

## **METHODS FOR FINGERPRINTING OF BIOLOGICAL SAMPLES**

### **CROSS-REFERENCE**

**[0001]** This application claims the benefit of U.S. Provisional Patent Application No. 62/681,642, filed June 6, 2018, entitled METHODS FOR FINGERPRINTING OF BIOLOGICAL SAMPLES, which is entirely incorporated herein by reference.

### **BACKGROUND**

**[0002]** The collection and assaying of biological samples obtained from subjects may often encounter challenges with reliable maintenance of sample identity throughout clinical and laboratory processes. For example, biological samples may often be inadvertently swapped in laboratory or clinical settings, thereby resulting in potentially incorrect clinical results if left undetected and uncorrected.

### **SUMMARY**

**[0003]** Methods for fingerprinting biological samples using panels of genetic loci may require sufficiently deep coverage to obtain genetic information at a desired sensitivity, specificity, or accuracy. For example, deep coverage may be required for a sufficiently high signal-to-noise ratio (SNR) to distinguish between fingerprints generated from different samples. Such samples may be longitudinal samples (e.g., obtained from the same subject at two different time points). Longitudinal samples processed using low-pass sequencing may encounter challenges with (1) correcting matching together samples from different time points and (2) identifying a panel of genetic loci suitable for sample fingerprinting despite relatively low read coverage at any one location.

**[0004]** Methods and systems are provided for generating and comparing fingerprints of biological samples. Sample fingerprints may be generated by sequencing one or more sets of nucleic acid molecules from biological samples obtained from a subject at each of one or more time points. Pairwise comparison of sample fingerprints may be performed to determine whether a sample mismatch (e.g., that the two samples were obtained from different subjects) or a sample match (e.g., that the two samples were obtained from the same subject) is present between the two biological samples from which the sample fingerprints were generated.

**[0005]** In an example, the present disclosure provides a method for identifying a sample mismatch, comprising: obtaining a first biological sample comprising a first plurality of nucleic acid molecules from a subject; processing, by a computer, the first plurality of nucleic acid molecules to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of

genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second biological sample comprising a second plurality of nucleic acid molecules from the subject; processing, by a computer, the second plurality of nucleic acid molecules to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold. Additionally, in this example, the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measures of the first plurality of nucleic acid molecules.

**[0006]** In another example, the present disclosure provides a method for identifying a sample mismatch, comprising: obtaining a first biological sample comprising a first plurality of nucleic acid molecules from a subject; processing, by a computer, the first plurality of nucleic acid molecules to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second biological sample comprising a second plurality of nucleic acid molecules from the subject; processing, by a computer, the second plurality of nucleic acid molecules to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold. Additionally, in this example, the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms. In one aspect, the present disclosure refers to a method for identifying a sample mismatch, comprising:

- obtaining a first biological sample comprising a first plurality of nucleic acid molecules from a subject;

- processing, by a computer, the first plurality of nucleic acid molecules to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs);

- obtaining a second biological sample comprising a second plurality of nucleic acid molecules from the subject;

- processing, by a computer, the second plurality of nucleic acid molecules to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci;

- determining a difference between the first sample fingerprint and the second sample fingerprint;
- and

- identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms.

**[0007]** In another example, the present disclosure provides a method for identifying a sample mismatch, comprising: obtaining a first biological sample comprising a first plurality of nucleic acid molecules from a subject; processing, by a computer, the first plurality of nucleic acid molecules to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second biological sample comprising a second plurality of nucleic acid molecules from the subject; processing, by a computer, the second plurality of nucleic acid molecules to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first

sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold. Additionally, in this example, the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a pre-determined threshold. In some examples where the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a particular threshold, the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds about 7.5%.

**[0008]** In some examples, the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules comprise cell-free DNA (cfDNA). In some examples, the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules comprise buffy coat DNA. In some examples, the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules comprise solid tumor DNA.

**[0009]** In some examples, the second biological sample is obtained from the subject at a later time after obtaining the first biological sample. In some examples, processing the first plurality of nucleic acid molecules comprises sequencing the first plurality of nucleic acid molecules to generate a first plurality of sequencing reads, and processing the second plurality of nucleic acid molecules comprises sequencing the second plurality of nucleic acid molecules to generate a second plurality of sequencing reads.

**[0010]** In some examples, the sequencing comprises whole genome sequencing (WGS). In some examples, the sequencing is performed at a depth of no more than about 10X. In some examples, the sequencing is performed at a depth of no more than about 8X. In some examples, the sequencing is performed at a depth of no more than about 6X. In some examples, the quantitative measure of the first plurality of nucleic acid molecules comprises a coverage of the first plurality of nucleic acid molecules at each of the plurality of genetic loci, and the quantitative measure of the second plurality of nucleic acid molecules comprises a coverage of the second plurality of nucleic acid molecules at each of the plurality of genetic loci.

**[0011]** In some examples, processing the first plurality of nucleic acid molecules comprises performing binding measurements of the first plurality of nucleic acid molecules, and processing the second plurality of nucleic acid molecules comprises performing binding measurements of the second plurality of nucleic acid molecules. In some examples, the quantitative measure of the first plurality of nucleic acid molecules at each of the plurality of genetic loci comprises a number of the first plurality of nucleic acid molecules containing the genetic locus, and the quantitative measure of the second plurality of nucleic acid molecules at

each of the plurality of genetic loci comprises a number of the second plurality of nucleic acid molecules containing the genetic locus.

**[0012]** In some examples, the method further comprises enriching the first plurality of nucleic acid molecules and/or the second plurality of nucleic acid molecules for at least a portion of the plurality of genetic loci. In some examples, the enrichment comprises amplifying at least a portion of the first plurality of nucleic acid molecules and/or the second plurality of nucleic acid molecules. In some examples, the amplification comprises selective amplification. In some examples, the amplification comprises universal amplification. In some examples, the enrichment comprises selectively isolating at least a portion of the first plurality of nucleic acid molecules and/or the second plurality of nucleic acid molecules.

**[0013]** In some examples, the plurality of genetic loci comprises at least about 50 distinct autosomal single nucleotide polymorphisms (SNPs). In some examples, the plurality of genetic loci comprises at least about 100 distinct autosomal single nucleotide polymorphisms (SNPs).

**[0014]** In some examples, generating the first sample fingerprint further comprises obtaining a third biological sample comprising a third plurality of nucleic acid molecules from the subject, and processing the third plurality of nucleic acid molecules to obtain a quantitative measure of the third plurality of nucleic acid molecules at each of a second plurality of genetic loci, wherein the second plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); and generating the second sample fingerprint further comprises obtaining a fourth biological sample comprising a fourth plurality of nucleic acid molecules from the subject, and processing the fourth plurality of nucleic acid molecules to obtain a quantitative measure of the fourth plurality of nucleic acid molecules at each of the second plurality of genetic loci.

**[0015]** In some examples, the third plurality of nucleic acid molecules and the fourth plurality of nucleic acid molecules comprise cell-free DNA (cfDNA). In some examples, the third plurality of nucleic acid molecules and the fourth plurality of nucleic acid molecules comprise buffy coat DNA. In some examples, the third plurality of nucleic acid molecules and the fourth plurality of nucleic acid molecules comprise solid tumor DNA. In some examples, generating the first sample fingerprint further comprises obtaining a fifth biological sample comprising a fifth plurality of nucleic acid molecules from the subject, and processing the fifth plurality of nucleic acid molecules to obtain a quantitative measure of the fifth plurality of nucleic acid molecules at each of a third plurality of genetic loci, wherein the third plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); and

generating the second sample fingerprint further comprises obtaining a sixth biological sample comprising a sixth plurality of nucleic acid molecules from the subject, and processing the sixth plurality of nucleic acid molecules to obtain a quantitative measure of the sixth plurality of nucleic acid molecules at each of the third plurality of genetic loci.

**[0016]** In some examples, the third plurality of nucleic acid molecules and the fourth plurality of nucleic acid molecules comprise cell-free DNA (cfDNA). In some examples, the third plurality of nucleic acid molecules and the fourth plurality of nucleic acid molecules comprise buffy coat DNA. In some examples, the third plurality of nucleic acid molecules and the fourth plurality of nucleic acid molecules comprise solid tumor DNA.

**[0017]** In some examples, the method comprises identifying the sample mismatch with a sensitivity of at least about 90%. In some examples, identifying the sample mismatch is performed with a sensitivity of at least about 95%. In some examples, the method comprises identifying the sample mismatch with a sensitivity of at least about 99%.

**[0018]** In some examples, the method comprises identifying the sample mismatch with a specificity of at least about 90%. In some examples, the method comprises identifying the sample mismatch with a specificity of at least about 95%. In some examples, the method comprises identifying the sample mismatch with a specificity of at least about 99%.

**[0019]** In some examples, the method comprises identifying the sample mismatch with a positive predictive value (PPV) of at least about 90%. In some examples, the method comprises identifying the sample mismatch with a positive predictive value (PPV) of at least about 95%. In some examples, the method comprises identifying the sample mismatch with a positive predictive value (PPV) of at least about 99%.

**[0020]** In some examples, the method comprises identifying the sample mismatch with a negative predictive value (NPV) of at least about 90%. In some examples, the method comprises identifying the sample mismatch with a negative predictive value (NPV) of at least about 95%. In some examples, the method comprises identifying the sample mismatch with a negative predictive value (NPV) of at least about 99%.

**[0021]** In some examples, the method comprises identifying the sample mismatch with an area under the curve (AUC) of at least about 0.90. In some examples, the method comprises identifying the sample mismatch with an area under the curve (AUC) of at least about 0.95. In some examples, the method comprises identifying the sample mismatch with an area under the curve (AUC) of at least about 0.99.

**[0022]** In some examples, the predetermined criterion is that the difference comprises a difference in genotype similarity greater than a predetermined threshold. In some examples, the predetermined threshold is about 0.8.

**[0023]** In some examples, the method further comprises excluding the second biological sample from further assaying based on the identified sample mismatch.

**[0024]** In some examples, the method further comprises identifying a sample match when the difference between the first sample fingerprint and the second sample fingerprint does not satisfy the predetermined criterion.

**[0025]** In some examples, the method comprises identifying the sample match with a sensitivity of at least about 90%. In some examples, the method comprises identifying the sample match with a sensitivity of at least about 95%. In some examples, the method comprises identifying the sample match with a sensitivity of at least about 99%.

**[0026]** In some examples, the method comprises identifying the sample match with a specificity of at least about 90%. In some examples, the method comprises identifying the sample match with a specificity of at least about 95%. In some examples, the method comprises identifying the sample match with a specificity of at least about 99%.

**[0027]** In some examples, the method comprises identifying the sample match with a positive predictive value (PPV) of at least about 90%. In some examples, the method comprises identifying the sample match with a positive predictive value (PPV) of at least about 95%. In some examples, the method comprises identifying the sample match with a positive predictive value (PPV) of at least about 99%.

**[0028]** In some examples, the method comprises identifying the sample match with a negative predictive value (NPV) of at least about 90%. In some examples, the method comprises identifying the sample match with a negative predictive value (NPV) of at least about 95%. In some examples, the method comprises identifying the sample match with a negative predictive value (NPV) of at least about 99%.

**[0029]** In some examples, the method comprises identifying the sample match with an area under the curve (AUC) of at least about 0.90. In some examples, the method comprises identifying the sample match with an area under the curve (AUC) of at least about 0.95. In some examples, the method comprises identifying the sample match with an area under the curve (AUC) of at least about 0.99.

**[0030]** In some examples, the method further comprises subjecting the second biological sample to further assaying based on the identified sample match. In some examples, the

method further comprises, based on the identified sample match, storing the second sample fingerprint in a database, and optionally, storing the first sample fingerprint in the database.

**[0031]** In another example, the present disclosure provides a non-transitory computer-readable medium comprising machine-executable code that, upon execution by one or more computer processors, implements a method for identifying a sample mismatch, comprising: receiving information of a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules of a first biological sample at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs), and wherein the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measures of the plurality of nucleic acid molecules; receiving information of a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules of a second biological sample at each of the plurality of genetic loci, wherein the second biological sample is obtained from the subject; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint satisfies a predetermined criterion. In one aspect, the present disclosure refers to a non-transitory computer-readable medium comprising machine-executable code that, upon execution by one or more computer processors, implements a method for identifying a sample mismatch, comprising:

- receiving information of a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules of a first biological sample at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms;

- receiving information of a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules of a second biological sample at each of the plurality of genetic loci, wherein the second biological sample is obtained from the subject;

- determining a difference between the first sample fingerprint and the second sample fingerprint; and

- identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint satisfies a predetermined threshold.

**[0032]** In another example, the present disclosure provides a computer-implemented method for identifying a sample mismatch, comprising: processing a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); processing the second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measures of the first plurality of nucleic acid molecules.

**[0033]** In another example, the present disclosure provides a computer-implemented method for identifying a sample mismatch, comprising: processing a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at

each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); processing the second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms.

**[0034]** In another example, the present disclosure provides a computer-implemented method for identifying a sample mismatch, comprising: processing a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); processing the second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a pre-determined threshold.

**[0035]** In another example, the present disclosure provides a system, comprising a controller comprising, or capable of accessing, computer readable media comprising non-transitory computer-executable instructions which, when executed by at least one electronic processor perform at least: processing a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); processing the second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying a sample mismatch when the difference between the

first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measures of the first plurality of nucleic acid molecules.

**[0036]** In another example, the present disclosure provides a system, comprising a controller comprising, or capable of accessing, computer readable media comprising non-transitory computer-executable instructions which, when executed by at least one electronic processor perform at least: processing a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); processing the second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying a sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms.

**[0037]** In another example, the present disclosure provides a system, comprising a controller comprising, or capable of accessing, computer readable media comprising non-transitory computer-executable instructions which, when executed by at least one electronic processor perform at least: processing a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); processing the second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying a sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a pre-determined threshold.

**[0038]** In another example, the present disclosure provides a computer-implemented method for identifying a sample mismatch, comprising: obtaining a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measures of the first plurality of nucleic acid molecules.

**[0039]** In another example, the present disclosure provides a computer-implemented method for identifying a sample mismatch, comprising: obtaining a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms.

**[0040]** In another example, the present disclosure provides a computer-implemented method for identifying a sample mismatch, comprising: obtaining a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold,

wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a pre-determined threshold.

**[0041]** In another example, the present disclosure provides a system, comprising a controller comprising, or capable of accessing, computer readable media comprising non-transitory computer-executable instructions which, when executed by at least one electronic processor perform at least: obtaining a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying a sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measures of the first plurality of nucleic acid molecules.

**[0042]** In another example, the present disclosure provides a system, comprising a controller comprising, or capable of accessing, computer readable media comprising non-transitory computer-executable instructions which, when executed by at least one electronic processor perform at least: obtaining a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject) at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying a sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms.

**[0043]** In another example, the present disclosure provides a system, comprising a controller comprising, or capable of accessing, computer readable media comprising non-transitory computer-executable instructions which, when executed by at least one electronic processor perform at least: obtaining a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules (e.g., from a first biological sample obtained from a subject)

at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); obtaining a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules (e.g., from a second biological sample obtained from the subject) at each of the plurality of genetic loci; determining a difference between the first sample fingerprint and the second sample fingerprint; and identifying a sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold, wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a pre-determined threshold.

**[0043a]** In one aspect, the present disclosure refers to a system, comprising:

- one or more processors;
- a non-transitory computer-readable medium comprising machine-executable code that, upon execution by the one or more processors, implements a method for identifying a sample mismatch, comprising:
  - receiving information of a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules of a first biological sample at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms;
  - receiving information of a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules of a second biological sample at each of the plurality of genetic loci, wherein the second biological sample is obtained from the subject;
  - determining a difference between the first sample fingerprint and the second sample fingerprint; and
  - identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint satisfies a predetermined threshold.

**[0044]** Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative examples of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different examples, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

## **INCORPORATION BY REFERENCE**

**[0045]** All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0046]** Some novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also “Figure” and “FIG.” herein), of which:

**[0047]** **FIG. 1** illustrates an example of a method for fingerprinting of biological samples, in accordance with some embodiments.

**[0048]** **FIG. 2** illustrates an example of a method for identifying sample mismatches based on fingerprinting a first biological sample and a second biological sample, in accordance with some embodiments.

## CLAIMS

### WHAT IS CLAIMED IS:

1. A method for identifying a sample mismatch, comprising:
  - obtaining a first biological sample comprising a first plurality of nucleic acid molecules from a subject;
  - processing, by a computer, the first plurality of nucleic acid molecules to generate a first sample fingerprint comprising a quantitative measure of the first plurality of nucleic acid molecules at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs);
  - obtaining a second biological sample comprising a second plurality of nucleic acid molecules from the subject;
  - processing, by a computer, the second plurality of nucleic acid molecules to generate a second sample fingerprint comprising a quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci;
  - determining a difference between the first sample fingerprint and the second sample fingerprint; and
  - identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint exceeds a pre-determined threshold,
  - wherein the autosomal single nucleotide polymorphisms comprise simple single nucleotide polymorphisms.
2. The method of claim 1, wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds about 7.5%.
3. The method of any of claims 1-2, wherein the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules comprise cell-free DNA (cfDNA), buffy coat DNA, or solid tumor DNA.
4. The method of any of claims 1-2, wherein the second biological sample is obtained from the subject at a later time after obtaining the first biological sample.
5. The method of any of claims 1-2, wherein processing the first plurality of nucleic acid molecules comprises sequencing the first plurality of nucleic acid molecules to generate a first plurality of sequencing reads, and wherein processing the second plurality of nucleic acid molecules comprises sequencing the second plurality of nucleic acid molecules to generate a second plurality of sequencing reads.

6. The method of claim 5, wherein the sequencing comprises whole genome sequencing (WGS).
7. The method of claim 6, wherein the sequencing is performed at a depth of no more than about 6X, no more than about 8X, or no more than about 10X.
8. The method of claim 5, wherein the quantitative measure of the first plurality of nucleic acid molecules comprises a coverage of the first plurality of nucleic acid molecules at each of the plurality of genetic loci, and wherein the quantitative measure of the second plurality of nucleic acid molecules comprises a coverage of the second plurality of nucleic acid molecules at each of the plurality of genetic loci.
9. The method of any of claims 1-2, wherein processing the first plurality of nucleic acid molecules comprises performing binding measurements of the first plurality of nucleic acid molecules, and wherein processing the second plurality of nucleic acid molecules comprises performing binding measurements of the second plurality of nucleic acid molecules.
10. The method of claim 9, wherein the quantitative measure of the first plurality of nucleic acid molecules at each of the plurality of genetic loci comprises a number of the first plurality of nucleic acid molecules containing the genetic locus, and wherein the quantitative measure of the second plurality of nucleic acid molecules at each of the plurality of genetic loci comprises a number of the second plurality of nucleic acid molecules containing the genetic locus.
11. The method of any of claims 1-10, further comprising enriching the first plurality of nucleic acid molecules and/or the second plurality of nucleic acid molecules for at least a portion of the plurality of genetic loci.
12. The method of claim 11, wherein the enrichment comprises amplifying at least a portion of the first plurality of nucleic acid molecules and/or the second plurality of nucleic acid molecules.
13. The method of claim 12, wherein the amplification comprises selective amplification or universal amplification.
14. The method of claim 11, wherein the enrichment comprises selectively isolating at least a portion of the first plurality of nucleic acid molecules and/or the second plurality of nucleic acid molecules.
15. The method of any of claims 1-2, wherein the plurality of genetic loci comprises at least about 50 or at least about 100 distinct autosomal single nucleotide polymorphisms (SNPs).
16. The method of any of claims 1-2, wherein generating the first sample fingerprint further comprises obtaining a third biological sample comprising a third plurality of nucleic acid molecules from the subject, and processing the third plurality of nucleic acid molecules to obtain

a quantitative measure of the third plurality of nucleic acid molecules at each of a second plurality of genetic loci, wherein the second plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); and wherein generating the second sample fingerprint further comprises obtaining a fourth biological sample comprising a fourth plurality of nucleic acid molecules from the subject, and processing the fourth plurality of nucleic acid molecules to obtain a quantitative measure of the fourth plurality of nucleic acid molecules at each of the second plurality of genetic loci.

17. The method of claim 16, wherein generating the first sample fingerprint further comprises obtaining a fifth biological sample comprising a fifth plurality of nucleic acid molecules from the subject, and processing the fifth plurality of nucleic acid molecules to obtain a quantitative measure of the fifth plurality of nucleic acid molecules at each of a third plurality of genetic loci, wherein the third plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs); and wherein generating the second sample fingerprint further comprises obtaining a sixth biological sample comprising a sixth plurality of nucleic acid molecules from the subject, and processing the sixth plurality of nucleic acid molecules to obtain a quantitative measure of the sixth plurality of nucleic acid molecules at each of the third plurality of genetic loci.

18. The method of any of claims 1-17, comprising identifying the sample mismatch with a sensitivity or specificity of at least about 90%.

19. The method of any of claims 1-17, comprising identifying the sample mismatch with a positive predictive value (PPV) of at least about 90%, a negative predictive value (NPV) of at least about 90%, or an area under the curve (AUC) of at least about 0.90.

20. The method of any of claims 1-17, wherein the predetermined threshold is that the difference comprises a difference in genotype similarity greater than a predetermined threshold.

21. The method of claim 20, wherein the predetermined threshold is about 0.8.

22. The method of any of claims 1-21, further comprising excluding the second biological sample from further assaying based on the identified sample mismatch.

23. The method of any of claims 1-2, further comprising identifying a sample match when the difference between the first sample fingerprint and the second sample fingerprint does not satisfy the predetermined threshold.

24. The method of claim 23, comprising identifying the sample match with a sensitivity of at least about 90%, a specificity of at least about 90%, a positive predictive value (PPV) of at least about 90%, a negative predictive value (NPV) of at least about 90%, or an area under the curve (AUC) of at least about 0.90.

25. The method of any of claims 23-24, further comprising: (a) subjecting the second biological sample to further assaying based on the identified sample match; or (b) based on the identified sample match, storing the second sample fingerprint in a database, and optionally, storing the first sample fingerprint in the database.
26. The method of claim 2, wherein the quantitative measure of the first plurality of nucleic acid molecules comprises no more than twelve independent measurements of the first plurality of nucleic acid molecules.
27. The method of claim 2, wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a predetermined threshold.
28. A non-transitory computer-readable medium comprising machine-executable code that, upon execution by one or more computer processors, implements a method for identifying a sample mismatch, comprising:
- receiving information of a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules of a first biological sample at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms;
  - receiving information of a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules of a second biological sample at each of the plurality of genetic loci, wherein the second biological sample is obtained from the subject;
  - determining a difference between the first sample fingerprint and the second sample fingerprint; and
  - identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint satisfies a predetermined threshold.
29. A system, comprising:
- one or more processors;
  - a non-transitory computer-readable medium comprising machine-executable code that, upon execution by the one or more processors, implements a method for identifying a sample mismatch, comprising:
- receiving information of a first sample fingerprint comprising a quantitative measure of a first plurality of nucleic acid molecules of a first biological sample at each of a plurality of genetic loci, wherein the plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms;

receiving information of a second sample fingerprint comprising a quantitative measure of a second plurality of nucleic acid molecules of a second biological sample at each of the plurality of genetic loci, wherein the second biological sample is obtained from the subject;

determining a difference between the first sample fingerprint and the second sample fingerprint; and

identifying the sample mismatch when the difference between the first sample fingerprint and the second sample fingerprint satisfies a predetermined threshold.