WHAT IS CLAIMED IS:

1. A method for determining a variable number tandem repeat (VNTR) status comprising:

under control of a hardware processor:

receiving a plurality of long sequence reads generated from a plurality of first samples obtained from a plurality of first subjects;

determining a plurality of haplotypes of a VNTR using long sequence reads of the plurality of long sequence reads aligned to the VNTR in a reference;

receiving a plurality of short sequence reads generated from a second sample obtained from a second subject;

for each of the plurality of haplotypes of the VNTR, realigning short sequence reads, of the plurality of short sequence reads aligned to the VNTR, to the haplotype to generate a realignment;

determining a probability indication of each of the plurality of haplotypes of the VNTR for the second subject using the realignment of the short sequence reads realigned to the haplotype; and

determining a status of the VNTR of the second subject based on the probability indications of each of the plurality of haplotypes.

2. The method of claim 1, wherein determining the plurality of haplotypes of the VNTR comprises building a database comprising the plurality of haplotypes of the VNTR.

3. The method of any one of claims 1-2, wherein determining the plurality of haplotypes of the VNTR comprises:

for each of the plurality of first samples:

extracting the long sequence reads of the plurality of long sequence reads of the first sample aligned to the VNTR in the reference;

realigning the long sequence reads extracted to a left flanking region and a right flanking region of the VNTR to determine aligned long sequence reads; and

determining a haplotype of the plurality of haplotypes based on the aligned long sequence reads each with an alignment score above an alignment threshold.

4. The method of claim 3, wherein at least one of the long sequence reads of the plurality of long sequence reads of the first sample is aligned to the VNTR and/or realigned to the left flanking region and the right flanking region span the VNTR.

5. The method of any one of claims 3-4, wherein determining the haplotype of the plurality of haplotypes of the VNTR comprises:

trimming sequences, of the aligned long sequence reads each with the alignment score above the alignment threshold, aligned to the left flanking region and the right flanking region to generate trimmed long sequence reads; and

determining the haplotype of the plurality of haplotypes based on the trimmed long sequence reads.

6. The method of any one of claims 3-5, wherein the first sample is homozygous for the VNTR, and wherein determining the haplotype of the plurality of haplotypes comprises: determining only one haplotype of the plurality of haplotypes based on the trimmed long sequence reads.

7. The method of claim 6, wherein determining the only one haplotype comprises:

clustering the trimmed long sequence reads into only one cluster, optionally wherein clustering the trimmed long sequence reads into the only one cluster comprises clustering the trimmed long sequence reads into the only one cluster based on lengths of the trimmed long sequence reads, optionally wherein the clustering comprises k-means clustering; and

determining the only one haplotype based on the trimmed long sequence reads.

8. The method of any one of claims 3-5, wherein the first sample is heterozygous for the VNTR, and wherein determining the haplotype of the plurality of haplotypes comprises: determining two haplotypes of the plurality of haplotypes of the VNTR based on the trimmed long sequence reads.

9. The method of claim 8, wherein determining the two haplotypes comprises:

clustering the trimmed long sequence reads into two clusters, optionally wherein clustering the trimmed long sequence reads into the two clusters comprises clustering the trimmed long sequence reads into the two clusters based on lengths of the trimmed long sequence reads, optionally wherein the clustering comprises k-means clustering;

determining a first haplotype of the two haplotypes based on the trimmed long sequence reads in a first cluster of the two clusters; and

determining a second haplotype of the two haplotypes based on the trimmed long sequence reads in a second cluster of the two clusters.

10. The method of claim 9, wherein the trimmed long sequence reads comprises a first plurality of trimmed long sequence reads and a second plurality of trimmed long sequence reads with different lengths, optionally wherein the different lengths differ by at least 5,000 base pairs, wherein the first cluster comprises all, substantially all, or a majority of the first plurality

of trimmed long sequence reads, and wherein the second cluster comprises all or majority of the second plurality of trimmed long sequence reads.

11. A system for determining a variable number tandem repeat (VNTR) status comprising:

non-transitory memory configured to store executable instructions and a plurality of haplotypes of a VNTR; and

a hardware processor in communication with the non-transitory memory, the hardware processor programmed by the executable instructions to perform:

receiving a plurality of short sequence reads generated from a test sample obtained from a test subject;

for each of the plurality of haplotypes of the VNTR, realigning short sequence reads, of the plurality of short sequence reads aligned to the VNTR, to the haplotype to generate a realignment;

determining a probability of each of the plurality of haplotypes for the test subject using the realignment of the short sequence reads realigned to the haplotype; and

determining a status of the VNTR of the test subject.

12. The system of claim 11, wherein the plurality of haplotypes of the VNTR is determined using long sequence reads of a plurality of long sequence reads aligned to the VNTR in a reference, and wherein the plurality of long sequence reads is generated from a plurality of reference samples obtained from a plurality of reference subjects.

13. The system of claim 12, wherein the plurality of haplotypes of the VNTR is determined by:

for each of the plurality of samples:

extracting the long sequence reads of the plurality of long sequence reads of the test sample aligned to the VNTR in the reference;

realigning the long sequence reads extracted to a left flanking region and a right flanking region of the VNTR to determine aligned long sequence reads; and

determining a haplotype of the plurality of haplotypes based on the aligned long sequence reads each with an alignment score above an alignment threshold.

14. The system of claim 13, wherein at least one of the long sequence reads of the plurality of long sequence reads of the test sample is aligned to the VNTR and/or realigned to the left flanking region and the right flanking region span the VNTR.

15. The system of any one of claims 13-14, wherein the haplotype of the plurality of haplotypes of the VNTR is determined by:

trimming sequences, of the aligned long sequence reads each with the alignment score above the alignment threshold, aligned to the left flanking region and the right flanking region to generate trimmed long sequence reads; and

determining the haplotype of the plurality of haplotypes based on the trimmed long sequence reads.

16. The system of any one of claims 13-15, wherein the reference sample is homozygous for the VNTR, and the haplotype of the plurality of haplotypes of the VNTR is determined to comprise only one haplotype of the plurality of haplotypes based on the trimmed long sequence reads.

17. The system of claim 16, wherein the only one haplotype is determined by:

clustering the trimmed long sequence reads into only one cluster, optionally wherein clustering the trimmed long sequence reads into the only one cluster comprises clustering the trimmed long sequence reads into the only one cluster based on lengths of the trimmed long sequence reads, optionally wherein the clustering comprises k-means clustering; and

determining the only one haplotype based on the trimmed long sequence reads.

18. The system of any one of claims 13-15, wherein the reference sample is heterozygous for the VNTR and wherein the haplotype of the plurality of haplotypes of the VNTR is determined to comprise two haplotypes of the plurality of haplotypes based on the trimmed long sequence reads.

19. The system of claim 18, wherein the two haplotypes is determined by:

clustering the trimmed long sequence reads into two clusters, optionally wherein clustering the trimmed long sequence reads into the two clusters comprises clustering the trimmed long sequence reads into the two clusters based on lengths of the trimmed long sequence reads, optionally wherein the clustering comprises k-means clustering;

determining a first haplotype of the two haplotypes based on the trimmed long sequence reads in a first cluster of the two clusters; and

determining a second haplotype of the two haplotypes based on the trimmed long sequence reads in a second cluster of the two clusters.

20. The system of claim 19, wherein the trimmed long sequence reads comprises a first plurality of trimmed long sequence reads and a second plurality of trimmed long sequence reads with different lengths, optionally wherein the different lengths differ by at least 5,000 base pairs, wherein the first cluster comprises all, substantially all, or a majority of the first plurality

of trimmed long sequence reads, and wherein the second cluster comprises all, substantially all, or a majority of the second plurality of trimmed long sequence reads.