Claims

What is claimed is:

Claim 1. A genomics method for constructing and using a chimeric reference standard specific to an individual subject for improving accuracy of analysis of a plurality of portions of a genome of the subject, the method comprising:

comparing a first portion of the genome of the subject to each reference standard of a plurality of reference standards stored in a database, wherein the genome of the subject is based on a biological sample obtained from the subject;

selecting a portion of a first reference standard that satisfies a predetermined matching threshold to a first portion of the genome of the subject based on the comparison of the first portion to the plurality of reference standards;

comparing a second portion of the genome of the subject to each reference standard of the plurality of reference standards;

selecting a portion of a second reference standard that satisfies a predetermined matching threshold to the second portion of the genome of the subject based on the comparison of the second portion to the plurality of references standards;

wherein the first portion of the genome is a portion of a first chromosome of the subject's DNA and the second portion of the genome is a portion of a second chromosome of the subject's DNA, or wherein the first portion of the genome is a portion of one strand of the subject's DNA and the second portion of the genome is a portion of the other strand of the subject's DNA;

creating a chimeric reference standard, whereby the creating comprises inserting the selected portions of the first reference standard and the second reference standard into respective portions of a hash table maintained in one or more memory devices; and

using the hash table and the genome of the subject in order to perform one or more of a mapping, aligning, and/or variant calling procedure.

Claim 2. The method of claim 1, wherein the first portion of the genome is a portion of a first chromosome of the subject’s DNA and the second portion of the genome is a portion of a second chromosome of the subject’s DNA, and wherein the first chromosome and the second chromosome are homologous chromosomes.

Claim 3. The method of claim 1 or claim 2, wherein each reference standard of the plurality of reference standards corresponds to i) one or more genomes specific to a

geographical place, ii) one or more genomes from one or more members of the subject's family, or (iii) one or more genomes from one or more of the subject's ancestors.

Claim 4. The method of any one of claim 1 through 3, wherein producing the reference standard comprises producing the chimeric reference standard in graph format.

Claim 5. The method of claim 4, wherein variations of the chimeric reference standard from one or both of the first reference standard and the second reference standard are represented by bubbles.

Claim 6. The method of claim 4, wherein variations of the chimeric reference standard from one or both of the first reference standard and the second reference standard are represented by branches.

Claim 7. The method of any one of claim 1 to 6, wherein producing the chimeric reference standard comprises performing a multi-region joint detection process.

Claim 8. The method of any one of claim 1 to 7, wherein the matching comprises a mapping procedure.

Claim 9. The method of claim 8, wherein the mapping procedure comprises an alt-aware type mapping procedure.

Claim 10. The method of any one of claim 1 to 7, wherein the matching comprises an aligning procedure.

Claim 11. The method of any one of claim 1 to 7, wherein the matching comprises a variant calling procedure.

Claim 12. The method of claim 11, wherein the matching comprises a Genome Analysis Tool Kit (GATK) haplotype caller procedure, a Hidden Markov Model (HHM) tool procedure, or a De Bruijn Graph function procedure.

Claim 13. The method of any one of claims 1 to 12, wherein the first portion and/or the second portion comprise(s) at least 1 or 2 million base pairs.

Claim 14 A non-transitory computer-readable medium storing software comprising instructions executable by one or more computers which, upon such execution, cause the one or more computers to perform operations for improving accuracy of analysis of a plurality of portions of a genome of a subject, the operations comprising the method of any one of claims 1 to 13.

Claim 15. A system for improving accuracy of nucleic acid sequence analysis of nucleic acid sequences from a subject, the system comprising:

one or more computers and one or more storage devices storing instructions that are operable, when executed by the one or more computers, to cause the one or more computers to perform operations comprising the method of any one of claims 1 to 13.