
29. Chen EZ, Chiu RW, Sun H, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One.* 2011;6(7):e21791.
30. Chiu RW, Akolekar R, Zheng YW, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ.* 2011;342:c7401.
31. Palomaki GE, Deciu C, Kloza EM, et al. DNA sequencing of maternal plasma reliably identifies trisomy 18 and trisomy 13 as well as Down syndrome: an international collaborative study. *Genet Med.* 2012;14(3):296-305.
32. Nicolaides KH, Syngelaki A, Ashoor G, Birdir C, Touzet G. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. *Am J Obstet Gynecol.* 2012;207(5):374.e371-376.
33. Pergament E, Cuckle H, Zimmermann B, et al. Single-nucleotide polymorphism-based noninvasive prenatal screening in a high-risk and low-risk cohort. *Obstet Gynecol.* 2014;124(2 Pt 1):210-218.
34. Canick JA, Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE. The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies. *Prenat Diagn.* 2013;33(7):667-674.
35. Dar P, Curnow KJ, Gross SJ, et al. Clinical experience and follow-up with large scale single-nucleotide polymorphism-based noninvasive prenatal aneuploidy testing. *Am J Obstet Gynecol.* 2014;211(5):527.e521-527.e517.
36. Norton ME, Wapner RJ. Cell-free DNA Analysis for Noninvasive Examination of Trisomy. *N Engl J Med.* 2015;373(26):2582.

37. Yared E, Dinsmoor MJ, Endres LK, et al. Obesity increases the risk of failure of noninvasive prenatal screening regardless of gestational age. *Am J Obstet Gynecol.* 2016;215(3):370.e371-376.
38. Nolin SL, Brown WT, Glicksman A, et al. Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am J Hum Genet.* 2003;72(2):454-464.
39. McLennan Y, Polussa J, Tassone F, Hagerman R. Fragile x syndrome. *Curr Genomics.* 2011;12(3):216-224.
40. Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YM. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem.* 2002;48(3):421-427.
41. Diegoli TM, Farr M, Cromartie C, Coble MD, Bille TW. An optimized protocol for forensic application of the PreCR™ Repair Mix to multiplex STR amplification of UV-damaged DNA. *Forensic Sci Int Genet.* 2012;6(4):498-503.
42. Ye J, Chen C, Yuan Y, et al. Haplotype-based Noninvasive Prenatal Diagnosis of Hyperphenylalaninemia through Targeted Sequencing of Maternal Plasma. *Sci Rep.* 2018;8(1):161.
43. Ma D, Yuan Y, Luo C, et al. Noninvasive prenatal diagnosis of 21-Hydroxylase deficiency using target capture sequencing of maternal plasma DNA. *Sci Rep.* 2017;7(1):7427.
44. Rajan-Babu IS, Lian M, Cheah FSH, et al. FMR1 CGG repeat expansion mutation detection and linked haplotype analysis for reliable and accurate preimplantation genetic diagnosis of fragile X syndrome. *Expert Rev Mol Med.* 2017;19:e10.
45. Tural S, Tekcan A, Kara N, Elbistan M, Güven D, Ali Tasdemir H. FMR1 gene mutation screening by TP-PCR in patients with premature ovarian failure and fragile-X. *Gynecol Endocrinol.* 2015;31(3):191-195.
46. Naumann A, Hochstein N, Weber S, Fanning E, Doerfler W. A distinct DNA-methylation boundary in the 5'-upstream sequence of the FMR1 promoter binds nuclear proteins and is lost in fragile X syndrome. *Am J Hum Genet.* 2009;85(5):606-616.
47. Hosomichi K, Shiina T, Tajima A, Inoue I. The impact of next-generation sequencing technologies on HLA research. *J Hum Genet.* 2015;60(11):665-673.

48. Sun K, Jiang P, Chan KC, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A*. 2015;112(40):E5503-5512.
49. Lehmann-Werman R, Neiman D, Zemmour H, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A*. 2016;113(13):E1826-1834.
50. Chan RW, Jiang P, Peng X, et al. Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing. *Proc Natl Acad Sci U S A*. 2014;111(49):E5302-5311.
- 10 51. Napirei M, Ludwig S, Mezrhah J, Klöckl T, Mannherz HG. Murine serum nucleases--contrasting effects of plasmin and heparin on the activities of DNase1 and DNase1-like 3 (DNase113). *FEBS J*. 2009;276(4):1059-1073.

WHAT IS CLAIMED IS:

- 1 1. A method of improving analysis of a first biological sample, the first
2 biological sample including cell-free nucleic acid molecules, the method comprising:
3 obtaining a plurality of double-stranded nucleic acid molecules from the cell-free
4 nucleic acid molecules to produce a second biological sample, wherein:
5 one or more double-stranded nucleic acid molecules of the plurality of
6 double-stranded nucleic acid molecules each have one or more defects, and
7 for each of the one or more double-stranded nucleic acid molecules having
8 a defect of the one or more defects:
9 the defect is present in the respective double-stranded nucleic acid
10 molecule at a location at least one nucleotide away from a closest end of the
11 respective double-stranded nucleic acid molecule;
12 adding a mixture comprising an enzyme to the second biological sample; and
13 repairing the one or more defects in each of the one or more double-stranded
14 nucleic acid molecules using the enzyme to produce a repaired set of double-stranded nucleic
15 acid molecules.
- 1 2. The method of claim 1, further comprising:
2 producing a sequencing library using the repaired set of double-stranded nucleic
3 acid molecules, and
4 detecting an aneuploidy or a sequence imbalance using the sequencing library.
- 1 3. The method of claim 2, further comprising:
2 determining whether a sequence imbalance exists in the second biological sample
3 by:
4 obtaining a plurality of sequence reads from the repaired set of double-
5 stranded nucleic acid molecules,
6 analyzing, by a computer system, the plurality of sequence reads, wherein
7 analyzing a sequence read includes:
8 identifying a location of the sequence read in a reference genome
9 by aligning the sequence read to the reference genome,

10 determining, by a computer system, an amount of sequence reads from a
11 genomic region using the identified locations,
12 obtaining a value of a normalized parameter for the amount of sequence
13 reads from the genomic region,
14 comparing the value of the normalized parameter to a cutoff value,
15 determining whether the sequence imbalance exists based on the
16 comparison.

1 4. The method of claim 3, wherein the genomic region is a chromosome, and
2 the sequence imbalance is an aneuploidy.

1 5. The method of claim 2, further comprising sequencing the sequencing
2 library or performing targeted capture enrichment of the sequencing library to determine a set of
3 reads.

1 6. The method of claim 5, wherein the set of reads does not comprise reads
2 from any strand of the one or more double-stranded nucleic acid molecules having one or more
3 defects.

1 7. The method of claim 1, wherein the one or more defects are selected from
2 the group consisting of a nick, gap, abasic site, thymidine dimer, oxidized pyrimidine,
3 deaminated cytosine, blocked 3' end defect, or a combination thereof.

1 8. The method of claim 1, wherein the repaired set of double-stranded
2 nucleic acid molecules are free of a nick, gap, abasic site, thymidine dimer, oxidized pyrimidine,
3 deaminated cytosine, blocked 3' end defect, or a combination thereof.

1 9. The method of claim 1, wherein the enzyme comprises at least one of a
2 polymerase, a ligase, an endonuclease, or a glycosylase.

1 10. The method of claim 1, wherein the mixture comprises at least one of *Taq*
2 DNA polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV,
3 T4 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), or uracil-DNA glycosylase (UDG).

1 11. The method of claim 1, wherein the mixture comprises *Taq* DNA
2 polymerase, *Bst* DNA polymerase, *Taq* DNA ligase, endonuclease VIII, endonuclease IV, T4
3 endonuclease V (PDG), 8-oxoguanine glycosylase (FPG), and uracil-DNA glycosylase (UDG).

1 12. The method of claim 1, wherein the first biological sample is maternal
2 plasma from a female subject pregnant with a fetus.

1 13. The method of claim 1, wherein the one or more double-stranded nucleic
2 acid molecules have lengths in a range from 251 to 600 bp, from 251 to 450 bp, or from 451 bp
3 to 600 bp.

1 14. The method of claim 1, wherein the one or more double-stranded nucleic
2 acid molecules have lengths associated with dinucleosomal nucleic acid molecules or
3 trinucleosomal nucleic acid molecules.

1 15. The method of claim 1, wherein repairing the one or more defects in each
2 of the one or more double-stranded nucleic acid molecules comprises repairing a greater number
3 of nucleic acid molecules having lengths from 251 to 600 bp than nucleic acid molecules having
4 lengths from 0 to 250 bp.

1 16. The method of claim 1, wherein:
2 the first biological sample is obtained from a female subject pregnant with a fetus,
3 the one or more double-stranded nucleic acid molecules comprise a plurality of
4 fetal-derived nucleic acid molecules,
5 after repairing the one or more defects, the second biological sample is
6 characterized by a fetal fraction calculated from reads obtained from analyzing the repaired set of
7 double-stranded nucleic acid molecules, and
8 the fetal fraction is greater than 0.05.

1 17. The method of claim 1, further comprising:
2 performing blunt-end ligation of the plurality of double-stranded nucleic acid
3 molecules or of the repaired set of double-stranded nucleic acid molecules.

1 18. The method of claim 1, further comprising:

2 haplotyping the HLA gene using the repaired set of double-stranded nucleic acid
3 molecules.

1 19. The method of claim 1, further comprising:
2 haplotyping a monogenic disease by:
3 obtaining a plurality of sequence reads from the enriched set of double-
4 stranded nucleic acid molecules,
5 aligning the plurality of sequence reads to a reference genome, and
6 identifying a mutation at a locus between two proximal SNP using the
7 aligned plurality of sequence reads.

1 20. The method of claim 19, wherein the monogenic disease is congenital
2 adrenal hyperplasia.

1 21. A method of determining a classification of whether an individual has a
2 disorder, the method comprising:
3 receiving a first sample comprising a first set of double-stranded nucleic acid
4 molecules derived from cell-free nucleic acid molecules in a biological sample, wherein:
5 one or more double-stranded nucleic acid molecules of the first set of
6 double-stranded nucleic acid molecules each have one or more defects, and
7 the one or more defects are present in the respective double-stranded
8 nucleic acid molecule at a location at least one nucleotide away from a closest end of the
9 respective double-stranded nucleic acid molecule;
10 receiving a second sample comprising a second set of double-stranded nucleic
11 acid molecules derived from the cell-free nucleic acid molecules in the biological sample;
12 adding a first mixture comprising an enzyme to the first sample;
13 repairing the one or more defects in each of the one or more double-stranded
14 nucleic acid molecules of the first set of double-stranded nucleic acid molecules using the
15 enzyme to produce a repaired first set of double-stranded nucleic acid molecules;
16 determining a value of a parameter characterizing a difference in defects between
17 the repaired first set of double-stranded nucleic acid molecules and the second set of double-
18 stranded nucleic acid molecules;

19 comparing the value of the parameter to a reference value; and
20 determining the classification of whether the individual has the disorder based on
21 the comparison of the value of the parameter to the reference value.

1 22. The method of claim 21, wherein:
2 the first mixture comprises a plurality of enzymes.

1 23. The method of claim 21, wherein the first sample and the second sample
2 are equal in volume.

1 24. The method of claim 21, further comprising:
2 adding a second mixture to a third sample comprising a third set of double-
3 stranded nucleic acid molecules derived from the cell-free nucleic acid molecules in the
4 biological sample to produce the second sample comprising the second set of double-stranded
5 nucleic acid molecules, the second mixture excluding the enzyme.

1 25. The method of claim 24, wherein:
2 the first mixture comprises a plurality of enzymes, and
3 the second mixture excludes the plurality of enzymes.

1 26. The method of claim 21, further comprising:
2 sequencing or performing target capture enrichment of the repaired first set of
3 double-stranded nucleic acid molecules to determine a first set of reads, and
4 sequencing or performing targeted capture enrichment of the second set of
5 double-stranded nucleic acid molecules to determine a second set of reads, wherein:
6 the first set of reads does not comprise reads from double-stranded nucleic
7 acid molecules having one or more defects,
8 the second set of reads does not comprise reads from double-stranded
9 nucleic acid molecules having one or more defects,
10 the first set of reads has a first amount of reads,
11 the second set of reads has a second amount of reads, and
12 the value of the parameter is determined using the first amount of reads
13 and the second amount of reads.

1 27. The method of claim 26, further comprising:
2 calculating a first statistical value characterizing the sizes of the first set of reads,
3 calculating a second statistical value characterizing the sizes of the second set of
4 reads,
5 wherein:
6 the value of the parameter is determined using the first statistical value and
7 the second statistical value.

1 28. The method of claim 26, wherein:
2 the value of the parameter is determined using a difference between the first
3 amount of reads and the second amount of reads, or
4 the value of the parameter is determined using a ratio of the first amount of reads
5 and the second amount of reads.

1 29. The method of claim 26, wherein:
2 the first amount of reads is a normalized amount of reads, and
3 the second amount of reads is a normalized amount of reads.

1 30. The method of claim 21, wherein:
2 the parameter is multidimensional,
3 a first dimension of the parameter is based on a size of a nucleic acid molecule,
4 and
5 a second dimension of the parameter comprises an amount of the nucleic acid
6 molecules.

1 31. The method of claim 30, wherein the first dimension is a range of sizes or
2 a range of ratios of the size of the nucleic acid molecule to a reference size.

1 32. The method of claim 30, wherein:
2 the reference value is multidimensional, and
3 the reference value specifies amounts of nucleic acid molecules at different sizes
4 of the nucleic acid molecules.

1 33. The method of claim 21, wherein:
2 the parameter is multidimensional,
3 a first dimension of the parameter comprises a location in a reference genome of a
4 nucleic acid molecule, and
5 a second dimension of the parameter comprises an amount of the nucleic acid
6 molecule.

1 34. The method of claim 21, wherein the reference value is determined using
2 one or more subjects identified to have the disorder or from one or more subjects identified to
3 not have the disorder.

1 35. The method of claim 34, wherein comparing the value of the parameter to
2 the reference value comprises determining whether the value of the parameter exceeds the
3 reference value.

1 36. The method of claim 21, wherein the disorder is systemic lupus
2 erythematosus, congenital adrenal hyperplasia (CAH), Fragile-X Syndrome, graft-versus-host
3 (GVHD) disease, a pregnancy-related disorder, an aneuploidy, or a sequence imbalance.

1 37. The method of claim 35, wherein:
2 determining the classification is based on a difference between the value of the
3 parameter and the reference value, and
4 the classification comprises a severity of the disorder, the classification indicating
5 a more severe disorder when the difference is greater than a cutoff value.

1 38. A computer product comprising a computer readable medium storing a
2 plurality of instructions for controlling a system to perform any one of claims 21 to 37.

1 39. A system comprising:
2 the computer product of claim 38; and
3 one or more processors for executing instructions stored on the computer readable
4 medium.

1 40. A system comprising means for performing any of the above methods.

- 1 41. A system comprising one or more processors configured to perform any of
2 the above methods.
- 1 42. A system comprising modules that respectively perform the steps of any
2 of the above methods.