Immunotherapy Methods for Patients Whose Tumors Carry a High Passenger Gene Mutation Burden

CROSS-REFERENCE TO RELATED APPLICATION

This application claims benefit of U.S. Provisional Application No. 62/560,955 filed September 20, 2017, and is hereby incorporated herein by reference in its entirety.

FIELD

The disclosure relates generally to the field of immunotherapy. More particularly, the disclosure relates to methods of administering an immunotherapy regimen to patients whose tumors have a high passenger gene mutation burden.

BACKGROUND

Various publications, including patents, patent applications, published patent applications, accession numbers, technical articles and scholarly articles are cited throughout the specification. Each of these cited publications is incorporated by reference, in its entirety and for all purposes, in this document.

Recent studies suggested that patients with higher overall tumor mutational burden (TMB) in their tumors are more likely to benefit from immunotherapy treatment due to the increase in neo-antigen presentation that could elicit an immune response. However, the overall mutational burden includes driver gene mutations that could actually suppress immunogenicity and decrease sensitivity to the treatment.

No existing method has been developed for the purpose of identifying passenger genes and their mutations to assess immunogenicity. These and other shortcomings are addressed in the present disclosure.

SUMMARY

It is to be understood that both the following general description and the following detailed description are exemplary and explanatory only and are not restrictive.

In a first aspect, the disclosure provides a method comprising receiving genetic sequence data, wherein the genetic sequence data comprises a plurality of genes and is

derived from a plurality of biological samples collected from subjects having a plurality of disease types, identifying a plurality of mutated genes for each of the plurality of biological samples, wherein each of the mutated genes comprises a genetic sequence having at least one non-synonymous somatic mutation, determining a tumor mutational burden for each biological sample based on a number of mutated genes in each biological sample, for each disease type, determining an average tumor mutational burden of the plurality of mutated genes in the plurality of biological samples based on the determined numbers of mutated genes in each biological sample, for each mutated gene and each disease type, determining a fraction of biological samples comprising the mutated gene, for each mutated gene, determining a correlation coefficient between the average tumor mutational burden and the fraction of biological samples comprising the mutated gene. A higher correlation coefficient indicates that a particular gene is more likely to acquire somatic mutations in the cancer types with higher overall mutation frequency (*e.g.*, passenger gene), whereas a lower correlation coefficient indicates that a particular gene is less likely to acquire somatic mutations in the cancer types with higher overall mutation frequency (*e.g.*, not a passenger gene).

In another aspect, the disclosure provides methods for selecting a cancer patient for immunotherapy. In general, the methods comprise establishing a total passenger gene mutation burden from a tumor of the cancer patient, generating a background distribution for the mutational burden of the tumor, normalizing the total passenger gene mutation burden against the background distribution, and categorizing the cancer patient as an immunotherapy responder when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution.

Generating a background distribution may comprise establishing the mutational burden from a plurality of samples of randomly selected genes obtained from the tumor, but the number of randomly selected genes in each sample preferably is equal to the number of passenger genes used to compute the total passenger gene mutation burden. Normalizing the total passenger gene mutation burden against the background distribution may comprise generating a z-score indicating the number of standard deviations from the mean of the background distribution.

The methods may further comprise categorizing a mutated gene in the tumor as a passenger gene. Categorizing a mutated gene in the tumor as a passenger gene may comprise selecting a mutated gene from the tumor and matching the mutated gene to a data structure comprising passenger genes established according to a passenger gene index. The passenger gene index may comprise a correlation coefficient between the fraction of samples comprising the mutated gene obtained from a cancer patient cohort and the median number of mutated genes in each type of tumor within the cancer patient cohort.

The methods may further comprise administering to the cancer patient an immunotherapy regimen. The immunotherapy regimen may comprise administering to the patient an inhibitor of a T cell inhibitory receptor. The immunotherapy regimen may comprise administering to the patient activator of T cell activating receptor.

The immunotherapy regimen may comprise administering to the patient an antibody that binds to PD1. The antibody that binds to PD1 may comprise at least the heavy chain variable region (HCVR) sequence of SEQ ID NO: 21 and a light chain variable region, or may comprise at least the light chain variable region (LCVR) sequence of SEQ ID NO: 22 and a heavy chain variable region. The antibody that binds to PD1 may comprise the HCVR or SEQ ID NO: 21 and the LCVR or SEQ ID NO: 22. That antibody that binds to PD1 may be administered in combination with an antibody that binds to LAG3.

The immunotherapy regimen may comprise administering to the patient an antibody that binds to PDL1. The antibody that binds to PDL1 may comprise at least the HCVR sequence of SEQ ID NO: 122 and a LCVR, or may comprise at least the LCVR sequence of SEQ ID NO: 123 and a HCVR. The antibody that binds to PDL1 may comprise the HCVR or SEQ ID NO: 122 and the LCVR or SEQ ID NO: 123. That antibody that binds to PDL1 may be administered in combination with an antibody that binds to LAG3.

The immunotherapy regimen may comprise administering to the patient an antibody that binds to LAG3. The antibody that binds to LAG3 may comprise at least the HCVR sequence of SEQ ID NO: 93 and a LCVR, or may comprise at least the LCVR sequence of SEQ ID NO: 94 and a HCVR. The antibody that binds to LAG3 may comprise the HCVR or SEQ ID NO: 93 and the LCVR or SEQ ID NO: 94. That antibody that binds to LAG3 may be

administered in combination with an antibody that binds to PD1 or with an antibody that binds to PDL1.

Additional advantages will be set forth in part in the description which follows or may be learned by practice. The advantages will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments and together with the description, serve to explain the principles of the methods and systems:

Figure 1 shows a flowchart illustrating an example method;

Figure 2 shows a flowchart illustrating an example method;

Figure 3 shows a flowchart illustrating another example method;

Figure 4 shows a flowchart illustrating another example method;

Figure 5 illustrates an overview of passenger gene characteristics;

Figure 6 shows scatter plots for the fraction of patients with the gene variant (y-axis) and average number of total mutated gene (x-axis) in each cancer type;

Figure 7 shows enrichment along the Passenger Gene Index (PGI) scale for cancer driver genes and various other gene groups;

Figure 8 shows the highest (left) and lowest (right) PGI CGC genes, and their corresponding cancer type with the highest percentage (>2%) of mutated sample;

Figures 9A-C are graphical representations of a) local immune cytolytic activities, b) TCR read count and c)clinical outcome of patient cohorts;

Figure 10 is a block diagram illustrating an exemplary operating environment for performing the disclosed methods;

Figure 11 is TMB of the patient cohort in phase 1 clinical study;

Figure 12 shows the top 500 passenger genes – highest Passenger Genes Index (PGI).

DETAILED DESCRIPTION

Various terms relating to aspects of disclosure are used throughout the specification and claims. Such terms are to be given their ordinary meaning in the art, unless otherwise indicated. Other specifically defined terms are to be construed in a manner consistent with the definition provided herein.

As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

Inhibiting comprises reducing, decreasing, blocking, preventing, delaying, inactivating, desensitizing, stopping, and/or downregulating activity or expression of a molecule or pathway of interest.

Embodiments of the methods and systems are described below with reference to block diagrams and flowchart illustrations of methods, systems, apparatuses and computer program products. It will be understood that each block of the block diagrams and flowchart illustrations, and combinations of blocks in the block diagrams and flowchart illustrations, respectively, can be implemented by computer program instructions. These computer program instructions may be loaded onto a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions which execute on the computer or other programmable data processing apparatus create a means for implementing the functions specified in the flowchart block or blocks.

These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture including computer-readable instructions for implementing the function specified in the flowchart block or blocks. The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions that execute on the computer or other programmable apparatus provide steps for implementing the functions specified in the flowchart block or blocks.

Accordingly, blocks of the block diagrams and flowchart illustrations support combinations of means for performing the specified functions, combinations of steps for performing the specified functions and program instruction means for performing the specified functions. It will also be understood that each block of the block diagrams and flowchart illustrations, and combinations of blocks in the block diagrams and flowchart illustrations, can be implemented by special purpose hardware-based computer systems that perform the specified functions or steps, or combinations of special purpose hardware and computer instructions.

The terms "subject" and "patient" are used interchangeably and include any animal. Mammals are preferred, including companion (e.g., cat, dog) and farm mammals (e.g., pig, horse, cow), as well as rodents, including mice, rabbits, and rats, guinea pigs, and other rodents. Non-human primates are more preferred, and human beings are highly preferred.

It has been observed in accordance with the disclosure that, in cancer, the total mutational burden of passenger genes, as opposed to the total mutational burden of all genes, serves as an accurate indicator of whether a cancer patient is likely to respond positively to immunotherapy. Tumor mutational burden (TMB) may refer to a number of mutations within the coding region of a tumor genome. Mutated genes were assessed and classified according to their status as a passenger gene by way of a passenger gene index, which was used as a metric to identify passenger genes from a large-scale cancer genome analysis. It was observed that identified passenger genes were enriched for gene families known for excessive passenger mutations, including genes encoding large proteins, genes with low expression level, and genes with late DNA replication time. The total mutational burden of passenger genes positively correlated with tumor immunogenicity and favorably predicted patient clinical outcomes. Accordingly, the disclosure features methods to classify patients according to their passenger gene mutation burden, as part of an immunotherapy regimen.

In cancer biology, driver mutations are understood to be at least casually implicated in cancer formation or cell transformation. And passenger mutations are understood to be those that do not confer a growth advantage or contribute to cancer development. *See* Stratton MR *et al.* (2009) Nature. 458:719-24. Thus, passenger genes include genes that

comprise passenger mutations. Non-limiting examples of mutations include substitutions, inversions, insertions, and deletions of one or more nucleotides, codons, genes, or chromosomes, as well as copy number variations.

In one aspect, the disclosure features methods and systems for identifying or classifying passenger genes. Identified passenger genes are enriched for families known for excessive passenger mutations, such as extremely large proteins and genes with low expression level or late DNA replication time. In some embodiments, passenger genes can be identified or classified according to a Passenger Gene Index (PGI). Thus, for example, passenger genes can be identified or classified according to a PGI that comprises a correlation coefficient between a fraction of samples obtained from a cancer patient cohort that comprises the mutated gene and the median number of mutated genes in each type of tumor within the cancer patient cohort. Based on identification of passenger genes, a data structure comprising passenger genes can be established.

Individual cancer patients can be screened to determine whether their tumors comprise passenger genes, as well as to determine the total passenger gene mutation burden of their tumor. Based on the patient's passenger gene mutation burden, the patient can be classified according to their capacity to respond positively to immunotherapy. Immunotherapy generally enhances the body's natural immune response to cancer, and includes, but is not limited to, the enhancement of the T cell response to the tumor.

An example of a methodology by which a cancer patient may be assessed for immunotherapy responsiveness is shown in **FIG. 1**. In general, the methods comprise establishing a total passenger gene mutation burden from a tumor of a cancer patient (**110**), generating a background distribution for the mutational burden of the tumor (**120**), normalizing the total passenger gene mutation burden against the background distribution (**130**), and categorizing the cancer patient as an immunotherapy responder (**140**).

Also disclosed are methods of treating a cancer patient with an immunotherapy after being assessed for immunotherapy responsiveness. For example, disclosed are methods of treating a cancer patient with an immunotherapy comprising determining if a cancer patient is an immunotherapy responder comprising establishing a total passenger gene mutation burden of the tumor of the patient; generating a background distribution for the mutational

burden of the tumor; normalizing the total passenger gene mutation burden against the background distribution; and categorizing the cancer patient as an immunotherapy responder genotype when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution; and administering an immunotherapy to the cancer patient categorized as an immunotherapy responder.

Further disclosed are methods for treating a patient with an inhibitor of a T cell inhibitory receptor or a receptor on a tumor cell or a non-immunotherapeutic treatment, wherein the patient is suffering from cancer, the method comprising the steps of: determining whether the patient is an immunotherapy responder by: obtaining or having obtained a biological sample from a tumor of the patient; performing or having performed a genotyping assay on the biological sample to determine if the patient has an immunotherapy responder genotype by, sequencing the biological sample to generate sequence data; establishing, based on the sequence data, a total passenger gene mutation burden of the tumor of the patient; generating, based on the sequence data, a background distribution for the mutational burden of the tumor; normalizing the total passenger gene mutation burden against the background distribution; and categorizing the patient as an immunotherapy responder genotype when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution; wherein if the patient has an immunotherapy responder genotype, then administering a therapeutically effective amount of an inhibitor of a T cell inhibitory receptor or a receptor on a tumor cell, wherein if the patient does not have an immunotherapy responder genotype, then administering a non-immunotherapeutic treatment. In some embodiments, a risk of unfavorable clinical outcome for a patient having an immunotherapy responder genotype is lower following the administration of the therapeutically effective amount of the inhibitor of a T cell inhibitory receptor or a receptor on a tumor cell than it would be if the patient were administered the non-immunotherapeutic treatment. In some embodiments, T cell activation and/or immune cytolytic activity in a patient having an immunotherapy responder genotype is higher following the administration of the therapeutically effective amount of the inhibitor of a T

cell inhibitory receptor or a receptor on a tumor cell than it would be if the patient were administered the non-immunotherapeutic treatment.

Disclosed are immunotherapies for use in the method of treating a cancer patient, the method comprising determining if a cancer patient is an immunotherapy responder by, establishing a total passenger gene mutation burden of a tumor of the patient; generating a background distribution for the mutational burden of the tumor; normalizing the total passenger gene mutation burden against the background distribution; categorizing the cancer patient as an immunotherapy responder genotype when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution; and administering the immunotherapy to the cancer patient categorized as an immunotherapy responder.

In some preferred embodiments, establishing a total passenger gene mutation burden from a tumor of a cancer patient (**110**) may comprise determining the total passenger gene mutation burden by any sequencing method that is used to determine the coding regions ("exome") of a tumor genome. Whole genome sequencing methods can also be used.

Exome mutations can be determined using sequencing methods known in the art. For example, US 2013/0040863, incorporated herein by reference, describes methods for determining the nucleic acid sequence of a target nucleic acid molecule, including sequencing by synthesis, sequencing by ligation or sequencing by hybridization, including for mutation detection, whole genome sequencing, and exon sequencing. If desired, various amplification methods can be used to generate larger quantities, particularly of limited nucleic acid samples, prior to sequencing.

Sequencing by synthesis (SBS) and sequencing by ligation can be performed using ePCR, as used by 454 Lifesciences (Branford, Conn.) and Roche Diagnostics (Basel, Switzerland). Nucleic acids such as genomic DNA or others of interest can be fragmented, dispersed in water/oil emulsions and diluted such that a single nucleic acid fragment is separated from others in an emulsion droplet. A bead, for example, containing multiple copies of a primer, can be used and amplification carried out such that each emulsion droplet serves as a reaction vessel for amplifying multiple copies of a single nucleic acid

fragment. Other methods can be used, such as bridging PCR (Illumina, Inc.; San Diego Calif.), or polony amplification (Agencourt/Applied Biosystems). US 2009/0088327; US 2010/0028885; and US 2009/0325172, each of which is incorporated herein by reference.

Methods for manual or automated sequencing are well known in the art and include, but are not limited to, Sanger sequencing, Pyrosequencing, sequencing by hybridization, sequencing by ligation and the like. Sequencing methods can be performed manually or using automated methods. Furthermore, the amplification methods set forth herein can be used to prepare nucleic acids for sequencing using commercially available methods such as automated Sanger sequencing (available from Applied Biosystems, Foster City, Calif.) or Pyrosequencing (available from 454 Lifesciences, Branford, Conn. and Roche Diagnostics, Basel, Switzerland); for sequencing by synthesis methods commercially available from Illumina, Inc. (San Diego, Calif.) or Helicos (Cambridge, Mass.) or sequencing by ligation methods being developed by Applied Biosystems in its Agencourt platform (see also Ronaghi et al., Science 281:363 (1998); Dressman et al., Proc. Natl. Acad. Sci. USA 100:8817-8822 (2003); Mitra et al., Proc. Natl. Acad. Sci. USA 100:55926-5931 (2003)), incorporated herein by reference,.

A population of nucleic acids in which a primer is hybridized to each nucleic acid such that the nucleic acids form templates and modification of the primer occurs in a template directed fashion. The modification can be detected to determine the sequence of the template. For example, the primers can be modified by extension using a polymerase and extension of the primers can be monitored under conditions that allow the identity and location of particular nucleotides to be determined. For example, extension can be monitored and sequence of the template nucleic acids determined using pyrosequencing, which is described in US 2005/0130173, US 2006/0134633, U.S. Pat. No. 4,971,903; U.S. Pat. No. 6,258,568 and U.S. Pat. No. 6,210,891, each of which is incorporated herein by reference, and is also commercially available. Extension can also be monitored according to addition of labeled nucleotide analogs by a polymerase, using methods described, for example, in U.S. Pat. No. 4,863,849; U.S. Pat. No. 5,302,509; U.S. Pat. No. 5,763,594; U.S. Pat. No. 5,798,210; U.S. Pat. No. 6,001,566; U.S. Pat. No. 6,664,079; U.S. 2005/0037398; and U.S. Pat. No. 7,057,026, each of which is incorporated herein by reference. Polymerases

useful in sequencing methods are typically polymerase enzymes derived from natural sources. It will be understood that polymerases can be modified to alter their specificity for modified nucleotides as described, for example, in WO 01/23411; U.S. Pat. No. 5,939,292; and WO 05/024010, each of which is incorporated herein by reference. Furthermore, polymerases need not be derived from biological systems. Polymerases that are useful in the invention include any agent capable of catalyzing extension of a nucleic acid primer in a manner directed by the sequence of a template to which the primer is hybridized. Typically polymerases will be protein enzymes isolated from biological systems.

Alternatively, exon sequences can be determined using sequencing by ligation as described, for example, in Shendure et al. Science 309:1728-1732 (2005); U.S. Pat. No. 5,599,675; and U.S. Pat. No. 5,750,341, each of which is incorporated herein by reference. Sequences of template nucleic acids can be determined using sequencing by hybridization methods such as those described in U.S. Pat. No. 6,090,549; U.S. Pat. No. 6,401,267 and U.S. Pat. No. 6,620,584, each of which is incorporated herein by reference.

If desired, exon sequence products are detected using a ligation assay such as oligonucleotide ligation assay (OLA). Detection with OLA involves the template-dependent ligation of two smaller probes into a single long probe, using a target sequence in an amplicon as the template. In a particular embodiment, a single-stranded target sequence includes a first target domain and a second target domain, which are adjacent and contiguous. A first OLA probe and a second OLA probe can be hybridized to complementary sequences of the respective target domains. The two OLA probes are then covalently attached to each other to form a modified probe. In embodiments where the probes hybridize directly adjacent to each other, covalent linkage can occur via a ligase. One or both probes can include a nucleoside having a label such as a peptide linked label. Accordingly, the presence of the ligated product can be determined by detecting the label. In particular embodiments, the ligation probes can include priming sites configured to allow amplification of the ligated probe product using primers that hybridize to the priming sites, for example, in a PCR reaction.

Alternatively, the ligation probes can be used in an extension-ligation assay wherein hybridized probes are non-contiguous and one or more nucleotides are added along with

one or more agents that join the probes via the added nucleotides. Furthermore, a ligation assay or extension-ligation assay can be carried out with a single padlock probe instead of two separate ligation probes.

In some preferred embodiments, generating a background distribution (**120**) comprises establishing the mutational burden from a plurality of samples of randomly selected genes obtained from the tumor, provided that the number of randomly selected genes in each sample is equal to the number of passenger genes used to compute the total passenger gene mutation burden.

In some preferred embodiments, normalizing the total passenger gene mutation burden against the background distribution (**130**) comprises generating a z-score indicating the number of standard deviations from the mean of the background distribution. In an alternative embodiment, p-values can be used. Z-scores can be correlated to p-values. For example, a z-score of 1.65 equals a p-value of $p<0.05$ and a z-score of 2.3 equals a p-value of $p<0.01$.

Categorizing the cancer patient as an immunotherapy responder may be according to the relationship of the total passenger gene mutation burden to the mean of the background distribution. For example, the patient may be categorized as an immunotherapy responder when the total passenger gene mutation burden is at least a number of standard deviations greater than the mean of the background distribution. The number of standard deviations can be, for example, at least about 1, at least about 1.5, at least about 2, at least about 2.5, at least about 3, or greater than 3 standard deviations greater than the mean of the background distribution.

In some embodiments, a cancer patient can be suffering from cutaneous squamous cell cancer (CSCC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), colon/rectum adenocarcinoma (CORE), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), liver hepatocellular carcinoma (LIHC), brain lower grade glioma (LGG), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), pheochromocytoma and

paraganglioma (PCPG), prostate adenocarcinoma (PRAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma.

In some embodiments, the methods further comprise categorizing a mutated gene in the tumor as a passenger gene. Categorizing a mutated gene in the tumor as a passenger gene may comprise selecting a mutated gene from the tumor and matching the mutated gene to a data structure comprising passenger genes established according to a passenger gene index. The passenger gene index may comprises a correlation coefficient between the fraction of samples comprising the mutated gene obtained from a cancer patient cohort and the median number of mutated genes in each type of tumor within the cancer patient cohort.

When a cancer patient is categorized as an immunotherapy responder, the method may further comprise administering to the cancer patient an immunotherapy regimen. In some embodiments, the immunotherapy regimen comprises administering to the patient an inhibitor of a T cell inhibitory receptor or a receptor on a tumor cell. In some embodiments, an inhibitor of T cell inhibitory receptor or receptor on a tumor cell can comprise an antibody or antigen-binding fragment thereof. In some embodiments, the immunotherapy regimen comprises administering to the patient an activator of a T cell receptor that promotes T cell activation and prolongs immune cytolytic activities.

In some embodiments, T cell inhibitory receptors or receptors on a tumor cell, which can be targeted with inhibitors for immunotherapy comprise one or more of PD1, PDL1, CTLA4, LAG3 and TIM3. Thus, in some embodiments, an inhibitor of a T cell inhibitory receptor or a receptor on a tumor cell comprises an antibody or antigen-binding fragment thereof that specifically binds to one or more of PD1, PDL1, CTLA4, LAG3, and TIM3. As part of an immunotherapy regimen, the cancer patient may be administered an antibody or antigen-binding fragment thereof that specifically binds to one or more of PD1, PDL1, CTLA4, LAG3, and TIM3, or may be administered any combination of two or more such antibodies or antigen-binding fragments thereof.

In some embodiments, the immunotherapy regimen comprises administering to the patient an antibody that binds to PD1. In some preferred embodiments, the antibody that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID

NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22. In embodiments, any of the antibodies or antigen-binding fragments thereof that bind PD1 can be any of the antibodies or antigen-binding fragments thereof described in US Application No. 14/603,776 (Publication No. US 2015-0203579), which is hereby incorporated by reference herein. For example, in some embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 1 and a LCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 1 and an HCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to PD1 comprises an HCVR and LCVR pair as shown in Table 1. Other antibodies that bind to PD1 can be used (or antigen-binding fragments thereof), and these include but are not limited to pembrolizumab, nivolumab, durvalumab, atezolizumab, pidilizumab, camrelizumab, PDR001, MED10680, JNJ-63723283, and MCLA-134.

Table 1: Amino Acid Sequence Identifiers for PD1 antibodies

HCVR SEQ ID NO:	LCVR SEQ ID NO:		HCVR SEQ ID NO:	LCVR SEQ ID NO:
1	2		28	26
3	4		29	26
5	6		30	26
7	8		31	26
9	10		32	26
11	12		33	26
13	14		34	26
15	16		35	26
17	18		36	26

19	20		37	26
21	22		38	24
23	24		39	24
25	26		40	24
27	26			

In some embodiments, the immunotherapy regimen comprises administering to the patient an antibody that binds to the LAG3 protein (aka CD223). In some embodiments, the antibody that binds to LAG3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94. In some embodiments, the antibodies or antigen-binding fragments thereof that bind LAG3 can be any of the antibodies or antigen-binding fragments thereof described in US Application No 15/289,032 (Publication No. US 2017-0101472), which is hereby incorporated by reference herein. For example, in some embodiments, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 2 and a LCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 2 and an HCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to LAG3 comprises an HCVR and LCVR pair as shown in Table 2. Other antibodies that bind to LAG3 can be used (or antigen-binding fragments thereof), and these include but are not limited to BMS-986016 and GSK2381781.

Table 2: Amino Acid Sequence Identifiers for LAG3 antibodies

HCVR SEQ ID NO:	LCVR SEQ ID NO:		HCVR SEQ ID NO:	LCVR SEQ ID NO:
41	42		81	82
43	44		83	84
45	46		85	86
47	48		87	88
49	50		89	90

51	52		91	92
53	54		93	94
55	56		95	96
57	58		97	98
59	60		99	98
61	62		100	98
63	64		101	98
65	66		102	98
67	68		103	98
69	70		104	105
71	72		106	105
73	74		107	105
75	76		108	109
77	78		110	111
79	80			

In some embodiments, the immunotherapy regimen comprises administering to the patient an antibody that binds to PDL1. In some preferred embodiments, the antibody that binds to PDL1 comprises at least the HCVR sequence of SEQ ID NO:122 and the LCVR sequence of SEQ ID NO:123. In some embodiments, the antibodies or antigen-binding fragments thereof that bind PDL1 can be any of the antibodies or antigen-binding fragments thereof described in US Application No 14/603,808 (Publication No. US 2015-0203580), which is hereby incorporated by reference herein. For example, in some embodiments, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 3 and a LCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 3 and an HCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to PDL1 comprises an HCVR and LCVR pair as shown in Table 3. Other antibodies that bind to PDL1 can be used (or antigen-binding fragments thereof), and these include but are not limited to, one or more of avelumab, atezolizumab, and durvalumab.

Table 3: Amino Acid Sequence Identifiers for PDL1 antibodies

HCVR SEQ ID NO:	LCVR SEQ ID NO:		HCVR SEQ ID NO:	LCVR SEQ ID NO:
112	113		137	138
114	115		139	140
116	117		141	142
118	119		143	144
120	121		145	146
122	123		147	146
124	125		148	146
126	127		149	146
128	129		150	146
130	131		151	146
132	133		152	146
134	133		153	146
135	136		154	146

In some embodiments, the immunotherapy regimen comprises administering to the patient an antibody that binds to CTLA4. In some embodiments, the antibodies or antigen-binding fragments thereof that bind CTLA4 can be any of the antibodies or antigen-binding fragments thereof described in US Provisional Application No 62/537,753, filed on July 27, 2017, which is hereby incorporated by reference herein. For example, in some embodiments, the antibody or antigen-binding fragment thereof that binds to CTLA4 comprises a HCVR having an amino acid sequence from among the sequences listed in Table 4 and a LCVR. In some embodiments, the antibody or antigen-binding fragment thereof that binds to CTLA4 comprises a LCVR having an amino acid sequence from among the sequences listed in Table 4 and an HCVR. In some embodiments, the antibody or antigen-

binding fragment thereof that binds to CTLA4 comprises an HCVR and LCVR pair as shown in Table 4. Other antibodies that bind to CTLA4 can be used (or antigen-binding fragments thereof), and these include but are not limited to, one or more of ipilimumab and tremelimumab, as well as any of the antibodies or antigen-binding fragments thereof disclosed in US Patent Nos. 6,984,720; 7,605, 238; or 7,034,121, all of which are hereby incorporated by reference herein.

Table 4: Amino Acid Sequence Identifiers for CTLA4 antibodies

HCVR	LCVR	HCVR	LCVR
155	156	187	188
157	158	189	190
159	160	191	192
161	162	193	192
163	164	195	194
165	166	197	196
167	168	199	198
169	170	201	200
171	172	203	202
173	174	205	204
175	176	207	206
177	178	209	208
179	180	211	210
181	182	213	212
183	184	215	214
185	186	217	216

In some embodiments, the immunotherapy regimen may comprise administering to the patient a combination of one or more inhibitors of a T cell inhibitory receptor. The combination may comprise a combination of antibodies or a combination of antigen-binding portions of such antibodies, or a combination of antibodies and antigen-binding portions. Thus, for example, the immunotherapy regimen may comprise administering to the patient an antibody that binds to PD1 in combination with a second immunotherapy regimen, such

as an antibody that binds to LAG3, or an antibody that binds to PDL1, or an antibody that binds to CTLA. The immunotherapy regimen may comprise administering to the patient an antibody that binds to PDL1 in combination with a second immunotherapy regimen, such as an antibody that binds to LAG3, or an antibody that binds to PD1, or an antibody that binds to CTLA. The immunotherapy regimen may comprise administering to the patient an antibody that binds to LAG3 in combination with a second immunotherapy regimen, such as an antibody that binds to PD1, or an antibody that binds to PDL1, or an antibody that binds to CTLA. The immunotherapy regimen may comprise administering to the patient an antibody that binds to CTLA4 in combination with a second immunotherapy regimen, such as an antibody that binds to LAG3, or an antibody that binds to PDL1, or an antibody that binds to PD1. The antibody that binds to PD1 may comprise any antibody or antigen binding domain described or exemplified herein. The antibody that binds to PD1 may comprise any antibody or antigen binding domain described or exemplified herein. The antibody that binds to PDL1 may comprise any antibody or antigen binding domain described or exemplified herein. The antibody that binds to LAG3 may comprise any antibody or antigen binding domain described or exemplified herein. The antibody that binds to CTLA4 may comprise any antibody or antigen binding domain described or exemplified herein.

In some preferred embodiments, the immunotherapy regimen comprises administering to the patient a combination of an antibody, or antigen binding portion thereof, that binds to PD1 and an antibody, or antigen-binding portion thereof, that binds to LAG3. In some preferred embodiments, the antibody that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22, and the antibody that binds to LAG3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

In some preferred embodiments, the immunotherapy regimen comprises administering to the patient a combination of an antibody, or antigen binding portion thereof, that binds to PDL1 and an antibody, or antigen-binding portion thereof, that binds to LAG3. In some preferred embodiments, the antibody that binds to PDL1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:122 and the light chain

variable region (LCVR) sequence of SEQ ID NO:123, and the antibody that binds to LAG3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

In some embodiments, the immunotherapy can be any of the known immunotherapies for cancer. For example, the immunotherapy can be cemiplimab, nivolumab, pembrolizumab, atezolizumab, durvalumab, avelumab, ipilimumab, IFN-alpha, IL-2, or a combination thereof. In some embodiments, the immunotherapy can be an immune checkpoint inhibitor as described throughout or those commonly known in the art. For example, cemiplimab, nivolumab, pembrolizumab, atezolizumab, durvalumab, avelumab are known immune checkpoint inhibitors.

In some alternative embodiments, the immunotherapy regimen comprises administering to the patient an activator of T cell activating receptor. In some preferred embodiments, a T cell activating receptor, which can be targeted with activators for immunotherapy comprise one or more of CD28, CD40L, ICOS and 4-1BB.

An example of a methodology for establishing a total passenger gene mutation burden from a tumor of a cancer patient is shown in **FIG. 2** and **FIG. 3**. A genetic sample can be obtained/received (**202**). The genetic sample can be from a cancer patient. The genetic sample can be from a tumor of the cancer patient. The genetic sample can be sequenced, resulting in genetic sequence data.

In some embodiments, the sequence data can be obtained or received through any method described herein. For example, the sequence data can be obtained directly, by performing a sequencing process on a sample. Alternatively, or additionally, the sequence data can be obtained indirectly, for example, from a third party, a database and/or a publication. In some embodiments, the sequence data are received at a computer system, for example, from a data storage device or from a separate computer system.

In some embodiments, the sequence data can comprise bulk sequence data. The term “bulk sequencing” or “next generation sequencing” or “massively parallel sequencing” refers to any high throughput sequencing technology that parallelizes the DNA and/or RNA sequencing process. For example, bulk sequencing methods are typically capable of

producing more than one million polynucleic acid amplicons in a single assay. The terms "bulk sequencing," "massively parallel sequencing," and "next generation sequencing" refer only to general methods, not necessarily to the acquisition of greater than 1 million sequence tags in a single run. Any bulk sequencing method can be implemented in the disclosed methods and systems, such as reversible terminator chemistry (e.g., Illumina), pyrosequencing using polony emulsion droplets (e.g., Roche), ion semiconductor sequencing (IonTorrent), single molecule sequencing (e.g., Pacific Biosciences), massively parallel signature sequencing, etc.

In some embodiments, the sequence data can be produced by any sequencing method known in the art. For example, in some embodiments the sequencing data are produced using chain termination sequencing, sequencing by ligation, sequencing by synthesis, pyrosequencing, ion semiconductor sequencing, single-molecule real-time sequencing, tag-based sequencing, dilute-`n`-go sequencing, and/or 454 sequencing.

In some embodiments, the sequence data are the result of a process whereby a nucleic acid amplification process is performed to amplify at least part of one or more genomic locus or transcript, followed by the sequencing of the resulting amplification product. Examples of nucleic acid amplification processes useful in the performance of methods disclosed herein include, but are not limited to, polymerase chain reaction (PCR), LATE-PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), transcription mediated amplification (TMA), self-sustained sequence replication (3SR), Qβ replicase based amplification, nucleic acid sequence-based amplification (NASBA), repair chain reaction (RCR), boomerang DNA amplification (BDA) and/or rolling circle amplification (RCA).

In some embodiments, the method includes the step of performing a sequencing process on a sample. Any sample can be used, so long as the sample contains DNA and/or RNA from a tumor of a patient. The source of the sample may be, for example, solid tissue, as from a fresh, frozen and/or preserved organ, tissue sample, biopsy, or aspirate; blood or any blood constituents, serum, blood; bodily fluids such as cerebral spinal fluid, amniotic fluid, peritoneal fluid or interstitial fluid.

The genetic sequence data can be analyzed (**204**) via a computing device to identity

driver genes and to determine a number of mutations in the driver genes. If the number of mutations in the driver genes is high (**206**), an indication can be generated that the patient will respond to the immunotherapy (**210**). If the number of mutations in the driver genes is low (**206**), an indication can be generated poor or no response will be seen with immunotherapy (**208**). Mutations on driver genes could promote "hallmarks of cancer" e.g., immune escape.

In an alternate embodiment, the genetic sequence data can be analyzed (**212**) via a computing device to identify passenger genes and to determine a number of mutations in the passenger genes. If the number of mutations in the passenger genes is high (**216**), an indication can be generated (**210**) that the patient will respond to the immunotherapy. If the number of mutations in the passenger genes is low (**216**), an indication can be generated (**208**) that poor or no response will be seen with immunotherapy. While passenger genes do not have any causal implication in cancer, mutations on passenger genes can be used to assess immunogenicity. In some embodiments, the genetic sequence data can be analyzed (**212**) to identify passenger genes, determine a number of mutations in the passenger genes, and determine a background distribution for the mutational burden of the tumor. The number of mutations in the passenger genes can be analyzed with regard to the background distribution to determine how many standard deviations (if any) the number of mutations in the passenger genes is from the mean. If the number of standard deviations is high (e.g., at least 1, 1.5, 2, 2.5) (**216**) the cancer patient can be categorized as a better immunotherapy responder (**210**). If the number of standard deviations is low (**216**) the cancer patient can be categorized as a poor immunotherapy responder (**208**).

In some embodiments, the passenger genes can be identified in large-scale cancer genome analysis according to a metric referred to herein as a Passenger Gene Index (PGI) (**212**). In some embodiments, the PGI is based on genetic mutation rates (GMR) of the passenger genes being highly correlated with overall cancer mutation frequencies, also referred to as tumor mutational burden (**214**). Identified passenger genes are enriched for families known for excessive passenger mutations, such as extremely large proteins and genes with low expression level or late DNA replication time. More passenger gene mutations will accumulate in cancer samples/types with higher mutations rates, and the

average number of mutated genes per sample in each cancer type can be a surrogate for likelihood of passenger mutations in that cancer type. Thus, PGI can be defined, for each gene, X_i, as a correlation between percentage sample with gene X_i mutation and average number of mutated gene per sample in each cancer type. A higher PGI score indicates that a particular gene is more likely to acquire somatic mutations in the cancer types with higher overall mutation frequency. Genes with a low PGI show a weak association between the two variables (e.g., and can be observed in canonical cancer driver genes such as TP53, PIK3CA and KRAS). Genes ranked at the top of the PGI are enriched for gene families known for excessive passenger mutations, e.g., extremely large proteins (>4,000 amino acids), genes spanning large genomic loci (>1Mb), genes with low expression level, genes with late DNA replication time, and the like. A cumulative distribution function (CDF) of these gene families show sharp uptrend at PGI>0.7, while the genes in Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) are more uniformly distributed. Two-sample Kolmogorov-Smirnov tests show significant difference in the rank distribution of passenger gene families as compared to that of CGC genes (p=8.3×10^{-19} for large proteins; p=2.9×10^{-12} for genomic locus >1Mb; p=6.4×10^{-35} for low expression; p=2.7×10^{-29} for late replication). Similar results are obtained when samples are grouped by mutation rate (instead of cancer type) for computing PGI.

The top passenger genes, based on highest PGI, can be agnostic to tumor type or specific for each tumor type. Thus, in some instances, the top passenger genes can be used generically, regardless of tumor type. Although these top passenger genes do not change regardless of tumor type, the top passenger genes can change over time due to accessibility of additional samples. In some instances, the top passenger genes can vary between tumor types. In some instances, the top passenger genes can be identical between tumor types. Furthermore, if the top passenger genes are identical between tumor types, the ranking within that top passenger gene list can vary. For example, the top 50 passenger genes for breast cancer can be identical to the top 50 passenger genes for lung cancer however the number one passenger gene (meaning highest PGI) for breast cancer can be the number five passenger gene for lung cancer. In some instances, the top 50 passenger genes of one tumor type can comprise 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, up to 100 percent of the top 50 passenger genes of a second tumor type. Depending on what range of PGI is included, a

list of top 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, or even 2000 plus passenger genes can be included on a top passenger gene list. In some instances, the top passenger genes do not vary among patients. All patients can use the same passenger gene list and each patient will have a different TMB score.

In an embodiment, illustrated in **FIG. 3**, a method (**300**) is disclosed comprising receiving genetic sequence data (**310**). The genetic sequence data can comprise a plurality of genes and can be derived from a plurality of biological samples collected from subjects having a plurality of disease types. The plurality of disease types can comprise cancers.

In some embodiments, the method (**300**) can identify a plurality of mutated genes for each of the plurality of biological samples (**320**), wherein each of the mutated genes comprises a genetic sequence having at least one non-synonymous somatic mutation.

In some embodiments, the method (**300**) can determine a tumor mutational burden for each biological sample based on a number of mutated genes in each biological sample (**330**). In a preferred embodiment, determining the tumor mutational burden for each biological sample based on a number of mutated genes in each biological sample can comprise adding a number of mutated genes in each patient sample.

The method (**300**) can identify a mutation in a gene (passenger or driver), for example, by aligning the mutated sequences with wild type or reference sequences. Various programs and alignment algorithms are described in: Smith and Waterman (1981) Adv. Appl. Math. 2:482; Needleman and Wunsch (1970) J. Mol. Biol. 48:443; Pearson and Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444; Higgins and Sharp (1988) Gene 73:237-244; Higgins and Sharp (1989) CABIOS 5:151-153; Corpet et al. (1988) Nucl. Acids Res. 16:10881-90; Huang et al. (1992) *Computer Appl. in the Biosci.* 8:155-65; and Pearson et al. (1994). *Meth. Mol. Biol.* 24:307-31, which are herein incorporated by reference. Altschul et al. (1994) *Nature Genet.* 6:119-29 (herein incorporated by reference) present a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) is available from several sources, including the National Center for Biological Information (NCBI, Bethesda, Md.) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at

</www.ncbi.nlmn.ih.gov/BLAST/>. A description of how to determine sequence identity using this program is available at </www.nebi.rlm.nih.gov/BLAST/blast-help.html>.

In some embodiments, for each disease type, the method (**300**) can determine an average tumor mutational burden of the plurality of mutated genes in the plurality of biological samples based on the determined numbers of mutated genes in each biological sample (**340**). In a preferred embodiment, determining the average tumor mutational burden of the plurality of mutated genes in the plurality of biological samples based on the determined numbers of mutated genes in each biological sample can comprise adding the tumor mutational burden from each patient sample and dividing by a number of patient samples for each disease type.

In some embodiments, for each mutated gene and each disease type, the method (**300**) can determine a fraction of biological samples comprising the mutated gene (**350**).

In some embodiments, for each mutated gene, the method (**300**) can determine a correlation coefficient between the average tumor mutational burden and the fraction of biological samples comprising the mutated gene (**360**).

In some embodiments, the method (**300**) can determine whether the mutated gene is a passenger gene based on the correlation coefficient (**370**). A higher correlation coefficient indicates that a particular gene is more likely to acquire somatic mutations in the cancer types with higher overall mutation frequency (e.g., passenger gene), whereas a lower correlation coefficient indicates that a particular gene is less likely to acquire somatic mutations in the cancer types with higher overall mutation frequency (e.g., not a passenger gene).

In an alternate embodiment, the method (**300**) can further comprise generating a list of the mutated genes identified as passenger genes. In aspect preferred embodiment, the list can represent an immunogenicity profile for the selected disease.

In some embodiments, illustrated in **FIG. 4**, a method for selecting a patient for cancer therapy (**400**) is disclosed comprising determining a plurality of passenger genes present in a tumor sample for a patient with a disease (**410**).

In some embodiments, the method (**400**) can compare the plurality of passenger

genes to an immunogenicity profile for the disease (**420**). In a preferred embodiment, the immunogenicity profile can be generated by performing steps comprising receiving genetic sequence data, wherein the genetic sequence data comprises a plurality of genes and is derived from a plurality of biological samples collected from subjects having a plurality of disease types, identifying a plurality of mutated genes for each of the plurality of biological samples, wherein each of the mutated genes comprises a genetic sequence having at least one non-synonymous somatic mutation, determining a tumor mutational burden for each biological sample based on a number of mutated genes in each biological sample, for each disease type, determining an average tumor mutational burden of the plurality of mutated genes in the plurality of biological samples based on the determined numbers of mutated genes in each biological sample, for each mutated gene and each disease type, determining a fraction of biological samples comprising the mutated gene, for each mutated gene, determining a correlation coefficient between the average tumor mutational burden and the fraction of biological samples comprising the mutated gene. In some embodiments, the mutated gene can be determined to be a passenger gene based on the correlation coefficient. A higher correlation coefficient indicates that a particular gene is more likely to acquire somatic mutations in the cancer types with higher overall mutation frequency (e.g., passenger gene), whereas a lower correlation coefficient indicates that a particular gene is less likely to acquire somatic mutations in the cancer types with higher overall mutation frequency (e.g., not a passenger gene).

A list of the mutated genes identified as passenger genes can be generated, wherein the list represents the immunogenicity profile for the selected disease. In a preferred embodiment, determining the tumor mutational burden for each biological sample based on a number of mutated genes in each biological sample can comprise adding a number of mutated genes in each patient sample. In a preferred embodiment, determining the average tumor mutational burden of the plurality of mutated genes in the plurality of biological samples based on the determined numbers of mutated genes in each biological sample can comprise adding the tumor mutational burden from each patient sample and dividing by a number of patient samples for each disease type.

In some embodiments, PGI can be used to identify passenger genes for particular

cancers and then using TMB of the passenger genes, patients can be identified that are responders to specific treatments, such as, but not limited to, anti-PD-1 or a combination of anti-PD-1 and another cancer therapeutic. TMB of passenger genes can also be used to identify responders to other cancer antibody treatments, such as, but not limited to, anti-CD20 (chronic lyphocytic leukemia), anti-HER2 (breast cancer), anti-EGFR (colorectal and head and neck cancer), anti-CD19 (B cell cancers), and anti-CD20 (lymphoma) or combinations of an antibody treatment and another cancer therapeutic. In some embodiments, the other cancer therapeutic can be chemotherapy, an immunomodulatory agent (e.g., a second antibody, a cytokine), radiation, or surgery.

In some embodiments, comparing the plurality of passenger genes to an immunogenicity profile for the disease can comprise determining a number of matches between the plurality of mutated genes and a list of mutated genes in the profile.

In some embodiments, if the plurality of passenger genes matches the immunogenicity profile for the disease (**430**), the method (**400**) can identify the patient as a candidate for immunotherapy.

In some embodiments, if the plurality of passenger genes does not match the immunogenicity profile for the disease (**440**), the method (**400**) can identify the patient as not a candidate for immunotherapy.

In an alternate embodiment, the method (**400**) can further comprise enrolling the patient in an immunotherapy program if the patient was identified as a candidate for immunotherapy.

The disclosed immunotherapies can be used in combination with other antibody or antigen-binding fragments thereof as well as other anti-cancer therapies. Combination therapies can be administered simultaneously or sequentially. In some embodiments, two or more therapies can be formulated together with a pharmaceutically acceptable carrier resulting in a pharmaceutical composition. In some embodiments, two or more therapies are formulated individually with a pharmaceutically acceptable carrier resulting in two or more pharmaceutical compositions. By “pharmaceutically acceptable” is meant a material or carrier that would be selected to minimize any degradation of the active ingredient and to minimize any adverse side effects in the subject, as would be well known to one of skill in

the art. Examples of carriers include dimyristoylphosphatidyl (DMPC), phosphate buffered saline or a multivesicular liposome. For example, PG:PC:Cholesterol:peptide or PC:peptide can be used as carriers in this invention. Other suitable pharmaceutically acceptable carriers and their formulations are described in Remington: The Science and Practice of Pharmacy (19th ed.) ed. A.R. Gennaro, Mack Publishing Company, Easton, PA 1995. Typically, an appropriate amount of pharmaceutically-acceptable salt is used in the formulation to render the formulation isotonic. Other examples of the pharmaceutically-acceptable carrier include, but are not limited to, saline, Ringer's solution and dextrose solution. The pH of the solution can be from about 5 to about 8, or from about 7 to about 7.5. Further carriers include sustained release preparations such as semi-permeable matrices of solid hydrophobic polymers containing the composition, which matrices are in the form of shaped articles, e.g., films, stents (which are implanted in vessels during an angioplasty procedure), liposomes or microparticles. It will be apparent to those persons skilled in the art that certain carriers may be more preferable depending upon, for instance, the route of administration and concentration of composition being administered. These most typically would be standard carriers for administration of drugs to humans, including solutions such as sterile water, saline, and buffered solutions at physiological pH.

Pharmaceutical compositions can also include carriers, thickeners, diluents, buffers, preservatives and the like, as long as the intended activity of the immunotherapy of the invention is not compromised. Pharmaceutical compositions may also include one or more active ingredients (in addition to the composition of the invention) such as antimicrobial agents, anti-inflammatory agents, anesthetics, and the like. The pharmaceutical composition may be administered in a number of ways depending on whether local or systemic treatment is desired, and on the area to be treated.

Preparations of parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated

Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

Formulations for optical administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids, or binders may be desirable. Some of the compositions may potentially be administered as a pharmaceutically acceptable acid- or base- addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide, potassium hydroxide, and organic bases such as mon-, di-, trialkyl and aryl amines and substituted ethanolamines.

Pharmaceutical compositions of the present invention suitable for injectable use include sterile aqueous solutions or dispersions. Furthermore, the compositions can be in the form of sterile powders for the extemporaneous preparation of such sterile injectable solutions or dispersions. Typically, the final injectable form should be sterile and should be effectively fluid for easy syringability. The pharmaceutical compositions should be stable under the conditions of manufacture and storage; thus, preferably should be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol and liquid polyethylene glycol), vegetable oils, and suitable mixtures thereof.

Injectable solutions, for example, can be prepared in which the carrier comprises saline solution, glucose solution or a mixture of saline and glucose solution. Injectable

suspensions may also be prepared in which case appropriate liquid carriers, suspending agents and the like may be employed. Also included are solid form preparations that are intended to be converted, shortly before use, to liquid form preparations.

Preparations of parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

Pharmaceutical compositions of the present invention can be in a form suitable for topical use such as, for example, an aerosol, cream, ointment, lotion, dusting powder, mouth washes, gargles, and the like. Further, the compositions can be in a form suitable for use in transdermal devices. These formulations can be prepared, utilizing a compound of the invention, or pharmaceutically acceptable salts thereof, via conventional processing methods. As an example, a cream or ointment is prepared by mixing hydrophilic material and water, together with about 5 wt% to about 10 wt% of the compound, to produce a cream or ointment having a desired consistency.

In the compositions suitable for percutaneous administration, the carrier optionally comprises a penetration enhancing agent and/or a suitable wetting agent, optionally combined with suitable additives of any nature in minor proportions, which additives do not introduce a significant deleterious effect on the skin. Said additives may facilitate the administration to the skin and/or may be helpful for preparing the desired compositions. These compositions may be administered in various ways, e.g., as a transdermal patch, as a spot on, as an ointment.

Pharmaceutical compositions of this invention can be in a form suitable for rectal administration wherein the carrier is a solid. It is preferable that the mixture forms unit dose suppositories. Suitable carriers include cocoa butter and other materials commonly used in

the art. The suppositories can be conveniently formed by first admixing the composition with the softened or melted carrier(s) followed by chilling and shaping in molds.

In addition to the aforementioned carrier ingredients, the pharmaceutical formulations described above can include, as appropriate, one or more additional carrier ingredients such as diluents, buffers, flavoring agents, binders, surface-active agents, thickeners, lubricants, preservatives (including anti-oxidants) and the like. Furthermore, other adjuvants can be included to render the formulation isotonic with the blood of the intended recipient. Compositions containing a disclosed immunotherapy, and/or pharmaceutically acceptable salts thereof, can also be prepared in powder or liquid concentrate form.

The exact dosage and frequency of administration depends on the particular disclosed peptide, a product of a disclosed method of making, a pharmaceutically acceptable salt, solvate, or polymorph thereof, a hydrate thereof, a solvate thereof, a polymorph thereof, or a stereochemically isomeric form thereof; the particular condition being treated and the severity of the condition being treated; various factors specific to the medical history of the subject to whom the dosage is administered such as the age; weight, sex, extent of disorder and general physical condition of the particular subject, as well as other medication the individual may be taking; as is well known to those skilled in the art. Furthermore, it is evident that said effective daily amount may be lowered or increased depending on the response of the treated subject and/or depending on the evaluation of the physician prescribing the compositions.

Depending on the mode of administration, the pharmaceutical composition will comprise from 0.05 to 99 % by weight, preferably from 0.1 to 70 % by weight, more preferably from 0.1 to 50 % by weight of the active ingredient, and, from 1 to 99.95 % by weight, preferably from 30 to 99.9 % by weight, more preferably from 50 to 99.9 % by weight of a pharmaceutically acceptable carrier, all percentages being based on the total weight of the composition.

In an exemplary embodiment, some or all of the methods and systems can be implemented on one or more computers, such as a computer (**1001**) as illustrated in **FIG. 10** and described below. In some embodiments, the methods and systems disclosed can utilize one or more computers to perform one or more functions in one or more locations. **FIG. 10**

shows a block diagram illustrating an exemplary operating environment for performing the disclosed methods. This exemplary operating environment is only an example of an operating environment and is not intended to suggest any limitation as to the scope of use or functionality of operating environment architecture. Neither should the operating environment be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment.

In some embodiments, the present methods and systems can be operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well-known computing systems, environments, and/or configurations that can be suitable for use with the systems and methods comprise, but are not limited to, personal computers, server computers, laptop devices, and multiprocessor systems. Additional examples comprise set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that comprise any of the above systems or devices, and the like.

In some embodiments, the processing of the disclosed methods and systems can be performed by software components. The disclosed systems and methods can be described in the general context of computer-executable instructions, such as program modules, being executed by one or more computers or other devices. Generally, program modules comprise computer code, routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. The disclosed methods can also be practiced in grid-based and distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules can be located in both local and remote computer storage media including memory storage devices.

Further, one skilled in the art will appreciate that the systems and methods disclosed herein can be implemented via a general-purpose computing device in the form of a computer **1001**. The components of the computer **1001** can comprise, but are not limited to, one or more processors **1003**, a system memory **1012**, and a system bus **1013** that couples various system components including the one or more processors **1003** to the

system memory **1012**. The system can utilize parallel computing.

The system bus **1013** represents one or more of several possible types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, or local bus using any of a variety of bus architectures. The bus **1013**, and all buses specified in this description can also be implemented over a wired or wireless network connection and each of the subsystems, including the one or more processors **1003**, a mass storage device **1004**, an operating system **1005**, PGI software **1006**, PGI data **1007**, a network adapter **1008**, the system memory **1012**, an Input/Output Interface **1010**, a display adapter **1009**, a display device **1011**, and a human machine interface **1002**, can be contained within one or more remote computing devices **1014a,b,c** at physically separate locations, connected through buses of this form, in effect implementing a fully distributed system.

The computer **1001** typically comprises a variety of computer readable media. Exemplary readable media can be any available media that is accessible by the computer **1001** and comprises, for example and not meant to be limiting, both volatile and non-volatile media, removable and non-removable media. The system memory **1012** comprises computer readable media in the form of volatile memory, such as random access memory (RAM), and/or non-volatile memory, such as read only memory (ROM). The system memory **1012** typically contains data such as the PGI data **1007** and/or program modules such as the operating system **1005** and the PGI software **1006** that are immediately accessible to and/or are presently operated on by the one or more processors **1003**. The PGI data **1007** can comprise read coverage data and/or expected read coverage data.

In some embodiments, the computer **1001** can also comprise other removable/non-removable, volatile/non-volatile computer storage media. By way of example, **FIG. 10** illustrates the mass storage device **1004** which can provide non-volatile storage of computer code, computer readable instructions, data structures, program modules, and other data for the computer **1001**. For example and not meant to be limiting, the mass storage device **1004** can be a hard disk, a removable magnetic disk, a removable optical disk, magnetic cassettes or other magnetic storage devices, flash memory cards, CD-ROM, digital versatile disks (DVD) or other optical storage, random access memories (RAM), read only memories

(ROM), electrically erasable programmable read-only memory (EEPROM), and the like.

Optionally, any number of program modules can be stored on the mass storage device **1004**, including by way of example, the operating system **1005** and the PGI software **1006**. Each of the operating system **1005** and the PGI software **1006** (or some combination thereof) can comprise elements of the programming and the PGI software **1006**. The PGI data **1007** can also be stored on the mass storage device **1004**. The PGI data **1007** can be stored in any of one or more databases known in the art. Examples of such databases comprise, DB2®, Microsoft® Access, Microsoft® SQL Server, Oracle®, mySQL, PostgreSQL, and the like. The databases can be centralized or distributed across multiple systems.

In an alternate embodiment, the user can enter commands and information into the computer **1001** via an input device (not shown). Examples of such input devices comprise, but are not limited to, a keyboard, pointing device (*e.g.*, a "mouse"), a microphone, a joystick, a scanner, tactile input devices such as gloves, and other body coverings, and the like. These and other input devices can be connected to the one or more processors **1003** via the human machine interface **1002** that is coupled to the system bus **1013**, but can be connected by other interface and bus structures, such as a parallel port, game port, an IEEE 1394 Port (also known as a Firewire port), a serial port, or a universal serial bus (USB).

In an alternate embodiment, the display device **1011** can also be connected to the system bus **1013** via an interface, such as the display adapter **1009**. It is contemplated that the computer **1001** can have more than one display adapter **1009** and the computer **1001** can have more than one display device **1011**. For example, a display device can be a monitor, an LCD (Liquid Crystal Display), or a projector. In addition to the display device **1011**, other output peripheral devices can comprise components such as speakers (not shown) and a printer (not shown) which can be connected to the computer **1001** via the Input/Output Interface **1010**. Any step and/or result of the methods can be output in any form to an output device. Such output can be any form of visual representation, including, but not limited to, textual, graphical, animation, audio, tactile, and the like. The display **1011** and computer **1001** can be part of one device, or separate devices.

The computer **1001** can operate in a networked environment using logical connections to one or more remote computing devices **1014a,b,c**. By way of example, a

remote computing device can be a personal computer, portable computer, smartphone, a server, a router, a network computer, a peer device or other common network node, and so on. Logical connections between the computer **1001** and a remote computing device **1014a,b,c** can be made via a network **1015**, such as a local area network (LAN) and/or a general wide area network (WAN). Such network connections can be through the network adapter **1008**. The network adapter **1008** can be implemented in both wired and wireless environments. Such networking environments are conventional and commonplace in dwellings, offices, enterprise-wide computer networks, intranets, and the Internet.

For purposes of illustration, application programs and other executable program components such as the operating system **1005** are illustrated herein as discrete blocks, although it is recognized that such programs and components reside at various times in different storage components of the computing device **1001**, and are executed by the one or more processors **1003** of the computer. In an aspect, at least a portion of the PGI software **1006** and/or the PGI data **1007** can be stored on and/or executed on one or more of the computing device **1001,** the remote computing devices **1014a,b,c,** and/or combinations thereof. Thus the PGI software **1006** and/or the PGI data **1007** can be operational within a cloud computing environment whereby access to the PGI software **1006** and/or the PGI data **1007** can be performed over the network **1015** (e.g., the Internet). Moreover, in an aspect the PGI data **1007** can be synchronized across one or more of the computing device **1001,** the remote computing devices **1014a,b,c,** and/or combinations thereof.

An implementation of the PGI software **1006** can be stored on or transmitted across some form of computer readable media. Any of the disclosed methods can be performed by computer readable instructions embodied on computer readable media. Computer readable media can be any available media that can be accessed by a computer. By way of example and not meant to be limiting, computer readable media can comprise "computer storage media" and "communications media." "Computer storage media" comprise volatile and non-volatile, removable and non-removable media implemented in any methods or technology for storage of information such as computer readable instructions, data structures, program modules, or other data. Exemplary computer storage media comprises,

but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by a computer.

The methods and systems can employ Artificial Intelligence techniques such as machine learning and iterative learning. Examples of such techniques include, but are not limited to, expert systems, case based reasoning, Bayesian networks, behavior based AI, neural networks, fuzzy systems, evolutionary computation (e.g. genetic algorithms), swarm intelligence (e.g. ant algorithms), and hybrid intelligent systems (e.g. Expert inference rules generated through a neural network or production rules from statistical learning).

Unless otherwise expressly stated, it is in no way intended that any method set forth herein be construed as requiring that its steps be performed in a specific order. Accordingly, where a method claim does not actually recite an order to be followed by its steps or it is not otherwise specifically stated in the claims or descriptions that the steps are to be limited to a specific order, it is in no way intended that an order be inferred, in any respect. This holds for any possible non-express basis for interpretation, including: matters of logic with respect to arrangement of steps or operational flow; plain meaning derived from grammatical organization or punctuation; the number or type of embodiments described in the specification.

The following examples are provided to describe the disclosure in greater detail. They are intended to illustrate, not to limit, the disclosure.

Example 1

Passenger Gene Index

The method of Passenger Gene Index (PGI) involves all TCGA samples binned by cancer type, and the median number of mutated genes was determined for each bin. Mutations were limited only to nonsilent somatic mutations by comparing solid tumors to the blood derived or solid normal counterparts, with the exception in acute myeloid leukemia where blood derived tumors were compared to the solid tissue normal. Mutation profiles were constructed as a binary matrix such that a bit is set if any locus correspond to the gene harbors a mutation in that patient. PGI is computed for each gene X_i, as Pearson

correlation between fraction of sample with gene X_i mutation and the median number of mutated gene in each cancer type. Before computing the correlations, an infinitesimal amount of uniformly distributed noise was added to the sample fraction with mutation to avoid issues with all zero entries.

Example 2

Z-Score for Driver and Passenger Gene Tumor Mutational Burden

The method of Z-score for driver/passenger TMB involves the following. To compute z-score for a TMB, background distribution of the TMB was first established using 1000 randomly selected gene sets of equal size. The numbers of mutated driver/passenger gene were then compared to the background distribution to compute the z-score, indicating how many standard deviations the number is from the mean of the background. Driver genes were downloaded from COSMIC Cancer Gene Census on January 22, 2015, and passenger genes were defined as top n genes ranked by PGI derived from TCGA data.

Example 3

Passenger Gene Tumor Mutational Burden and Immunotherapy Responsiveness

To compute Passenger Gene Index (PGI), a list was compiled of nonsilent somatic mutations from 6,685 samples across 20 TCGA tumor types. Somatic mutations were determined by comparing tumor genome to that of the germ line, e.g. blood derived normal sample from the same patient. The median number of altered genes per sample ranged from 9 in acute myeloid leukemia to 289 in skin cutaneous melanoma, representing over 32-fold difference between the lowest and the highest mutation rate cancers (**FIG. 5**). This is consistent with previous observations that skin and lung cancer samples have the highest mutation rates, due to the exposure to environmental mutagens. FIG. 5 shows the number of nonsilent somatic mutations per sample in 6,685 TCGA cancer exomes.

The 20 cancer types included in this study are bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), colon/rectum adenocarcinoma (CORE), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), liver hepatocellular carcinoma (LIHC), brain lower grade glioma (LGG), lung adenocarcinoma

(LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), and uterine corpus endometrioid carcinoma (UCEC).

It was hypothesized that more passenger mutations will accumulate in cancer types with higher overall mutation rates, and the average number of altered genes per sample in each cancer type can be a surrogate for likelihood of passenger mutations in that cancer type. PGI was defined for each gene, X_i, as a correlation between percentage sample with gene X_i variant and average number of altered gene per sample in each cancer type. A higher PGI score indicated that a particular gene is more likely to acquire somatic mutations in the cancer types with higher overall mutation frequency. Passenger genes show strong linear relationship of the two variables, while weak associations were observed in canonical cancer genes such as TP53, PIK3CA and KRAS (**FIG. 6**). **FIG. 6** illustrates scatter plots for a fraction of patients with the gene variant (y-axis) and average number of total mutated gene (x-axis) in each cancer type. The top row shows a strong linear relationship in the top passenger genes (MUC16 r=0.979; ADAM2 r=0.972; COL5A2 r=0.968), and the bottom row shows weak association of the 2 variables in canonical cancer genes (TP53 r=0.301; PIK3CA r=0.120; KRAS r=0.222).

Genes ranked at the top of the PGI are enriched for gene families known for excessive passenger mutations, e.g., extremely large proteins (>4,000 amino acids), genes spanning large genomic loci (>1Mb), genes with low expression level, and genes with late DNA replication time. A cumulative distribution function (CDF) of these gene families show a sharp uptrend at PGI>0.7, while the driver genes in Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) are more uniformly distributed (**FIG. 7**). **FIG. 7** illustrates enrichment along the PGI scale for cancer driver genes and various other gene groups. The dotted lines (upper) show the fraction of genes at different PGI, and the vertical lines (lower) indicate the rank of individual genes. A two-sample Kolmogorov-Smirnov test was used to examine the difference in gene distribution for each group, as compared to the cancer genes distribution. The two-sample Kolmogorov-Smirnov tests showed significant difference in the rank distribution of passenger gene families as compared to that of CGC

genes (p=8.3×10^{-19} for large proteins; p=2.9×10^{-12} for genomic locus >1Mb; p=6.4×10^{-35} for low expression; p=2.7×10^{-29} for late replication). Similar results were observed when samples were grouped by mutation rate (instead of cancer type) for computing PGI. While some CGC genes also possess high PGI score, they are not validated for their top altered TCGA cancer types. For instance, KDR (kinase insert domain receptor) has the highest rate of mutation in melanoma (14% in SKCM) but KDR is only known for its causal implication in non-small-cell lung carcinoma and angiosarcoma. Similarly, we did not see any validated cases in the highest altered cancer type for the top 30 CGC genes. In contrast, 16 out of the 30 CGC driver genes with the lowest PGI are validated in their corresponding altered cancer type (**FIG. 8**). **FIG. 8** shows that low PGI CGC genes are more likely to be validated in the altered cancer type. **FIG. 8** shows the highest (left) and lowest (right) PGI CGC genes, and their corresponding cancer type with the highest percentage (>2%) of mutated sample. The acronyms marked with an asterisk are the cancer type validated by CGC for the gene. None of the cancer in the highest PGI CGC genes is validated, and 16/30 cancer types in the lowest PGI CGC genes are validated.

PGI was applied as a metric to select passenger genes and use the tumor mutational burden (TMB) of the selected passenger genes to stratify patient cohort that is more likely to respond to immunotherapy. To demonstrate this approach, local immune cytolytic activities and T-cell receptor (TCR) read count were used as the surrogate for immunogenicity, and tested if there is any immunogenicity difference between high and low TMB patients in TCGA data. For each patient, the TMB was computed in 3 different approaches, namely (i) conventional total TMB, (ii) TMB by driver genes, and (iii) TMB by passenger genes. To quantify the cytolytic activities, we adopted a simple RNA-based metric that based on gene expression level of two key cytolytic effectors, granzyme A (GZMA) and perforin (PRF1). The cytolytic activities were found significantly different ($p<0.05$ by Mann–Whitney U test) between high and low passenger TMB patients in 7 different cancer types (colon adenocarcinoma, $p<4.6x10^{-11}$; breast invasive carcinoma, $p<5.0x10^{-4}$; lung adenocarcinoma, $p<7.7x10^{-4}$; uterine corpus endometrioid carcinoma, $p<9.9x10^{-4}$; cervical squamous cell carcinoma, $p<2.2x10^{-3}$; lung squamous cell carcinoma, $p<5.7x10^{-3}$; prostate adenocarcinoma, $p<2.1x10^{-2}$), and the differences are more significant as compared to those using total TMB and driver gene TMB in the corresponding cancer types (**FIG. 9A**).

TCR is responsible for the recognition of peptide-MHC complexes and its diversity is directly associated to the number of foreign or mutated proteins, e.g. neo-antigens from cancer cells. TCRβ repertoire analysis was performed using TCGA RNA-seq data, and compared TCRβ read count between high and low passenger TMB patients. As shown in **FIG. 9B**, the detected number of TCRβ read count were significantly different between high and low passenger TMB patients in 8 different cancer types (uterine corpus endometrioid carcinoma, $p<3.2\times10^{-6}$; colon adenocarcinoma, $p<4.5\times10^{-6}$; cervical squamous cell carcinoma, $p<2.4\times10^{-3}$; breast invasive carcinoma, $p<7.3\times10^{-3}$; skin cutaneous melanoma, $p<9.2\times10^{-3}$; lung adenocarcinoma, $p<1.5\times10^{-2}$; ovarian serous cystadenocarcinoma, $p<2.7\times10^{-2}$; prostate adenocarcinoma, $p<3.8\times10^{-2}$). In agreement with the observation in cytolytic activities, the TCRβ differences are more significant between the groups segregated by passenger TMB, as compared to those using total TMB and driver TMB.

Finally, tests were done to see if there is any survival advantage associated with TMB in TCGA data. In cervical and lung squamous cell carcinoma (CESC and LUSC), although not statistically significant, total TMB shows positive association trend with better survival outcome, while driver TMB is associated with worse prognosis (**FIG. 9C**). **FIG. 9C** illustrates clinical outcome of patient cohorts segregated by mutation load of: (i) all genes, (ii) driver genes, and (iii) passenger genes in skin cutaneous melanoma (SKCM), cervical squamous cell carcinoma (CESC) and endocervical adenocarcinoma , and lung squamous cell carcinoma (LUSC) yield significant survival differences only in patient cohorts segregated by passenger TMB but not total/driver TMB. SKCM shows significant difference in patient survival between high and low total/passenger TMB groups.

Only when the passenger TMB was used, the difference of the survival outcome between the high and low TMB patient groups was statistically significant in both CESC and LUSC. In skin cutaneous melanoma (SKCM), patient stratification using passenger and total TMB both show similar significant separation in survival curves, indicating that there is very little or no driver gene mutation that strongly impact the immunogenicity suppression in melanoma. Using an independent dataset of CTLA-4 blockade in metastatic melanoma, 110 patients were segregated into two groups of equal size by mutation burden. Stratification using TMB of 200 passenger genes improved the clinical benefit rate of the selected patient

group from a baseline of 24.55% to 36.36% (Fisher's Exact Test p=0.0035). Patient stratification using the total TMB yields the same improvement in clinical benefit, further corroborate the observations in cytolytic activities, TCR detection, and survival advantage for melanoma using TCGA data.

FIG. 11 shows the TMB of the patient cohort in an anti-PD1 phase 1 clinical study. Closed circles, closed squares, and closed triangles indicate patients with partial response (PR), stable disease (SD), and progressive disease (PD) respectively. Hollow shapes show data for individual patients, and solid shapes show the average in each PR/SD/PD groups. Total TMBs are shown as the total number of mutated genes (left y-axis) and driver/passenger TMBs are shown in z-score (right y-axis). Both PR-vs-PD and PR-vs-PD+SD show statistically significant differences in the passenger TMB, invariantly of the number of top (50/100/1000) passenger genes used. PR group does not yield significant difference in the total or driver TMB.

Example 4

Passenger Gene Tumor Mutational Burden and Clinical Immunotherapy Response

To evaluate the clinical response of various malignancies to immunotherapy, the somatic mutation data from phase 1 study of a monoclonal human antibody to PD-1 (Programmed Death - 1) were used, as single therapy and in combination with other anti-cancer therapies. In total, clinical response data are available from 74 patients with advanced malignancies (n=8 with partial response, PR; n=29 with stable disease, SD; n=37 with progressive disease, PD). Total TMB is denoted as the total number of mutated genes, and driver/passenger TMBs are represented in z-scores in order to normalize for the total TMB of the patients. Z-scores are computed by comparing the number of mutated driver/passenger genes to the background distribution using randomly selected genes of equal size. Higher z-scores indicate higher mutation burden in the selected driver/passenger gene sets despite the total TMB background. The total TMB and z-score of driver TMB do not differentiate PR from other patient groups, while the passenger TMB z-scores in PR patients are significantly higher as compared to those in PD or PD+SD patient groups. Results are consistent in TMBs computed using top 50, 100, 500 (Fig. 12) and 1000 passenger genes.

We claim:

1. A method of treating a cancer patient with an immunotherapy comprising

determining if a cancer patient is an immunotherapy responder by,

establishing a total passenger gene mutation burden of the tumor of the patient;

generating a background distribution for the mutational burden of the tumor;

normalizing the total passenger gene mutation burden against the background distribution;

categorizing the cancer patient as an immunotherapy responder genotype when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution; and

administering an immunotherapy to the cancer patient categorized as an immunotherapy responder.

2. The method according to claim 1, wherein generating a background distribution comprises establishing the mutational burden from a plurality of samples of randomly selected genes obtained from the tumor, provided that the number of randomly selected genes in each sample is equal to the number of passenger genes used to compute the total passenger gene mutation burden.

3. The method according to claim 1 or 2, wherein normalizing the total passenger gene mutation burden against the background distribution comprises generating a z-score indicating the number of standard deviations from the mean of the background distribution.

4. The method according to any one of claims 1 to 3, further comprising categorizing a mutated gene in the tumor as a passenger gene.

5. The method according to claim 4, wherein categorizing a mutated gene in the tumor as a passenger gene comprises selecting a mutated gene from the tumor

and matching the mutated gene to a data structure comprising passenger genes established according to a passenger gene index.

6. The method according to claim 5, wherein the passenger gene index comprises a correlation coefficient between the fraction of samples comprising the mutated gene obtained from a cancer patient cohort and the median number of mutated genes in each type of tumor within the cancer patient cohort.

7. The method according to any one of claims 1-6, wherein the immunotherapy comprises an inhibitor of a T cell inhibitory receptor.

8. The method according to any one of claims 1-7, wherein the immunotherapy comprises an antibody that binds to PD1.

9. The method according to claim 8, wherein the antibody that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22

10. The method according to any one of claims 8 to 9, wherein the immunotherapy further comprises an antibody that binds to LAG-3.

11. The method according to claim 10, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

12. The method according to any one of claims 1-7, wherein the immunotherapy comprises an antibody that binds to PDL1.

13. The method according to claim 12, wherein the antibody that binds to PDL1 comprises at least the HCVR sequence of SEQ ID NO:122 and the LCVR sequence of SEQ ID NO:123.

14. The method according to any one of claims 12 to 13, wherein the immunotherapy further comprises an antibody that binds to LAG-3.

15. The method according to claim 14, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

16. The method according to any one of claims 1-7, wherein the immunotherapy comprises an antibody that binds to CTLA4.

17. The method according to claim 1, wherein determining if a cancer patient is an immunotherapy responder further comprises receiving genetic sequence data, wherein the genetic sequence data comprises a plurality of genes and is derived from the tumor of the patient.

18. The method according to any one of claims 1-7, wherein the immunotherapy comprises an activator of T cell activating receptor.

19. The method according to claim 1, wherein a cancer patient has cutaneous squamous cell cancer (CSCC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), colon/rectum adenocarcinoma (CORE), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), liver hepatocellular carcinoma (LIHC), brain lower grade glioma (LGG), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), ovarian serous cystadenocarcinoma (OV), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), skin cutaneous melanoma (SKCM), or stomach adenocarcinoma.

20. The method according to 19, wherein the immunotherapy is cemiplimab, nivolumab, pembrolizumab, atezolizumab, durvalumab, avelumab, ipilimumab, IFN-alpha, IL-2, or a combination thereof.

21. The method according to 20, wherein the cancer patient has CSCC.

22. The method of claim 21, wherein the immunotherapy is cemiplimab.

23. A method for selecting a cancer patient for immunotherapy, comprising:

establishing a total passenger gene mutation burden from a tumor of a cancer patient;

generating a background distribution for the mutational burden of the tumor;

normalizing the total passenger gene mutation burden against the background distribution; and

categorizing the cancer patient as an immunotherapy responder when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution.

24. The method according to claim 23, wherein generating a background distribution comprises establishing the mutational burden from a plurality of samples of randomly selected genes obtained from the tumor, provided that the number of randomly selected genes in each sample is equal to the number of passenger genes used to compute the total passenger gene mutation burden.

25. The method according to claim 23 or 24, wherein normalizing the total passenger gene mutation burden against the background distribution comprises generating a z-score indicating the number of standard deviations from the mean of the background distribution.

26. The method according to any one of claims 23 to 25, further comprising categorizing a mutated gene in the tumor as a passenger gene.

27. The method according to claim 26, wherein categorizing a mutated gene in the tumor as a passenger gene comprises selecting a mutated gene from the tumor and matching the mutated gene to a data structure comprising passenger genes established according to a passenger gene index.

28. The method according to claim 27, wherein the passenger gene index comprises a correlation coefficient between the fraction of samples comprising the mutated gene obtained from a cancer patient cohort and the median number of mutated genes in each type of tumor within the cancer patient cohort.

29. The method according to any one of claims 23-2328 wherein the immunotherapy comprises an inhibitor of a T cell inhibitory receptor.

30. The method according to any one of claims 23-29, wherein the immunotherapy comprises an antibody that binds to PD1.

31. The method according to claim 30, wherein the antibody that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22

32. The method according to any one of claims 30 to 31, wherein the immunotherapy further comprises an antibody that binds to LAG-3.

33. The method according to claim 32, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

34. The method according to any one of claims 23-29, wherein the immunotherapy comprises an antibody that binds to PDL1.

35. The method according to claim 34, wherein the antibody that binds to PDL1 comprises at least the HCVR sequence of SEQ ID NO:122 and the LCVR sequence of SEQ ID NO:123.

36. The method according to any one of claims 34 to 35, wherein the immunotherapy further comprises an antibody that binds to LAG-3.

37. The method according to claim 36, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

38. The method according to any one of claims 23-29, wherein the immunotherapy comprises an antibody that binds to CTLA4.

39. The method according to any one of claims 23-29, wherein the immunotherapy comprises an activator of T cell activating receptor.

40. A computer readable medium comprising processor-executable instructions adapted to cause one or more computing devices to:

establish a total passenger gene mutation burden from a tumor of a cancer patient;

generating a background distribution for the mutational burden of the tumor;

normalize the total passenger gene mutation burden against the background distribution; and

categorize the cancer patient as an immunotherapy responder when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution.

41. The computer readable medium according to claim 40, wherein the processor-executable instructions adapted to cause one or more computing devices to generate a background distribution comprise processor-executable instructions adapted to cause one or more computing devices to establish the mutational burden from a plurality of samples of randomly selected genes obtained from the tumor, provided that the number of randomly selected genes in each sample is equal to the number of passenger genes used to compute the total passenger gene mutation burden.

42. The computer readable medium according to claim 40 or 41, wherein the processor-executable instructions adapted to cause one or more computing devices to normalize the total passenger gene mutation burden against the background distribution comprise processor-executable instructions adapted to cause one or more computing devices to generate a z-score indicating the number of standard deviations from the mean of the background distribution.

43. The computer readable medium according to any one of claims 40 to 42, wherein the processor-executable instructions are further adapted to cause one or more computing devices to categorize the cancer patient as an immunotherapy responder when the total passenger gene mutation burden is at least about two standard deviations greater than the mean of the background distribution.

44. The computer readable medium according to any one of claims 40 to 43, further comprising processor-executable instructions adapted to cause one or more computing devices to categorize a mutated gene in the tumor as a passenger gene.

45. The computer readable medium according to claim 44, wherein the processor-executable instructions adapted to cause one or more computing devices to categorize a mutated gene in the tumor as a passenger gene comprise processor-executable instructions adapted to cause one or more computing devices to

select a mutated gene from the tumor and match the mutated gene to a data structure comprising passenger genes established according to a passenger gene index.

46. The computer readable medium according to claim 45, wherein the passenger gene index comprises a correlation coefficient between the fraction of samples comprising the mutated gene obtained from a cancer patient cohort and the median number of mutated genes in each type of tumor within the cancer patient cohort.

47. The computer readable medium according to any one of claims 40-46, wherein the cancer patient is categorized as an immunotherapy responder, and wherein the processor-executable instructions are further adapted to cause one or more computing devices to recommend administering to the cancer patient an immunotherapy regimen.

48. The computer readable medium according to claim 47, wherein the immunotherapy regimen comprises administering to the patient an inhibitor of a T cell inhibitory receptor.

49. The computer readable medium according to claim 47 or 48, wherein the immunotherapy regimen comprises administering to the patient an antibody that binds to PD1.

50. The computer readable medium according to claim 49, wherein the antibody that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22

51. The computer readable medium according to any one of claims 49 to 50, wherein the immunotherapy regimen further comprises administering to the patient an antibody that binds to LAG-3.

52. The computer readable medium according to claim 51, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

53. The computer readable medium according to claim 47 or 48, wherein the immunotherapy regimen comprises administering to the patient an antibody that binds to PDL1.

54. The computer readable medium according to claim 53, wherein the antibody that binds to PDL1 comprises at least the HCVR sequence of SEQ ID NO:122 and the LCVR sequence of SEQ ID NO:123.

55. The computer readable medium according to any one of claims 53 to 54, wherein the immunotherapy regimen further comprises administering to the patient an antibody that binds to LAG-3.

56. The computer readable medium according to claim 55, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

57. The computer readable medium according to claim 47 or 48, wherein the immunotherapy regimen comprises administering to the patient an antibody that binds to CTLA4.

58. The computer readable medium according to claim 47, wherein the immunotherapy regimen comprises administering to the patient activator of T cell activating receptor.

59. An apparatus, comprising:

one or more processors; and

a memory storing processor executable instructions that, when executed by the one or more processors, cause the apparatus to:

establish a total passenger gene mutation burden from a tumor of a cancer patient;

generating a background distribution for the mutational burden of the tumor;

normalize the total passenger gene mutation burden against the background distribution; and

categorize the cancer patient as an immunotherapy responder when the total passenger gene mutation burden is at least about one and a half standard deviations greater than the mean of the background distribution.

60. The apparatus according to claim 59, wherein the processor executable instructions that, when executed by the one or more processors, cause the apparatus to generate a background distribution comprise processor executable instructions that, when executed by the one or more processors, cause the apparatus to establish the mutational burden from a plurality of samples of randomly selected genes obtained from the tumor, provided that the number of randomly selected genes in each sample is equal to the number of passenger genes used to compute the total passenger gene mutation burden.

61. The apparatus according to claim 59 or 60, wherein the processor executable instructions that, when executed by the one or more processors, cause the apparatus to normalize the total passenger gene mutation burden against the background distribution comprise processor executable instructions that, when executed by the one or more processors, cause the apparatus to generate a z-score indicating the number of standard deviations from the mean of the background distribution.

62. The apparatus according to any one of claims 59 to 61, wherein the processor executable instructions, when executed by the one or more processors, cause the apparatus to categorize the cancer patient as an immunotherapy responder when the total passenger gene mutation burden is at least about two standard deviations greater than the mean of the background distribution.

63. The apparatus according to any one of claims 59 to 62, wherein the processor executable instructions, when executed by the one or more processors, further cause the apparatus to categorize a mutated gene in the tumor as a passenger gene.

64. The apparatus according to claim 63, wherein the processor executable instructions, when executed by the one or more processors, cause the apparatus to categorize a mutated gene in the tumor as a passenger gene comprise

processor executable instructions that, when executed by the one or more processors, cause the apparatus to select a mutated gene from the tumor and matching the mutated gene to a data structure comprising passenger genes established according to a passenger gene index.

65. The apparatus according to claim 64, wherein the passenger gene index comprises a correlation coefficient between the fraction of samples comprising the mutated gene obtained from a cancer patient cohort and the median number of mutated genes in each type of tumor within the cancer patient cohort.

66. The apparatus according to any one of claims 59 to 65, wherein the cancer patient is categorized as an immunotherapy responder, and the processor executable instructions, when executed by the one or more processors, cause the apparatus to recommend administering to the cancer patient an immunotherapy regimen.

67. The apparatus according to claim 66, wherein the immunotherapy regimen comprises administering to the patient an inhibitor of a T cell inhibitory receptor.

68. The apparatus according to claim 66 or 67, wherein the immunotherapy regimen comprises administering to the patient an antibody that binds to PD1.

69. The apparatus according to claim 68, wherein the antibody that binds to PD1 comprises at least the heavy chain variable region (HCVR) sequence of SEQ ID NO:21 and the light chain variable region (LCVR) sequence of SEQ ID NO:22

70. The apparatus according to any one of claims 68 to 69, wherein the immunotherapy regimen further comprises administering to the patient an antibody that binds to LAG-3.

71. The apparatus according to claim 70, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

72. The apparatus according to claim 66 or 67, wherein the immunotherapy regimen comprises administering to the patient an antibody that binds to PDL1.

73. The apparatus according to claim 72, wherein the antibody that binds to PDL1 comprises at least the HCVR sequence of SEQ ID NO:122 and the LCVR sequence of SEQ ID NO:123.

74. The apparatus according to any one of claims 72 to 73, wherein the immunotherapy regimen further comprises administering to the patient an antibody that binds to LAG-3.

75. The apparatus according to claim 74, wherein the antibody that binds to LAG-3 comprises at least the HCVR sequence of SEQ ID NO:93 and the LCVR sequence of SEQ ID NO:94.

76. The apparatus according to claim 66 or 67, wherein the immunotherapy regimen comprises administering to the patient an antibody that binds to CTLA4.

77. The apparatus according to claim 66, wherein the immunotherapy regimen comprises administering to the patient activator of T cell activating receptor.