CLAIMS

1. A method comprising:

determining first gene data associated with a plurality of genes;

determining second gene data associated with the plurality of genes, wherein the plurality of genes are sequenced from a plurality of tumor samples, wherein each tumor sample of the plurality of tumor samples is labeled as a responder or a non-responder;

determining, based on the first gene data and the second gene data, a plurality of features for a predictive model;

training, based on a first portion of the second gene data, the predictive model according to the plurality of features;

testing, based on a second portion of the second gene data, the predictive model; and

outputting, based on the testing, the predictive model.

2. The method of claim 1, wherein determining the first gene data associated with a plurality of genes comprises retrieving the first gene data from a public data source.

3. The method of any one of claims 1-2, wherein the plurality of genes comprise one or more of an immune cell type/function gene set, a tumor microenvironment component and signaling gene set, or a cancer cell proliferation and DNA repair gene set.

4. The method of any one of claims 1-3, wherein determining the first gene data associated with the plurality of genes comprises:

determining, based on the second gene data, the plurality of genes;

determining, based on the plurality of genes, one or more gene data sets that comprise at least one gene of the plurality of genes; and

generating, based on the one or more gene data sets, the first gene data.

5. The method of any one of claims 1-4, wherein the first gene data is comprised of gene data from a plurality of different gene data sets.

6. The method of any one of claims 1-5, wherein determining the second gene data associated with the plurality of genes comprises:

determining baseline gene expression levels for each tumor associated with the plurality of tumor samples;

treating each tumor associated with the plurality of tumor samples with a therapeutic;

determining, post-treatment, which tumors associated with the plurality of tumor samples are responders or non-responders to the therapeutic;

labeling the baseline gene expression levels for each tumor associated with the plurality of tumor samples, as responder or non-responder; and

generating, based on the labeled baseline gene expression levels, the second gene data.

7. The method of any one of claims 5-6, wherein determining, based on the first gene data and the second gene data, the plurality of features for the predictive model comprises:

determining, from the first gene data, genes present in two or more of the plurality of different gene data sets as a first set of candidate genes;

determining, from the second gene data, genes of the first set of candidate genes expressed at greater than or equal to 2 Transcripts Per Million (TPM) in at least half of the plurality of tumor samples as a second set of candidate genes; and

determining, from the second gene data, genes of the second set of candidate genes with a statistically significant increase in expression level between responders and non-responders as a third set of candidate genes,

wherein the plurality of features comprises the third set of candidate genes.

8. The method of claim 7, wherein determining, based on the first gene data and the second gene data, the plurality of features for the predictive model comprises:

determining, for the third set of candidate genes, a tumor mutational burden (TMB) value for each of the plurality of tumors associated with the third set of candidate genes; and

determining, based on the TMB values, a fourth set of candidate genes,

wherein the plurality of features comprises the fourth set of candidate genes.

9. The method of any one of claims 1-8, wherein training, based on the first portion of the second gene data, the predictive model according to the plurality of features results in determining a gene signature indicative of a responder.

10. The method of any one of claims 1-9, further comprising:

receiving baseline gene data associated with a plurality of genes for a subject, wherein the plurality of genes are sequenced from a tumor of the subject; and

providing, to the predictive model, the baseline gene data.

11. The method of claim 10, further comprising:

determining, based on the predictive model, that the subject is a candidate for a therapeutic treatment; and

treating the subject with the therapeutic treatment.

12. The method of claim 11, further comprising:

determining, based on the predictive model, that the subject is not a candidate for a therapeutic treatment; and

treating the subject with an alternative therapeutic treatment.

13. One or more non-transitory computer-readable media storing processor-executable instructions thereon that, when executed by a processor, cause the processor to perform the methods of any of claims 1-12.

14. A system comprising:

a computing device configured to perform the methods of any of claims 1-12; and

a predictive model configured to receive a plurality of features.

15. An apparatus comprising:

one or more processors; and

memory storing processor-executable instructions that, when executed by the one or more processors, cause the apparatus to perform the methods of any of claims 1-12.