diagnose cancers while also providing information pertaining to the presence/abundance of biochemical capacities could help elucidate how intratumoral microbiota contribute to tumor-specific biology by either providing or consuming tumor required or produced metabolites, respectively.

[0004] Other prior art that is relevant to this field is as follows: U.S. Publication No. 2018/0223338 describes using the solid tissue microbiome or salvia microbiome in identifying and diagnosing head and neck cancer; and U.S. Publication No. 2018/0258495Al describes using the solid tissue microbiome or fecal microbiome to detect colon cancer, some kinds of mutations associated with colon cancer, and a kit to collect and amplify the corresponding microbes. PCT WO 2019/191649 describes using cell-free microbial DNA and machine learning models to distinguish subjects having advanced adenoma and/or colorectal cancer from healthy subjects, wherein the machine learning algorithms rely upon DNA sequence reads mapping to reference genomes as input for analysis.

SUMMARY

[0004a] In one aspect, provided herein is a method of determining the presence or lack thereof cancer of a subject, the method comprising: (a) providing one or more sequencing reads of a biological sample obtained from a subject; (b) filtering the sequencing reads with a genome database to produce a set of filtered non-human sequencing reads; (c) decontaminating the filtered non-human sequencing reads to remove contaminant non-human sequencing reads from the filtered non-human sequencing reads; (d) translating the decontaminated non-human sequencing reads to non-human proteins; (e) mapping the non-human proteins to a protein database, thereby producing a set of protein database associations; and (f) processing the set of protein database associations with a trained model thereby determining the presence of cancer of the subject.

[0005] The disclosure provided herein describes systems and methods capable of accurately diagnosing or determining the presence or lack thereof cancer and other diseases, its subtypes, and its likelihood to respond to certain therapies solely using nucleic acids of non-human origin from a tissue or liquid biopsy sample. Specifically, the present invention provides methods that may identify the presence and abundance of microbial functional genes (and fragments thereof) and biochemical pathways present in a biopsy sample (e.g., a liquid or tissue biopsy). In some cases, the microbial functional

genes and biochemical pathways may be utilized to train one or more models and/or predictive models, described elsewhere herein. Such trained models may output a determination of the presence or lack thereof a subject's cancer or the likelihood of therapeutic response and/or efficacy when a subject receives a treatment.

[0006] The methods of the present invention disclosed herein provide a method to generate a diagnostic model capable of diagnosing and classifying cancer whilst also providing information pertaining to the presence and or abundance of biochemical capacities to elucidate intratumoral microbiota contributions to tumor-specific biology. In some cases, tumor-specific biology may pertain to how intratumoral microbiota contribute to consuming tumor required or produced metabolites. For example, pathway-