

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), 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 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;

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, 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, 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;

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 that have only two alleles that have no insertions or deletions.

[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

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 first sample fingerprint comprises a first plurality of sequencing reads obtained by sequencing the first plurality of nucleic acid molecules, 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 plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms that have only two alleles that have no insertions or deletions;

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 sample fingerprint comprises a second plurality of sequencing reads obtained by sequencing the second plurality of nucleic acid molecules, 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, and 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

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 first sample fingerprint comprises a first plurality of sequencing reads obtained by sequencing the first plurality of nucleic acid molecules, 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 plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms that have only two alleles that have no insertions or deletions;

- 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 sample fingerprint comprises a second plurality of sequencing reads obtained by sequencing the second plurality of nucleic acid molecules, 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, and 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), 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 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;
 - 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, 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, 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;
 - 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 that have only two alleles that have no insertions or deletions.
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 1-4, 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.
6. The method of claim 5, 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.
7. The method of claim 6, wherein the amplification comprises selective amplification or universal amplification.
8. The method of claim 5, 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.
9. 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).
10. 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.
11. The method of claim 10, 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.

12. The method of any of claims 1-11, comprising identifying the sample mismatch with a sensitivity or specificity of at least about 90%.
13. The method of any of claims 1-11, 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.
14. The method of any of claims 1-11, wherein the predetermined threshold is that the difference comprises a difference in genotype similarity greater than a predetermined threshold.
15. The method of claim 14, wherein the predetermined threshold is about 0.8.
16. The method of any of claims 1-15, further comprising excluding the second biological sample from further assaying based on the identified sample mismatch.
17. 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.
18. The method of claim 17, 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.
19. The method of any of claims 17-18, 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.
20. 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.
21. The method of claim 2, wherein the autosomal single nucleotide polymorphisms have a minor allele fraction that exceeds a predetermined threshold.
22. 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 first sample fingerprint comprises a first plurality of sequencing reads obtained by sequencing the first plurality of nucleic acid molecules, 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 plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms that have only two alleles that have no insertions or deletions;

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 sample fingerprint comprises a second plurality of sequencing reads obtained by sequencing the second plurality of nucleic acid molecules, 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, and 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.

23. 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 first sample fingerprint comprises a first plurality of sequencing reads obtained by sequencing the first plurality of nucleic acid molecules, 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 plurality of genetic loci comprises autosomal single nucleotide polymorphisms (SNPs) that comprise simple single nucleotide polymorphisms that have only two alleles that have no insertions or deletions;

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 sample fingerprint comprises a second plurality of sequencing reads obtained by sequencing the second plurality of nucleic acid molecules, 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, and 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.