CLAIMS:

[Claim 1]. A system for determining microsatellite instability, comprising:

a processor; and

a memory storing instructions that, when executed by the processor, cause the processor to:

access genomic sequence data of a sample of interest, the sample of interest being derived from a tumor sample for which a matched normal sample is unavailable, wherein the sequence data comprises nucleotide identity information for a plurality of microsatellite regions;

receive sample information related to the sample of interest;

select an associated reference sample dataset from a plurality of reference sample datasets based on the sample information, wherein the associated reference sample dataset is an unmatched normal cohort generated from nucleotide identity information for the plurality of microsatellite regions and from a plurality of individuals, wherein the nucleotide identity information for an individual microsatellite region of the plurality of microsatellite regions and from an individual of the plurality of individuals is subjected to a quality metric based on a sequencing depth at the individual microsatellite region, wherein the individual microsatellite region is included in the associated reference sample dataset upon passing the quality metric;

classify microsatellite instability for the sample of interest based on a comparison of the sequence data from the sample of interest to the associated reference sample dataset; and

provide an indication representative of microsatellite instability of the sample of interest based on the classification.

[Claim 2]. The system of claim 1, wherein the sample information comprises sample of interest origin information, wherein the plurality of reference sample datasets differ from one another based on origin, and wherein the associated reference sample dataset is selected based on a match between the sample of interest origin information and the origin of the associated reference sample dataset.

[Claim 3]. The system of claim 1 or claim 2, wherein the associated reference sample dataset is generated from formalin fixed paraffin embedded samples from a plurality of individuals and the sample of interest is a formalin fixed paraffin embedded sample.

[Claim 4]. The system of claim 1 or claim 2, wherein the associated reference sample dataset is generated from fresh frozen samples from a plurality of individuals and the sample of interest is a fresh frozen sample.

[Claim 5]. The system of claim 1 or claim 2, wherein the associated reference sample dataset is generated from cell lines from a plurality of individuals and the sample of interest is a cell line.

[Claim 6]. The system of any one of claims 1 to 5, wherein the sample information comprises tissue type information, wherein the plurality of reference sample datasets differ from one another based on tissue type, and wherein the associated reference sample dataset is further selected based on a match between the tissue type information and the tissue type of the associated reference sample dataset.

[Claim 7]. The system of any one of claims 1 to 6, wherein the sample information comprises sequencing panel information used to generate the sequence data, wherein the plurality of reference sample datasets differ from one another based on a sequencing panel used to generate the reference sample datasets, and wherein the

associated reference sample dataset is further selected based on a match between the sequencing panel information and the sequencing panel used to generated the associated reference sample dataset.

[Claim 8]. The system of any one of claims 1 to 7, wherein the associated reference sample dataset is a pooled dataset from the plurality of individuals.

[Claim 9]. The system of any one of claims 1 to 8, wherein the unmatched normal cohort is a universal matched sample.

[Claim 10]. A computer-implemented method for detecting microsatellite instability in a sample of interest, comprising:

providing sequence data of the sample of interest, the sample of interest being derived from a tumor sample for which a matched normal sample is unavailable, wherein the sequence data comprises nucleotide identity information for a plurality of microsatellite regions;

providing sample information related to the sample of interest;

selecting an associated reference sample dataset from a plurality of reference sample datasets based on the sample information, wherein the associated reference sample dataset is an unmatched normal cohort generated from nucleotide identity information for the plurality of microsatellite regions and from a plurality of individuals, wherein the nucleotide identity information for an individual microsatellite region of the plurality of microsatellite regions and from an individual of the plurality of individuals is subjected to a quality metric based on a sequencing depth at the individual microsatellite region, wherein the individual microsatellite region is included in the associated reference sample dataset upon passing the quality metric; and

assessing microsatellite instability of the sample of interest based on a computer-implemented comparison of the sequence data from the sample of interest to the associated reference sample dataset.

[Claim 11]. The method of claim 10, wherein the method uses a system according to any one of claims 1 to 9.

[Claim 12]. Use of a system according to any one of claims 1 to 9 to detect microsatellite instability in a sample of interest and/or to determine a microsatellite instability type in the sample of interest, wherein the sample has been derived from a tumor sample for which a matched normal sample is unavailable.