CLAIMS

1. A method of determining a tumor mutational burden (TMB) in a subject, the method comprising:

(a) determining an observed mutational count from sequence information obtained from one or more nucleic acids in a sample from the subject;

(b) determining a tumor fraction and/or a coverage of the nucleic acids to generate sequencing parameters;

(c) determining an expected mutational fraction and/or an expected distribution of the expected mutational fraction given the sequencing parameters to generate an expected result, wherein determining the expected mutation fraction and/or the expected distribution of the expected mutational fraction comprises calculating a probability of identifying a mutation in a given mutant allele fraction (MAF) across a distribution of expected MAFs, wherein the distribution of expected MAFs is determined from an expected distribution of relative MAFs; and

(d) adjusting the observed mutational count given the expected result to generate an adjusted result, thereby determining the TMB in the subject.

2. The method of claim 1, wherein the observed mutational count:

(a) comprises a number of synonymous mutations, a number of nonsynonymous mutations, and/or a number of non-coding mutations identified in the nucleic acids;

(b) comprises a number of mutations selected from the group consisting of: single nucleotide variants (SNVs), insertions or deletions (indels), copy number variants (CNVs), fusions, transversions, translocations, frame shifts, duplications, repeat expansions, and epigenetic variants;

(c) excludes driver mutations and/or clonal hematopoiesis-derived mutations; and/or

(d) comprises a number of somatic mutations identified in the nucleic acids, optionally wherein the observed mutational count excludes one or more known cancer driver and/or passenger mutations.

3. The method of any one of the preceding claims, comprising:

(a) using pooled evidence of one or more likely mutations that are below a limit of detection for a given single nucleotide variant (SNV) or a given insertion or deletion (indel) to determine the observed mutational count;

(b) using the expected mutational fraction as an observed fraction of an actual mutational count;

(c) comparing the sequence information with one or more reference sequences to identify the observed mutational count, optionally wherein the reference sequences comprise at least subsequences of hg19 and/or hg38; and/or

(d) further comprising determining a TMB score using the adjusted result.

4. The method of any one of the preceding claims, wherein the tumor fraction:

(a) comprises a maximum MAF of all somatic mutations identified in the nucleic acids; and/or

(b) is below about 0.05%, about 0.1 %, about 0.2%, about 0.5%, about 1%, about 2%, about 3%, about 4%, or about 5% of all nucleic acids in the sample.

5. The method of any one of the preceding claims:

(a) comprising identifying a number of unique cfDNA fragments comprising a given nucleotide position in the nucleic acids to determine the coverage, and/or identifying a median number of unique cell-free DNA (cfDNA) molecules comprising a given nucleotide position in the nucleic acids to determine the coverage;

(b) wherein the coverage is between 10 and 50,000 cfDNA fragments at a given nucleotide position in the nucleic acids present in the sample;

(c) wherein the expected mutational fraction and/or the expected distribution of the expected mutational fraction comprises about a 95% or more confidence interval for the mutational fraction; and/or

(d) comprising using an upper bound of a 95% confidence interval of the expected mutational fraction to generate a lower bound of the observed mutational count.

6. The method of any one of the preceding claims:

(a) comprising determining the expected result using the equation of:

$$Fraction\ of\ mutations\ observed = \sum_{MAF} (P(call\ a\ mutation \mid MAF) * P\ (mutation\ at\ MAF))$$

where *P* is probability, and *MAF* is the mutant allele fraction;

(b) comprising multiplying the distribution of expected relative MAFs by the tumor fraction to generate MAFs; and/or

(c) wherein the distribution of expected MAFs is calculated using a binomial proportion confidence interval of:

$$f_{upper\ bound} = f + z\ * \sqrt{(f\ *\ (1 - f)\ /\ n_true)}, \text{ and}$$

$$f_{lower\ bound} = f - z\ * \sqrt{(f\ *\ (1 - f)\ /\ n_true)},$$

where *f* is the expected fraction of mutations called, *n_true* is the expected actual mutations, which is equal to the number of mutations observed given *f*, and *z* is the confidence level, optionally wherein:

(i) if *f* is less than $threshold_f$, then TMB is not determined, or

(ii) if *f* is greater than $threshold_f$, then TMB is determined.

7. The method of any one of the preceding claims, wherein the expected distribution of the relative MAFs is obtained from one or more datasets of control sample dataset, optionally:

(a) wherein the control sample dataset comprises at least about 25, at least about 50, at least about 100, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1,000, at least about 5,000, at least about 10,000, at least about 15,000, at least about 20,000, at least about 25,000, at least about 30,000, or more control samples;

(b) wherein the max MAFs observed in the control sample dataset comprises about 0.5%, about 1%, about 2%, about 5% or about 10%; and/or

(c) comprising fitting the relative MAFs to a curve using the equation of:

$$F = 1/(1 + (P_50/relative\text{-}MAF)^n),$$

where *F* is the cumulative distribution function, *P_50* is the median relative MAF, *relative-MAF* is relative MAF, and *n* is the exponent to fit the shape of the relative distribution.

8. The method of any one of the preceding claims, comprising dividing the observed mutational count:

(a) by expected mutation fraction or an upper/ lower bound of a confidence interval of the expected mutational fraction in the sample to generate the adjusted result; and/or

(b) by the expected result to generate the adjusted result.

9. The method of any one of the preceding claims, wherein the adjusted result comprises:

(a) a number of mutations detected in the nucleic acids over a range of mutant allele fractions;

(b) a prediction of the highest-likely actual mutational count;

(c) a prediction of the lowest-likely actual mutational count; and/or

(d) an adjusted mutational count, optionally wherein the adjusted mutational count is greater than or equal to the observed mutational count.

10. The method of any one of the preceding claims, optionally:

(a) wherein the sample:

(i) is selected from the group consisting of: tissue, blood, plasma, serum, sputum, urine, semen, vaginal fluid, feces, synovial fluid, spinal fluid, and saliva;

(ii) comprises tissue; or

(iii) comprises blood, plasma, and/or serum;

(b) wherein the subject is a mammalian subject, optionally wherein the mammalian subject is a human subject;

(c) wherein the nucleic acids:

(i) comprise cell-free nucleic acids;

(ii) comprise cellular nucleic acids;

(iii) comprise circulating tumor nucleic acids;

(iv) are obtained from circulating tumor cells; and/or

(v) comprise deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA); and/or

(d) comprising amplifying at least one segment of the nucleic acids in the sample to generate at least one amplified nucleic acid, optionally comprising sequencing the amplified nucleic acid to generate the sequence information, further optionally:

(i) sequencing at least about 50,000, about 100,000, about 150,000, about 200,000, about 250,000, about 500,000, about 750,000, about 1,000,000, about 1,500,000, about 2,000,000, or more nucleotides of the nucleic acids to generate the sequence information; and/or

(ii) wherein the sequencing is selected from the group consisting of: targeted sequencing, intron sequencing, exome sequencing, and whole genome sequencing.

11. The method of any one of the preceding claims:

(a) wherein the sequence information is obtained by:

(i) amplifying nucleic acids in the sample to generate amplified nucleic acids; and

(ii) sequencing the amplified nucleic acids to generate the sequence information;

(b) comprising obtaining the sequence information from the nucleic acids in the sample from the subject, optionally wherein the sequence information is obtained from targeted segments of the nucleic acids, further optionally wherein the targeted segments comprise between about 1 and about 100,000 different and/or overlapping genomic regions;

(c) wherein the sequence information comprises sequence reads of the nucleic acids generated by a nucleic acid sequencer, optionally wherein the nucleic acid sequencer performs pyrosequencing, single-molecule sequencing, nanopore sequencing,

semiconductor sequencing, sequencing-by-synthesis, sequencing-by-ligation or sequencing-by hybridization on the nucleic acids to generate sequencing reads;

(d) further comprising selectively enriching one or more regions from a genome or transcriptome prior to sequencing, optionally further comprising amplifying the selectively enriched regions prior to sequencing; and/or

(e) wherein the sequence information is obtained from targeted segments of the nucleic acids, wherein the targeted segments are obtained by selectively enriching one or more regions from a genome or transcriptome prior to sequencing, optionally further comprising amplifying the obtained targeted segments prior to sequencing.

12. The method of any one of the preceding claims, further comprising attaching one or more adapters comprising molecular barcodes to the nucleic acids prior to sequencing, optionally:

(a) wherein the nucleic acids are uniquely tagged;

(b) wherein the nucleic acids are non-uniquely tagged;

(c) wherein the adapters comprise between 2 and 1,000,000 molecular barcodes;

(d) comprising randomly attaching an adapter comprising a molecular barcode to each end of a nucleic acid;

(e) wherein the adapters are attached to the nucleic acids by blunt end ligation or sticky end ligation; or

(f) wherein the adapters are T-tailed and/or C-tailed adapters.

13. The method of any one of the preceding claims,

(a) further comprising grouping the sequence reads into families of sequence reads, each family comprising sequence reads generated from a nucleic acid in the sample;

(b) wherein at least a portion of said method is computer implemented;

(c) further comprising generating a report in an electronic format which provides one or more TMB scores;

(d) further comprising classifying the sample as a TMB-High sample;

(e) further comprising classifying the sample as a TMB-High sample if the TMB score of the sample is greater than a $\text{threshold}_{\text{TMB score}}$ and classifying the sample as a TMB-Low sample if the TMB score of the sample is lower than the $\text{threshold}_{\text{TMB score}}$; and/or

(f) wherein the mutational count is determined in a set of genes or genomic regions comprising the genes listed in Table 2.

14. The method of any one of the preceding claims, further comprising:

(a) identifying one or more clonotypes of an immune repertoire in at least the second portion of the sequence information; and,

(b) correlating the TMB score with the one or more clonotypes to identify one or more neoantigen-orphan immune receptors in the subject.

15. The method of any one of the preceding claims, further comprising reporting the TMB score to the subject or physician.