TAXONOMY-INDEPENDENT CANCER DIAGNOSTICS AND CLASSIFICATION USING MICROBIAL NUCLEIC ACIDS AND SOMATIC MUTATIONS

CROSS-REFERENCE

[0001] This application claims benefit of U.S. Provisional Patent Application No. 63/128,971 filed December 22, 2020, which is entirely incorporated by reference.

BACKGROUND

[0002] An ideal diagnostic test for the detection of cancer in a subject would have the following characteristics: (i) it should identify, with high confidence, the tissue/body site location(s) of the cancer; (ii) it should identify the presence of somatic mutations that account for or are tightly associated with the cancerous state; (iii) it should detect the occurrence of cancer early (e.g., Stages I – II) to enable early-stage medical intervention; (iv) it should be minimally invasive; and (vi) it should be both highly sensitive and specific with respect to the cancer being diagnosed (i.e., there should be a high probability that the test will be positive when the cancer is present and a high probability that the test will be negative when the cancer is not present). Today, liquid biopsy-based diagnostics—both commercialized and in development—fall into two broad, non-overlapping categories – those that can detect cancer-associated somatic mutations and those that can detect the tissue/body site location of a cancer on the basis of tissue-unique molecular patterns, such as DNA methylation. Neither category of existing diagnostics therefore provides the full complement of data that would otherwise tell a physician where to focus medical intervention and which medicaments should be selected.

[0003] Thus, there remains a need in the art for early-stage cancer diagnostics that can detect the tissue/body site location(s) of cancer with high analytic sensitivity and specificity while also determining somatic mutations associated with the detected cancer.

SUMMARY

[0003a] In one aspect, provided herein is a method of diagnosing cancer of a subject, comprising: (a) receiving a plurality of somatic mutation and non-human nucleic acid molecule k-mer sequences of a sample obtained from a subject; (b) comparing the plurality of somatic mutation and non-human nucleic acid molecule k-mer sequences of the subject with a plurality of somatic mutation and non-human nucleic acid molecule k-mer sequences for a given cancer; and (c) diagnosing cancer of the subject by providing a probability of the cancer based at least in part on the

comparison of the subject's plurality of somatic mutation and non-human nucleic acid molecule k-mer sequences and the plurality of somatic mutation and non-human nucleic acid molecule k-mer sequences for the given cancer.

[0004] The disclosure of the present invention provides a method to accurately diagnose cancer, its location, and predict a cancer's likelihood of responding to certain therapies, using nucleic acids of non-human origin from a human tissue or liquid biopsy sample in combination with identified human somatic mutations present in the sample. Specifically, the present invention provides methods for identifying the presence and abundance of cancer-associated nucleic acid sequence mutations in the human genome, the presence, and abundance of non-human nucleic