

## **SYSTEMS AND METHODS FOR CLINICAL DECISION SUPPORT**

5

### **FIELD OF THE DISCLOSURE**

[0001] The present disclosure relates to systems and methods for bioinformatics and data processing. In particular, in a first aspect, the present disclosure relates to methods and systems for generating a personalized  
10 treatment guideline for a patient and for selecting a treatment for a patient. In another aspect, the present disclosure relates to methods and systems for selecting patients for a clinical trial of a treatment.

### **BACKGROUND OF THE DISCLOSURE**

[0002] A large number of publications exist regarding human disease  
15 etiology and progression, discussing various molecular entities such as proteins, small molecules such as metabolites, nutrients, drugs, transporters, enzymes, pathways, and other information. Additionally, with revolutionary advances occurring in profiling technologies, the amount of new literature is constantly increasing. With such a large mass of data, it may be difficult for researchers to  
20 easily and quickly perform analyses and for clinicians to identify personalized patient treatment options. With such a large mass of data, it may also be difficult for researchers and clinicians to select patients for a clinical trial of a given treatment.

[0003]        There exist attempts to combine knowledge about the mode of action of treatments, specifically the targets of targeted drugs, with the results of molecular profiling of a patient in order to select an appropriate treatment. In "Personalized Medicine in a Phase I Clinical Trials Program: the MD Anderson Cancer Center Initiative" by Tsimberidou, Apostolia-Maria, Nancy G Iskander, David S Hong, Jennifer J Wheler, Gerald S Falchook, Siqing Fu, Sarina Piha-Paul, et al. (Clinical Cancer Research: an Official Journal of the American Association for Cancer Research 18, no. 22 (November 15, 2012): 6373–6383. doi:10.1158/1078-0432.CCR-12-1627), the authors test a small set of drug-target- genes for a pre-defined set of aberrations, specifically mutations. Treatments are selected that target a protein encoded by a gene with an aberration. However, their data shows that more than 7% of the patients had more than one "actionable" aberration. With more extensive profiling this percentage is expected to increase. In such cases, it is not clear, which protein / aberration should best be targeted by the treatment to be selected.

#### **BRIEF SUMMARY OF THE DISCLOSURE**

[0004]        According to a first aspect of the invention, a method for generating a personalized treatment guideline for a patient, comprising the steps:

                 retrieving an identification of a disease indication of the patient as a user's input, from an electronic health record or electronic medical record or from a patient database comprising medical records of a candidate set of patients,

                 assigning an indication-specific score to each gene or protein of a first set of genes or proteins, the first set of genes or proteins being part of the

human microbiome, the indication-specific score representing an evidence of the gene or protein being associated with the disease indication of the patient,

wherein the indication-specific score of each gene or protein is computed by first assigning numerical values to attributes indicating degrees of potential relationship to the indication and then by forming a weighted sum of the numerical values or of any monotonic transformation of the numerical values over the attributes, the attributes comprising at least one of the following list:

- a) The gene or protein is a drug target,
- b) The gene or protein is a biomarker or part of a biomarker,
- 10 c) The gene or protein is disease-associated,
- d) The gene or protein is an oncogene or a product of an oncogene, respectively,
- e) The gene or protein is a tumor suppressor,
- f) The gene or protein has a cancer pathway association,
- 15 g) The gene or protein is gene ontology annotated for cancer relevant processes,
- h) The gene or protein is part of a cancer-associated gene fusion,
- i) The gene or protein has a tractable domain,
- 20 j) The gene or protein is embryonic lethal, and

k) The gene or protein is highly mutated in the specific indication;

retrieving a molecular profile of the patient, wherein the molecular profile is a set of molecular measurements of DNA or RNA extracted from the patient, indexed by a second set of genes or proteins, the second set of genes or  
5 proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins,

inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations as the differences of the molecular profile to a reference profile from a control sample of the patient  
10 or from other individuals,

assigning a profile-specific score to each gene or protein of the first set of genes or proteins, the profile-specific score representing a degree of how much the function of the genes or proteins is altered by the aberrations,

wherein assigning the profile-specific score to each gene or protein  
15 of the first set of genes or proteins comprises the following steps:

computing an aberration-specific functional impact score for each aberration taking into account the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the aberration-specific functional impact score correspond to an activating and  
20 inactivating impact, respectively, and

computing a protein- or gene-specific direct functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a



maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with the gene or protein, or

computing a protein- or gene-specific indirect functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a  
5 maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with another set of genes or proteins, the another set including the gene or protein and a set of genes or proteins that are upstream of the gene or protein according to a molecular interaction network, the molecular interaction network being a directed graph, consisting of nodes each of which  
10 represents a gene or a protein, and of edges each of which connects two nodes,

and wherein the profile-specific score is set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these,

assigning a patient-specific score to each gene or protein of the first  
15 set of genes or proteins, the patient-specific score of each gene or protein being a sum, a weighted sum or a product of its indication-specific score and its profile-specific score,

retrieving a set of targeted treatments, together with associated target genes or proteins, as a user's input or by selecting treatments that are  
20 approved or under investigation for the indication of the patient or for related indications from a treatment database, the treatment database storing treatments, in particular drugs and drug combinations, together with information on the genes or proteins that are the targets of the drugs,

assigning a score to each targeted treatment from the set of targeted treatments by calculating a sum or a weighted sum over the patient-specific scores of the associated target genes or proteins, and

generating the personalized treatment guideline as an ordered list of  
5 the targeted treatments, wherein the order of the targeted treatments is determined by their score

**[0005]** Also disclosed herein is a method for generating a personalized treatment guideline for a patient comprises retrieving an identification of a disease indication of the patient. The method further comprises assigning an indication-  
10 specific score to each gene or protein of a first set of genes or proteins, the indication-specific score reflecting an evidence of the gene or protein being associated with the disease indication of the patient. The method further comprises retrieving a molecular profile of the patient, wherein the molecular profile is a set of molecular measurements indexed by a second set of genes or  
15 proteins, the second set of genes or proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins. The method further comprises inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations. The method further comprises assigning a profile-specific score to each gene or protein of the  
20 first set of genes or proteins, the profile-specific score reflecting a degree of how much the function of the genes or proteins is altered by the aberrations. The method further comprises assigning a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being based on both its indication-specific score and its profile-specific  
25 score. The method further comprises retrieving a set of targeted treatments,

together with associated target proteins and/or the corresponding genes, from a treatment database. The method further comprises assigning a score to each targeted treatment from the set of targeted treatments by aggregating the patient-specific scores at least over its associated target proteins and/or corresponding  
5 genes that are in the first set of genes or proteins. The method further comprises generating the personalized treatment guideline as an ordered list of the targeted treatments, wherein the order of the targeted treatments is determined by their score.

**[0006]** In comparison to existing methods for treatment selection and/or  
10 assignment of patients to trials, the approach outlined above has qualitative advantages. The invention also allows to prioritize the most important mutations in the mechanistically most important proteins, thus providing a new level of actionability beyond established biomarkers. In a patient where no previously described biomarker is found, this approach may be used to define the most likely  
15 actionable novel biomarker in the patient.

**[0007]** One advantage over the approach described by Tsimberidou et al. is that it provides a principled, bio-medically founded solution to the case of molecular aberrations being found in more than one gene or protein. This case is expected to be the standard case, once the profiling is performed to cover a large  
20 number of genes and/or proteins, which is desirable with respect to a comprehensive assessment and understanding of the patient and his/her disease. The method further comprises assigning a profile-specific score to each gene or protein of the first set of genes or proteins, the profile-specific score reflecting a degree of how much the function of the genes or proteins is altered as compared  
25 to a fixed reference state. Specifically, this score allows for going beyond a simple

binary attribution of genes and/or proteins as either being mutated or not. In some embodiments, this invention further allows to match treatments to aberrations based on indirect, so-called downstream effects, hence providing treatment selection guidance even in cases in which no treatment directly targets a gene or  
5 protein with an aberration.

**[0008]** The invention thus resolves cases in which patients have more than one "actionable" aberration by combining the patient-specific molecular information and the treatment-specific molecular information further with a clinico-molecular disease model, specifically a scoring of genes and/or proteins that  
10 represents several aspects of their involvement into the disease. In this way, treatments can be prioritized that are most likely to impact the disease mechanism.

**[0009]** While much of this invention is described in the context of oncology, which is a preferred area of its application, it should be understood that the  
15 invention is of a general nature that extends to all kinds of diseases and conditions that have a molecular foundation (as opposed to, for instance, bone fractures, which are physical or mechanical in nature).

**[0010]** In embodiments of this invention that are directed towards treatment ranking or prioritization, an important choice is the selection of a set of treatments  
20 to be ranked. An obvious choice is to consider all treatments that are approved for the indication of the present patient. An alternative choice is to consider all treatments that are either approved or under development for that indication. In some application areas, for instance some rare cancers, there may not exist any approved treatments, or the set of existing approved treatments may already be

exhausted for a given patient, meaning that these treatments have been tried unsuccessfully or that they are not applicable due to some counter-indication like an idiosyncrasy or a toxic interaction with another vital medication. Then it can be useful to consider treatments that are approved for a broader disease indication area, which is here defined as a set of related indications. Relatedness of indications may be judged according to a disease ontology, for instance MeSH or MedDRA. The necessary and sufficient degree of relatedness of indications may depend on the numbers of treatments available for related indications.

**[0011]** According to a first embodiment, the treatments are manually entered. According to another embodiment, the treatments are selected based on the disease indication of the patient or on a disease indication area of the patient.

**[0012]** The method may further comprise a step of outputting the personalized treatment guideline to a user as a list comprising one or more of the highest ordered targeted treatments, in particular by displaying them on a screen or display or by printing them onto paper.

**[0013]** According to a variant, a method for selecting a treatment for a patient comprises the steps: generating a personalized treatment guideline for the patient by executing the method steps as specified herein, and selecting the first targeted treatment from the ordered list of the targeted treatments of the personalized treatment guideline. The method may also comprise administering a selected treatment to the patient.

**[0013A]** According to a second aspect of the invention, a method for selecting a treatment for a patient, comprising the steps:

generating a personalized treatment guideline for the patient by  
executing the method steps described in the first aspect, and

selecting the first targeted treatment from the ordered list of the  
targeted treatments of the personalized treatment guideline

- 5    **[0013B]**      According to a third aspect there is provided a method for selecting  
patients for a clinical trial of a treatment, comprising the steps:

retrieving identifications of disease indications of a set of patients as  
a user's input, from an electronic health record EHR or electronic medical record  
EMR or from a patient database comprising medical records of a candidate set of  
10    patients,

assigning, for every patient, an indication-specific score to each gene  
or protein of a first set of genes or proteins, the first set of genes or proteins being  
part of the human microbiome, the indication-specific score representing an  
evidence of the gene or protein being associated with the disease indication of the  
15    patient,

wherein the indication-specific score of each gene or protein is  
computed by first assigning numerical values to attributes indicating degrees of  
potential relationship to the indication and then by forming a weighted sum of the  
numerical values or of any monotonic transformation of the numerical values over  
20    the attributes, the attributes comprising at least one of the following list:

- a)    The gene or protein is a drug target,
- b)    The gene or protein is a biomarker or part of a biomarker,
- c)    The gene or protein is disease-associated,

- d) The gene or protein is an oncogene or a product of an oncogene, respectively,
  - e) The gene or protein is a tumor suppressor,
  - f) The gene or protein has a cancer pathway association,
  - 5 g) The gene or protein is gene ontology annotated for cancer relevant processes,
  - h) The gene or protein is part of a cancer-associated gene fusion,
  - i) The gene or protein has a tractable domain,
  - 10 j) The gene or protein is embryonic lethal, and
  - k) The gene or protein is highly mutated in the specific indication;
- retrieving, for every patient, a molecular profile of the patient, wherein the molecular profile is a set of molecular measurements of DNA or RNA extracted from the patient, indexed by a second set of genes or proteins, the
- 15 second set of genes or proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins,
- inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations as the differences of the molecular profile to a reference profile from a control sample of the patient
- 20 or from other individuals,

assigning, for every patient, a profile-specific score to each gene or protein of the first set of genes or proteins, the profile-specific score representing a degree of how much the function of the genes or proteins is altered by the aberrations,

5                    wherein assigning the profile-specific score to each gene or protein of the first set of genes or proteins comprises the following steps:

                  computing an aberration-specific functional impact score for each aberration taking into account the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the  
10    aberration-specific functional impact score correspond to an activating and inactivating impact, respectively, and

                  computing a protein- or gene-specific direct functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores  
15    over the aberrations associated with the gene or protein, or

                  computing a protein- or gene-specific indirect functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with another set of genes or proteins, the another  
20    set including the gene or protein and a set of genes or proteins that are upstream of the gene or protein according to a molecular interaction network, the molecular interaction network being a directed graph, consisting of nodes each of which represents a gene or a protein, and of edges each of which connects two nodes,



and wherein the profile-specific score is set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these,

5 assigning, for every patient, a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being a sum, a weighted sum or a product of its indication-specific score and its profile-specific score,

retrieving, for the treatment, associated target genes or proteins from a treatment database, the treatment database storing treatments, in particular  
10 drugs and drug combinations, together with information on the genes or proteins that are the targets of the drugs, which are also in the first set of genes or proteins,

assigning a score to every patient from the set of patients by calculating a sum or a weighted sum over the patient-specific scores of the associated target genes or proteins, and

15 generating an ordered list of the patients, wherein the order of the patients is determined by their score.

**[0013C]** According to a fourth aspect there is provided a system for generating a personalized treatment guideline for a patient and/or for selecting patients for a clinical trial of a treatment, comprising a processing unit with:

20 an interface for providing a first set of genes or proteins to the system (1), the first set of genes or proteins being part of the human microbiome,

an interface for providing identifications of disease indications of patients to the system,

a module for assigning, for an identification of a disease indication of a patient, an indication-specific score to each gene or protein of the first set of  
5 genes or proteins, the indication-specific score representing an evidence of the gene or protein being associated with the disease indication of the patient,

wherein the indication-specific score of each gene or protein is computed by first assigning numerical values to attributes indicating degrees of potential relationship to the indication and then by forming a weighted sum of the  
10 numerical values or of any monotonic transformation of the numerical values over the attributes, the attributes comprising at least one of the following list:

- a) The gene or protein is a drug target,
- b) The gene or protein is a biomarker or part of a biomarker,
- c) The gene or protein is disease-associated,
- 15 d) The gene or protein is an oncogene or a product of an oncogene, respectively,
- e) The gene or protein is a tumor suppressor,
- f) The gene or protein has a cancer pathway association,
- g) The gene or protein is gene ontology annotated for cancer  
20 relevant processes,

h) The gene or protein is part of a cancer-associated gene fusion,

i) The gene or protein has a tractable domain,

j) The gene or protein is embryonic lethal, and

5 k) The gene or protein is highly mutated in the specific indication;

an interface for providing molecular profiles of patients to the system, wherein a molecular profile is a set of molecular measurements of DNA or RNA extracted from a patient, indexed by a second set of genes or proteins, the second set of genes or proteins being equal to the first set of genes or proteins or being a  
10 subset of the first set of genes or proteins,

a module for assigning, for a molecular profile of a patient, a profile-specific score to each gene or protein of the first set of genes or proteins, wherein the module for assigning the profile-specific score is adapted to inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a  
15 possibly empty set of aberrations as the differences of the molecular profile to a reference profile from a control sample of the patient or from other individuals,

wherein assigning the profile-specific score to each gene or protein of the first set of genes or proteins comprises the following steps:

computing an aberration-specific functional impact score for each  
20 aberration taking into account the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the

aberration-specific functional impact score correspond to an activating and inactivating impact, respectively, and

computing a protein- or gene-specific direct functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores  
5 over the aberrations associated with the gene or protein, or

computing a protein- or gene-specific indirect functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores  
10 over the aberrations associated with another set of genes or proteins, the another set including the gene or protein and a set of genes or proteins that are upstream of the gene or protein according to a molecular interaction network, the molecular interaction network being a directed graph, consisting of nodes each of which represents a gene or a protein, and of edges each of which connects two nodes,

15 and wherein the profile-specific score is set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these,

a module for assigning a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene  
20 or protein being a sum, a weighted sum or a product of its indication-specific score and its profile-specific score,

an interface for providing targeted treatments, together with associated target genes or proteins, as a user's input or by selecting treatments

that are approved or under investigation for the indication of the patient or for related indications from a treatment database, the treatment database storing treatments, in particular drugs and drug combinations, together with information on the genes or proteins that are the targets of the drugs, to the system, and

5                   a module for assigning a score to patients and to treatments by calculating a sum or a weighted sum over the patient-specific scores of the associated target genes or proteins, wherein the module for assigning the score to patients and to treatments is adapted to generating ordered lists of patients or treatments, wherein the order of the patients or treatments is determined by their  
10   score.

**[0013D]**       Also disclosed herein is a method for selecting patients for a clinical trial of a treatment comprises retrieving identifications of disease indications of a set of patients. The method further comprises assigning, for every patient, an indication-specific score to each gene or protein of a first set of genes or proteins,  
15   the indication-specific score reflecting an evidence of the gene or protein being associated with the disease indication of the patient. The method further comprises retrieving, for every patient, a molecular profile of the patient, wherein the molecular profile is a set of molecular measurements indexed by a second set of genes or proteins, the second set of genes or proteins being equal to the first  
20   set of genes or proteins or being a subset of the first set of genes or proteins. The method further comprises inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations. The method further comprises assigning, for every patient, a profile-specific score to each gene or protein of the first set of genes or proteins, the profile-specific score  
25   reflecting a degree of how much the function of the genes or

proteins is altered by the aberrations. The method further comprises assigning, for every patient, a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being based on both its indication-specific score and its profile-specific score. The method

5 further comprises retrieving, from a treatment database, for the treatment those of its associated target proteins and/or corresponding genes that are also in the first set of genes and proteins. The method further comprises assigning a score to every patient from the set of patients by aggregating the patient-specific scores at least over the associated target proteins and/or the corresponding genes, and

10 generating an ordered list of the patients, wherein the order of the patients is determined by their score.

**[0014]** The inventive methods denoted by first and to second aspects only differ from their input. In the first method, a set of treatments is being related to a single patient, whereas in the second method, a set of patients is being related to

15 a single treatment. The involved scores may be computed analogously in both cases.

**[0015]** The indication-specific score of any gene or protein is computed by first assigning numerical values to attributes that indicate degrees of potential relationship of the gene or protein to the disease indication and then aggregating

20 the numerical values over the attributes, in particular by forming a weighted sum of the values or of any monotonic transformation of the values, such as logarithms or exponentials. The indication-specific score reflects the evidence of the gene or protein being associated with the disease indication of the patient.

**[0016]** The indication-specific score may be based on scores assigned to at least one of the following attributes labeled by a) to k).

**[0017]** a) The gene or protein is a drug target.

The attribute “drug target” may involve a-specific sub-scoring schema involving further attributes. The score for the attribute “drug target”, also referred to as a drug target score in the following, may be a sum, a product, or any function of sub-scores for the involved attributes. A first relevant attribute may be the relation of the drug to the indication. If the gene or protein is a drug target for a drug used in the-specific indication, then the drug target score may be attributed a high value. If the gene or protein is a drug target in a related indication, then the drug target score may be attributed a medium value. If the gene or protein is a drug target in an unrelated indication, then the drug target score may be attributed a low value. A second relevant attribute may be the developmental stage of the drug, e.g. experimental or approved. A low drug target score is related to the experimental stage and a high drug target score is related to the approved stage. A third relevant attribute may be a total number of targeting drugs, the drug target score thus being a function of the total number of targeting drugs. A fourth relevant attribute may be if the drug acts directly or indirectly on its target, a direct acting being related with a higher drug target score than an indirect acting. A fifth relevant attribute may be related to trials, in particular the total number of trials and the phase of a trial.

**[0018]** b) The gene or protein is a biomarker or part of a biomarker.

Biomarkers are values that indicate diagnosis, prognosis, response to treatment, or other clinically relevant information and that can be directly measured or

computed from measurement values. A gene or protein is said to be a biomarker if the measurement of a value directly related to that gene or protein, for instance its expression level or its sequence, is a biomarker. Analogously, a gene or protein is said to be part of a biomarker when a measurement value directly related to it is  
5 required for the computation of a biomarker that may also involve other values. Similarly to the drug target score, the biomarker score of a gene or protein may be formed by combining several sub-scores that relate to attributes of the biomarker. Specifically, there may be sub-scores that reflect: the level of validation of the biomarker; the strength of the biomarker, in other words the effect size associated  
10 with a change in the biomarker value; and the relationship between the indications that the biomarker is relevant for and the patient indication.

**[0019]**        c)        The gene or protein is disease-associated.

The attribute “disease-associated” may involve another specific sub-scoring schema involving further attributes. The score for the attribute “disease  
15 associated”, also referred to as a disease associated score in the following, may be a sum, a product or any function of sub-scores for the involved attributes. A first attribute may be if the gene or protein is part of an Online Mendelian Inheritance in Man (OMIM). A second attribute may be if the gene or protein is part of a disease ontology. A third attribute may be if the gene or protein has a text data mining co-  
20 occurrence with the indication, a related disease indication or an unrelated indication. For all of these attributes, the sub-score may depend on whether the gene or protein is associated with the specific indication, a related disease indication or an unrelated indication. If the gene or protein is associated with the specific indication, then the disease associated score may be attributed a high  
25 value. If the gene or protein is associated with a related indication, then the



disease associated score may be attributed a medium value. If the gene or protein is associated with an unrelated indication, then the disease associated score may be attributed a low value. A fourth attribute may involve text data mining (TDM) metrics and statistics. A fifth attribute may be a validity, e.g. cell line, animal or  
5 clinical study.

**[0020]** d) The gene or protein is an oncogene or a product of an oncogene, respectively.

The attribute “oncogene” is associated with a fixed numerical contribution to the indication-specific score.

10 **[0021]** e) The gene or protein is a tumor suppressor.

The attribute “tumor suppressor” is associated with a fixed numerical contribution to the indication-specific score.

**[0022]** f) The gene or protein has a cancer pathway association.

The attribute “cancer pathway association” may involve counting of memberships  
15 in pathways. A specific cancer pathway association may be the Vogelstein core cancer association.

**[0023]** g) The gene or protein is gene ontology annotated for cancer relevant processes.

The attribute “gene ontology annotated for cancer relevant processes” is  
20 associated with a fixed numerical contribution to the indication-specific score.

**[0024]** h) The gene or protein is part of a cancer-associated gene fusion.

The attribute “part of a cancer-associated gene fusion” is associated with a fixed numerical contribution to the indication-specific score.

**[0025]** i) The gene or protein has a tractable domain.

With regard to “tractable domains”, their contribution to the indication-specific  
5 score may be high if there are one or more tractable domains known, medium if they potentially exist and low if there are none.

**[0026]** j) The gene or protein is embryonic lethal.

The attribute “embryonic lethal” is associated with a fixed numerical contribution to the indication-specific score.

10 **[0027]** k) The gene or protein is highly mutated in the specific indication.

The attribute “Indication specific mutation rate” is associated with a fixed numerical contribution to the indication-specific score.

**[0028]** It is useful to define the indication-specific score such that it assumes only non-negative values; this can be achieved by defining it as a weighted sum of  
15 non-negative attribute scores with non-negative weights.

**[0029]** According to an embodiment, the indication-specific score is based on scores assigned to at least one of the attributes labeled by a) to c), which are not specific for cancer.

**[0030]** According to an embodiment of the method, the patient-specific  
20 score is a sum or a weighted sum of its indication-specific score and its profile-specific score.

**[0031]** It is preferred, however, that the patient-specific score is a product of its indication-specific score and its profile-specific score.

**[0032]** According to an embodiment, the profile-specific score is set to zero if the gene or protein has no aberrations.

5 **[0033]** The profile-specific score may simply be set to 1 if the gene or protein has one or more aberrations or may involve more sophisticated scoring schemes with aberration-specific functional impact scores.

**[0034]** Aberration-specific functional impact scores (aFIS) are scores that quantify the presumed impact of an aberration of a given gene or protein onto its  
10 function. In some embodiments, this score is signed (that is, it may assume negative values), such that the sign indicates the directionality of the impact, with a positive sign indicating an enhancing or activating effect, and a negative sign indicating an inhibitory or deactivating effect. In preferred embodiments, this score is normalized, such that a value of +1 would indicate permanent activation, and -1  
15 would indicate complete deactivation of the function.

**[0035]** According to an embodiment, assigning a profile-specific score to each gene or protein of the first set of genes or proteins comprises computing an aberration-specific functional impact score for each aberration and computing a protein- or gene-specific direct functional impact score for each protein or gene by  
20 aggregating the aberration-specific functional impact score over the aberrations associated with the gene or protein.

**[0036]** Alternatively or additionally, assigning a profile-specific score to each gene or protein of the first set of genes or proteins comprises computing an

aberration-specific functional impact score for each aberration and computing a protein- or gene-specific indirect functional impact score for each protein or gene by aggregating the aberration-specific functional impact score over the aberrations associated with a set of second genes or proteins, the set including the first gene  
5 or protein and a set of genes or proteins that are upstream of the first gene or protein according to a molecular interaction network.

**[0037]** These embodiments that consider effects of aberrations and/or treatments on genes or proteins downstream of the gene or protein that has the aberration or is a target of the treatment, respectively, enable to identify treatments  
10 that cannot be found by the direct matching of treatment targets to aberrations.

**[0038]** The profile-specific score may be set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these.

**[0039]** In some embodiments, the aberration-specific functional impact  
15 score also comprises an indication of the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the aberration-specific functional impact score correspond to an activating and inactivating impact, respectively.

**[0040]** Aggregating the aberration-specific functional impact scores over the  
20 aberrations may be formed by computing a weighted sum of all aberration-specific functional impact scores, in particular the plain sum of all functional impact scores or the mean of all functional impact scores. Alternatively, the aggregate of the aberration-specific functional impact scores may be formed as a weighted sum of all aberration-specific functional impact scores. Alternatively, aggregating the

aberration-specific functional impact score over the aberrations may be formed by taking a quantile of all aberration-specific functional impact scores, in particular the median of all aberration-specific functional impact scores or the maximum of all aberration-specific functional impact scores. Alternatively, aggregating the

5 aberration-specific functional impact score over the aberrations may be formed as the (possibly weighted) geometric mean of all aberration-specific functional impact scores.

**[0041]** If the aberration-specific functional impact scores represent

probabilities of each of the aberrations impacting the function of the gene or

10 protein, alternatively, aggregating the aberration-specific functional impact score over the aberrations may be performed by estimating the probability of the union of all aberrations impacting the function of the gene or protein, which can be implemented using the well-known inclusion-exclusion principle from combinatorics, in particular based on an assumption of mutual independence of  
15 the functional impact of the aberrations.

**[0042]** According to an embodiment, the aggregation of aberration-specific

functional impact scores takes into account an indication of the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the aberration-specific functional impact score correspond to an

20 activating and inactivating impact, respectively.

**[0043]** According to an embodiment, assigning a score to a treatment

further takes into account a directionality of the impact of the treatment on its targets in aggregating over them by adding profile-specific scores of targets that

are inhibited or antagonized by the treatment and subtracting profile-specific scores of targets that are activated or stimulated or agonized by the treatment.

**[0044]** According to an embodiment, assigning a score to each treatment is extended such that patient-specific scores are further aggregated over genes  
5 and/or proteins downstream of its targets according to a molecular interaction network.

**[0045]** In some embodiments, this invention takes advantage of an interaction network of proteins and/or genes. Such networks are frequently represented as graphs, consisting of nodes each of which represents a gene or a  
10 protein, and of edges each of which connects two nodes. In a directed network, edges have directions, pointing from a source node to a sink node. The source is said to be "upstream" of the sink; the sink is said to be "downstream" of the source. The relationships "upstream" and "downstream" are extended transitively: if, for instance, node A is upstream of node B and B is upstream of node C, then A  
15 is upstream of C and C is downstream of A. A directed path from one node X to another node Y is a sequence of nodes starting with X and ending with Y such that for each pair of successive nodes in the sequence there exists a directed edge from the first node in the pair to the second node in the pair.

**[0046]** In preferred embodiment, the molecular interaction network shall be  
20 represented by a directed graph with optional binary edge annotations, the annotations labeling an edge as either activating or inhibitory. An activating edge indicates that an increase in activity of the protein or gene represented by the source node triggers an increase in activity of the protein or gene represented by the sink node; analogously with decrease of activity. An inhibitory edge indicates

that an increase in activity of the protein or gene represented by the source node triggers a decrease in activity of the protein or gene represented by the sink node; and vice versa.

**[0047]** Specifically, in some embodiments, the profile-specific score of a gene or protein is computed by aggregating aberration-specific functional impact scores (aFIS) in a way that takes into account the direction of the impact of an aberration onto the function of that gene or protein. Let X denote the gene or protein for which the profile-specific score shall be computed, then the aggregation is formed as a weighted sum of aFIS of aberrations of X and genes or proteins upstream of X. For any aberration of X itself, the weight is +1. For any aberration of a protein U upstream of X, the weight depends on the number of inhibitory edges in the considered path from U to X: the weight is -1 for odd numbers, and +1 for even numbers. This way, the directionality of the effect of U on X is reflected. In general there may exist more than one directed path from a node U to another node X; a preferred way of dealing with this is to only consider shortest paths, and to only consider upstream nodes U that have a single (unambiguous) shortest path.

**[0048]** According to the invention, a system for generating a personalized treatment guideline for a patient and/or for selecting patients for a clinical trial of a treatment comprises a processing unit with an interface for providing a first set of genes or proteins to the system. The system further comprises an interface for providing identifications of disease indications of patients to the system. The system further comprises a module for assigning, for an identification of a disease indication of a patient, an indication-specific score to each gene or protein of the first set of genes or proteins, the indication-specific score reflecting an evidence of

the gene or protein being associated with the disease indication of the patient. The system further comprises an interface for providing molecular profiles of patients to the system, wherein a molecular profile is a set of molecular measurements indexed by a second set of genes or proteins, the second set of genes or proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins. The system further comprises a module for assigning, for a molecular profile of a patient, a profile-specific score to each gene or protein of the first set of genes or proteins. The module for assigning the profile-specific score is adapted to inferring, from a molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations. The system further comprises a module for assigning a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being based on both its indication-specific score and its profile-specific score. The system comprises an interface for providing targeted treatments, together with associated target proteins and/or the corresponding genes, from a treatment database to the system, wherein the associated target proteins and/or the corresponding genes are another subset of the first set of genes or proteins. The system further comprises a module for assigning a score to patients and to treatments by aggregating the patient-specific scores at least over the associated target proteins and/or the corresponding genes. The module for assigning the score to patients and to treatments is adapted to generating ordered lists of patients or treatments, wherein the order of the patients or treatments is determined by their score.

**[0049]** In an embodiment, the module for assigning the profile-specific score is further adapted to assigning an aberration-specific functional impact score for



each aberration and to aggregating the functional impact score over the aberrations if there are two or more aberrations in the gene or protein and assigning zero to the gene or protein if the gene or protein has no aberrations.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

5   **[0050]**       The foregoing and other objects, aspects, features, and advantages of the disclosure will become more apparent and better understood by referring to the following description taken in conjunction with the accompanying drawings, in which:

**[0051]**       FIG. 1 is a block diagram of a system for generating a personalized  
10 treatment guideline for a patient or for selecting patients for a clinical trial of a treatment according to one implementation,

**[0052]**       FIG. 2 is a flow diagram depicting a method for generating a personalized treatment guideline for a patient according to one implementation,

**[0053]**       FIG. 3 shows an intermediate step for generating a personalized  
15 treatment guideline for a patient according to one implementation and

**[0054]**       FIG. 4 shows a directed graph with labeled nodes and edges, illustrating a part of a molecular interaction network and the corresponding computation of scores.

### **DETAILED DESCRIPTION**

20   **[0055]**       Prior to discussing specifics of methods and systems utilizing prioritization of patient treatment options, it may be helpful to briefly define a few terms as used herein. These definitions are not intended to limit the use of the

terms, but rather may provide additional or alternate definitions for use of the terms within some contexts.

**[0056]** The term “aberration” refers to differences of a specific instance of a gene or protein to a reference version of that gene or protein. In the context of this invention, the specific instances of interest are those found in a given patient. The reference version is intended to represent what is “normal”, i.e. what corresponds to a healthy state, typically with respect to humans in general, a specific human population or cohort, or the given patient. As an example, the reference state may be defined by what is found in a control tissue sample of the patient, whereas the aberrations are looked for in a tumor sample. As another example, the reference state may be defined as what is frequently found in a representative sample of humans. The difference that constitutes an aberration may be with respect to any biochemically or medically relevant property of the gene or protein. Examples include the sequence of a gene or protein, the structure of a protein, the DNA methylation status of a gene, the expression of a gene or protein, and post-translational modifications of a protein. A typical aberration in an ontology context is a somatic mutation, i.e. a sequence alteration in the tumor as compared to the germline

**[0057]** The term “identification of a disease indication” is understood according to a disease ontology, eg MeSH, MedDRA, ICD-9, etc. For certain classes of indications there may also be specialized ontologies that may offer advantages like more precise categorization of the indication. For example, in oncology it may be beneficial to use ICD-O-3 and/or the TNM staging system.

**[0058]** The term “molecular profile” denotes a set of measurements indexed by a set of genes. The set of measurements is thus formed by measurements, wherein each measurement is associated with a gene of the human genome or a protein of the human proteome. The human genome and proteome are meant to include the genomes and proteomes of human symbionts, like gut bacteria.

Examples of molecular profiles include the sequences of genes or, equivalently, the sets of sequence variants. Examples of molecular profiles further include the expression levels of genes and the expression levels of proteins. They can be measured by standard technology, including next generation sequencing (NGS) or microarrays. The measurement values are stored in a database or in files, from where they can easily be retrieved.

**[0059]** The term “set of genes or proteins” may generally denote a set of genes, a set of proteins, or a mixed set of genes and proteins.

**[0060]** A “weighted sum” of n summands  $S_1, \dots S_n$  may be any sum of the type

$$S = \sum_{i=1}^n c_i \cdot S_i ,$$

wherein  $c_i$  are defined real numbers. The weights  $c_i$  may be defined with further constraints, that they must be positive and/or that they must lay in a certain range of values.

**[0061]** Figure 1 shows a system 1 for selecting a treatment for a patient or for selecting patients for a clinical trial of a treatment according to one exemplary implementation. The system 1 comprises a processing unit 2 equipped with interfaces 3, 4, 5, 6 for retrieving data from input sources 7. The retrieved data

comprises identifications of disease indications of patients, sets of genes or proteins, molecular profiles of patients and targeted treatments, together with associated target proteins and/or the corresponding genes.

**[0062]** Processing unit 2 may operate in at least two operation modes. In a first mode, the input is an identification of a disease indication of a patient and several candidate targeted treatments. The processing unit 2 then generates a personalized treatment guideline for the patient. In a second mode, the input is identifications of a disease indication of several patients and one targeted treatment. The processing unit 2 then generates an ordered list of patients, providing a hint on their suitability for a clinical trial. The processing unit 2 may comprise a switch (not depicted), in particular a software switch for switching between the two operation modes. The switch may control an input mask, limiting a user's input in one or the other direction, or may control the amount of information retrieved from one or more databases, for instance by detailing an SQL query.

**[0063]** Interface 3 is adapted to providing an identification of a patient disease indication to the processing unit 2. The identification of a patient disease indication may be retrieved via a user, e. g. may be typed on a keyboard 8 or may be deduced from a free text typed on a keyboard 8, or selected from a multiple-choice element in a GUI (graphical user interface, not depicted). The patient disease indication may also be retrieved directly from an electronic health record (EHR) or electronic medical record (EMR), possibly on a chip-card or in a database (not depicted). The identification of a patient disease indication may also be retrieved from a patient database comprising medical records of a candidate set of patients (not depicted).

**[0064]** Interface 4 is adapted to providing a first set genes or proteins to the processing unit 2. The corresponding input source 7 consists of a database 9 of genes or proteins, which are part of the human genome or proteome in a broad sense, i.e. possibly including the genomes and proteomes of human symbionts or parasites, in particular the human microbiome. The database 9 further comprises a number of attributes of the genes or proteins which may be provided to the processing unit 2 via interface 4. Such attributes may include whether the gene or protein is a drug target, whether the gene or protein is a biomarker, whether the gene or protein is disease-associated, whether the gene or protein is an oncogene, whether the gene or protein is a tumor suppressor, whether the gene or protein has an association with a disease-relevant pathway (for instance, in a cancer pathway or in a Vogelstein core cancer pathway), whether the gene or protein is gene ontology annotated for disease-relevant processes, whether the gene or protein is part of a indication associated gene fusion, whether the gene or protein has a tractable domain, whether the gene or protein is embryonic lethal, whether the gene or protein is part of a Online Mendelian Inheritance in Man (OMIM), whether the gene or protein is part of a disease ontology, whether the gene or protein shows substantial co-occurrence with the specific indication, any related indication, or any indication at all, in a body of literature, as can be assessed by the use of text data mining and whether the gene or protein is found to be frequently mutated in that particular indication, using resources such as the COSMIC database. The database 9 may further hold information on normal or reference states of the proteins or genes, in particular the sequences of a reference genome (for instance, "hg19" or "GRCh37" by the Genome Reference Consortium) or "standard" versions of proteins (for instance, those stored in

SwissProt) and/or commonly observed expression levels of genes or proteins (for instance, as stored in ArrayExpress or Gene Expression Omnibus).

**[0065]** Interface 5 is adapted to providing a molecular profile of the patient to the processing unit 2. Such molecular profiles may comprise measurements of DNA or RNA extracted from the patient, produced by molecular measurement technology, for instance NGS or microarrays. Molecular profiles may be stored in a database 10 or in a computer file (not shown), from which they are retrieved via interface 5.

**[0066]** Interface 6 is adapted to providing a set of targeted treatments, together with their associated target proteins and/or the corresponding genes, to the processing unit 2. A treatment may be provided by a user, e. g. may be typed on a keyboard or may be deduced from a free text typed on a keyboard, or selected from a multiple-choice element in a GUI (graphical user interface, not depicted). Interface 6 may then provide associated target proteins and/or the corresponding genes for the treatment from a treatment database 11. The treatment database 11 is a database that holds available treatments, in particular drugs and drug combinations, together with information on the proteins or genes that are the targets of the drugs. Alternatively, a set of targeted treatments, together with their associated target proteins and/or the corresponding genes, may directly be retrieved from a treatment database 11, without user's input.

Specifically, treatments may be selected based on the disease indication of a patient, for instance by selecting treatments that are approved or under investigation for the indication of the patient or for related indications.

**[0067]** Processing unit 2 comprises a module 12 adapted to assigning an indication-specific score to genes or proteins, also referred to as module 12 for assigning the indication-specific score. The module 12 for assigning the indication-specific score retrieves data from interface 3 and interface 4. From database 9, a number of attributes of the genes or proteins are retrieved, the attributes indicating degrees of potential relationship to the indication. The indication-specific score of any gene or protein is computed by first assigning numerical values to the attributes and then aggregating the numerical values over the attributes, in particular by forming a weighted sum.

10 **[0068]** Processing unit 2 comprises another module 13 adapted to assigning a profile-specific score to genes or proteins, also referred to as module 13 for assigning the profile-specific score. The module 13 for assigning the profile-specific score retrieves data from interface 4 and interface 5. The module 13 for assigning the profile-specific score infers from the molecular profile for each gene or protein from a given molecular profile a possibly empty set of aberrations as compared to a reference. This reference may be provided by a second molecular profile, which corresponds to a reference state; an example of this is the common practice in oncology research to sequence both a tumor sample and a control sample, for instance, taken from blood. Alternatively, the reference may consist of measurements from other individuals that have been gathered before and from which typical measurement values of corresponding molecular profiles can be derived; such reference data may be stored with other gene- or protein-associated data in database 9.

**[0069]** The profile-specific score of any gene or protein is computed by assigning an aberration-specific functional impact score to each inferred aberration

and aggregating the aberration-specific functional impact score over the aberrations associated with the gene or protein and/or associated with a set of second genes or proteins, the set including the first gene or protein and a set of genes or proteins that are upstream of the first gene or protein according to a molecular interaction network, if there are two or more aberrations in the gene or protein and assigning zero to the gene or protein if the gene or protein has no aberrations. Alternatively, the profile-specific score may be set to 1 if the gene or protein has one or more aberrations and set to zero if the gene or protein has no aberrations.

- 10 **[0070]** Processing unit 2 comprises another module 14 adapted to assigning a patient-specific score to genes or proteins, also referred to as module 14 for assigning a patient-specific score. The module 14 for assigning a patient-specific score retrieves data from the module 12 for assigning the indication-specific score and from the module 13 for assigning the profile-specific score.
- 15 Module 14 is adapted to aggregating the indication-specific score and the profile-specific score. Preferably, the patient-specific score of every gene or protein is computed by multiplying the indication-specific score and the profile-specific score.

- [0071]** Processing unit 2 comprises another module 15 adapted to assigning a score to patients and treatments, also referred to as module 15 for assigning the score to patients and treatments. The module 15 for assigning the score to patients and treatments retrieves data from interface 6 and from module 14 for assigning the patient-specific score. Module 15 for assigning the score to patients and treatments is adapted to aggregating, for each targeted treatment, the patient-specific scores at least over its target proteins and/or the corresponding
- 20



genes. The aggregation is performed by calculating a sum or a weighted sum over the patient-specific scores of the involved proteins and/or genes.

**[0072]** Module 15 for assigning the score to patients and treatments is further adapted to generating a personalized treatment guideline as an ordered list of the targeted treatments and to generating an ordered list of candidate patients for a clinical trial of a treatment. Module 15 for assigning the score to patients and treatments further makes personalized treatment guidelines or lists of candidate patients for a clinical trial of a treatment available to another interface 16 of the processing unit, which may be connected with a GUI or a printer.

10 **[0073]** Figure 2 shows a flow diagram depicting a method for generating a personalized treatment guideline for a patient according to one specific implementation.

**[0074]** In a first step 20, the processing unit 2 may be switched into the first operation mode as described with regard to figure 1. The first step 20 is optional, however, as the method may also start with retrieving an identification of a disease indication directly.

**[0075]** In step 21, an identification of a disease indication of the patient is provided to the processing unit 2. In some embodiments, the method involves a step 22, wherein sub-scores are generated from gene or protein attributes. Generating the sub-scores may further involve a step 23 of network propagation for gene or protein attributes which are considered with respect to related genes or proteins. In step 24, an indication-specific score is assigned to each gene or protein of a first set of genes or proteins as described with regard to figure 1. Step

24 produces a clinico-molecular disease model 25 representing several aspects of the involvement of genes and proteins into the disease, with may be outputted.

**[0076]** In step 26, a molecular profile of the patient is retrieved, as described with regard to figure 1.

5 **[0077]** In some embodiments, the method involves a step 27 of inferring from the molecular profile for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations, a step 28 of assigning a functional impact score to each aberration and another step 29 of aggregating the functional impact score over the aberrations, as described with regard to figure 1.

10 **[0078]** In step 30, a profile-specific score is assigned to each gene or protein of the first set of genes or proteins, as described with regard to figure 1.

**[0079]** In step 31, a patient-specific score is assigned to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein combining its indication-specific score and its profile-specific score. With  
15 the patient-specific scores assigned to each gene or protein of the first set of genes or proteins, a patient model 32 is established and may be outputted.

**[0080]** In step 33, a set of targeted treatments, together with associated target proteins and/or the corresponding genes, is retrieved from a treatment database, as described with regard to figure 1.

20 **[0081]** In step 34, a score is assigned to each targeted treatment from the set of targeted treatments by aggregating the patient-specific scores at least over its associated target proteins and/or the corresponding genes.

**[0082]** In step 35, an ordered list of the targeted treatments is generated, wherein the order of the targeted treatments is determined by their score. The ordered list is outputted as the personalized treatment guideline 36.

**[0083]** Figure 3 shows a flow diagram depicting a method for selecting  
5 patients for a clinical trial of a treatment according to one implementation.

**[0084]** In a first step 40, the processing unit 2 may be switched into the second operation mode as described with regard to figure 1. The first step 40 is optional, however, as the method may also start with retrieving identifications of disease indications directly.

10 **[0085]** In step 41, a set of identifications of disease indications of patients is provided to the processing unit 2. In some embodiments, the method involves a step 42, wherein sub-scores for each disease indication are generated from gene or protein attributes. Generating the sub-scores may further involve a step 43 of network propagation for gene or protein attributes which are considered with  
15 respect to related genes or proteins. In step 44, for every patient, an indication-specific score is assigned to each gene or protein of a first set of genes or proteins as described with regard to figure 1. Step 44 produces a clinico-molecular disease model 45 representing several aspects of the involvement of genes and proteins into the disease, which may be outputted.

20 **[0086]** In step 46, molecular profiles of the patients are retrieved, as described with regard to figure 1.

**[0087]** In some embodiments, the method involves, for every patient, a step 47 of inferring from the molecular profile for each gene or protein of the first set of

genes or proteins a possibly empty set of aberrations, a step 48 of assigning a functional impact score to each aberration and another step 49 of aggregating the functional impact score over the aberrations, as described with regard to figure 1.

**[0088]** In step 50, for every patient, a profile-specific score is assigned to  
5 each gene or protein of the first set of genes or proteins, as described with regard to figure 1.

**[0089]** In step 51, for every patient, a patient-specific score is assigned to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein combining its indication-specific score and its profile-  
10 specific score. With the patient-specific scores assigned to each gene or protein of the first set of genes or proteins, a patient model 52 is established and may be outputted.

**[0090]** In step 53, for a targeted treatment associated target proteins and/or the corresponding genes are retrieved from a treatment database, as described  
15 with regard to figure 1.

**[0091]** In step 54, for every patient, a score is assigned to the targeted treatment from the set of targeted treatments by aggregating the patient-specific scores at least over its associated target proteins and/or the corresponding genes as described with regard to figure 1.

20 **[0092]** In step 55, an ordered list of the patients is generated, wherein the order of the targeted treatments is determined by their score. From the ordered list, patients may be selected for a clinical trial of the treatment and outputted.

**[0093]** Figure 4 shows a part of a molecular interaction network, represented as a directed graph with labeled nodes and edges. The three types of nodes are circles to represent proteins, diamonds to represent targeted drugs (hence treatments), and a square with clipped corners to represent an aberration.

- 5 There are two types of edges, where an arrow represents activation, and a ball represents inhibition.

**[0094]** In this example, the indication-specific score shall be non-negative, which is the preferred embodiment. Specifically, let  $s_1$ ,  $s_2$ ,  $s_3$ , and  $s_4$  denote the indication-specific scores of the proteins P1, P2, P3, and P4. Further,

- 10 aggregations are performed by computing weighted sums with positive weights, which is the preferred embodiment.

**[0095]** Treatments T1 and T2 act by inhibiting their respective target proteins P1 and P2. The aberration A3 deactivates protein P3. P2 acts on P4 in such a way that increased or decreased activity of P2 cause changes in the

15 activity of P4 in the same direction. P1 and P3 act on P4 in such a way that P4 reacts with a change in opposite direction to a change in the activity of P1 or P3.

**[0096]** The signed indirect functional impact scores of the proteins, determined by the aberration A3, are negative for P3 (as its function is inhibited by the aberration), positive for P4 (as its activity behaves reciprocally to that of P3),

20 and zero for P1 and P2 (as they are not downstream of P3). Specifically, the inFIS may take the values (0,0,-1,+1) for the proteins (P1,P2,P3,P4).

**[0097]** The patient-specific score  $t_1$  for treatment T1 is computed by aggregating over its target proteins and the proteins downstream to its target

proteins, which are P1 and P4 in this example. Hence  $t1 = 0 \cdot p1 - (+1) \cdot p4$ , which is a negative value.

**[0098]** The patient-specific score  $t2$  for treatment T2 is computed by aggregating over its target proteins and the proteins downstream to its target

5 proteins, which are P2 and P4 in this example. Hence  $t2 = 0 \cdot p2 + (+1) \cdot p4$ , which is a positive value.

**[0099]** As a result, treatment T2 is preferred over treatment T1: it will be placed higher in the ranked list of the patient-specific treatment guideline than T1.

This corresponds to the fact that treatment with T1 has a similar effect on P4 as

10 does the aberration A3, whereas treatment T2 has an effect of opposite directionality. Hence treating the patient with T2 has the potential of canceling part of the supposedly disease-causing effect of A3. Note that the degree of causal contribution to the disease is modeled and taken into account by the indication-specific scores.

15

### EXAMPLES

**[00100]** Table 1

| Attribute              | Indication | Subscore | Gene A |
|------------------------|------------|----------|--------|
| Cancer Biomarker       | same       | 10       | yes    |
|                        | related    | 5        | yes    |
|                        | other      | 3        | Yes    |
| Cancer Drug Target     | same       | 10       | No     |
|                        | related    | 5        | Yes    |
|                        | other      | 3        | Yes    |
| Oncogene               | ---        | 8        | No     |
| Suppressor             | ---        | 5        | Yes    |
| In core cancer pathway | ---        | 2        | Yes    |

**[00101]** Table 1 shows details from step 22 described with regard to figure 2. Via table 1, an indication-specific score, in the specific case an Oncoscore, is calculated for an exemplary gene from the first set of genes or proteins. In table 1, the exemplary gene is indicated by "Gene A".

5 **[00102]** In table 1, gene or protein attributes are listed, here exemplarily "cancer biomarker", "cancer drug target", "oncogene", "suppressor" and "in core cancer pathway". Some of the attributes, here exemplarily "cancer biomarker" and "cancer drug target", involve further attributes. The further attributes exemplarily  
10 comprise information about the indication which the cancer biomarker or the cancer drug target is associated with. These further attributes are "same indication", "related indication", "other indication". The gene or protein attributes are associated with sub-scores, here exemplarily ranging from 2 to 10. The candidate gene "Gene A" is associated with the attributes per "yes" or "no".

**[00103]** Assigning the indication-specific score to the candidate gene may be  
15 performed by calculating a scalar product from the information about the candidate gene's attributes and the associated sub-scores. This results in if the candidate gene has the attribute, then its associated sub-score will be counted, otherwise not. In the example case, the oncoscore is  $10 + 5 + 3 + 5 + 3 + 5 + 2 = 33$ .

**[00104]** While the invention is particularly shown and described with  
20 reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein.

## CLAIMS

1. Method for generating a personalized treatment guideline for a patient, comprising the steps:

retrieving an identification of a disease indication of the patient as a user's input, from an electronic health record or electronic medical record or from a patient database comprising medical records of a candidate set of patients,

assigning an indication-specific score to each gene or protein of a first set of genes or proteins, the first set of genes or proteins being part of the human microbiome, the indication-specific score representing an evidence of the gene or protein being associated with the disease indication of the patient,

wherein the indication-specific score of each gene or protein is computed by first assigning numerical values to attributes indicating degrees of potential relationship to the indication and then by forming a weighted sum of the numerical values or of any monotonic transformation of the numerical values over the attributes, the attributes comprising at least one of the following list:

- a) The gene or protein is a drug target,
- b) The gene or protein is a biomarker or part of a biomarker,
- c) The gene or protein is disease-associated,
- d) The gene or protein is an oncogene or a product of an oncogene, respectively,
- e) The gene or protein is a tumor suppressor,
- f) The gene or protein has a cancer pathway association,
- g) The gene or protein is gene ontology annotated for cancer relevant processes,



- h) The gene or protein is part of a cancer-associated gene fusion,
- i) The gene or protein has a tractable domain,
- j) The gene or protein is embryonic lethal, and
- k) The gene or protein is highly mutated in the specific indication;

retrieving a molecular profile of the patient, wherein the molecular profile is a set of molecular measurements of DNA or RNA extracted from the patient, indexed by a second set of genes or proteins, the second set of genes or proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins,

inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations as the differences of the molecular profile to a reference profile from a control sample of the patient or from other individuals,

assigning a *profile-specific score* to each gene or protein of the first set of genes or proteins, the profile-specific score representing a degree of how much the function of the genes or proteins is altered by the aberrations,

wherein assigning the profile-specific score to each gene or protein of the first set of genes or proteins comprises the following steps:

computing an aberration-specific functional impact score for each aberration taking into account the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the aberration-specific functional impact score correspond to an activating and inactivating impact, respectively, and

computing a protein- or gene-specific direct functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with the gene or protein, or

computing a protein- or gene-specific indirect functional impact score for each protein or gene by computing a weighted sum, a quantile, a

median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with another set of genes or proteins, the another set including the gene or protein and a set of genes or proteins that are upstream of the gene or protein according to a molecular interaction network, the molecular interaction network being a directed graph, consisting of nodes each of which represents a gene or a protein, and of edges each of which connects two nodes,

and wherein the profile-specific score is set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these,

assigning a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being a sum, a weighted sum or a product of its indication-specific score and its profile-specific score,

retrieving a set of targeted treatments, together with associated target genes or proteins, as a user's input or by selecting treatments that are approved or under investigation for the indication of the patient or for related indications from a treatment database, the treatment database storing treatments, in particular drugs and drug combinations, together with information on the genes or proteins that are the targets of the drugs,

assigning a score to each targeted treatment from the set of targeted treatments by calculating a sum or a weighted sum over the patient-specific scores of the associated target genes or proteins, and

generating the personalized treatment guideline as an ordered list of the targeted treatments, wherein the order of the targeted treatments is determined by their score.

2. Method for selecting a treatment for a patient, comprising the steps:

generating a personalized treatment guideline for the patient by executing the method steps according to claim 1, and

selecting the first targeted treatment from the ordered list of the targeted treatments of the personalized treatment guideline.

3. Method for selecting patients for a clinical trial of a treatment, comprising the steps:

retrieving identifications of disease indications of a set of patients as a user's input, from an electronic health record EHR or electronic medical record EMR or from a patient database comprising medical records of a candidate set of patients,

assigning, for every patient, an indication-specific score to each gene or protein of a first set of genes or proteins, the first set of genes or proteins being part of the human microbiome, the indication-specific score representing an evidence of the gene or protein being associated with the disease indication of the patient,

wherein the indication-specific score of each gene or protein is computed by first assigning numerical values to attributes indicating degrees of potential relationship to the indication and then by forming a weighted sum of the numerical values or of any monotonic transformation of the numerical values over the attributes, the attributes comprising at least one of the following list:

- a) The gene or protein is a drug target,
- b) The gene or protein is a biomarker or part of a biomarker,
- c) The gene or protein is disease-associated,
- d) The gene or protein is an oncogene or a product of an oncogene, respectively,
- e) The gene or protein is a tumor suppressor,
- f) The gene or protein has a cancer pathway association,
- g) The gene or protein is gene ontology annotated for cancer relevant processes,
- h) The gene or protein is part of a cancer-associated gene fusion,
- i) The gene or protein has a tractable domain,

- j) The gene or protein is embryonic lethal, and
- k) The gene or protein is highly mutated in the specific indication;

retrieving, for every patient, a molecular profile of the patient, wherein the molecular profile is a set of molecular measurements of DNA or RNA extracted from the patient, indexed by a second set of genes or proteins, the second set of genes or proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins,

inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations as the differences of the molecular profile to a reference profile from a control sample of the patient or from other individuals,

assigning, for every patient, a profile-specific score to each gene or protein of the first set of genes or proteins, the profile-specific score representing a degree of how much the function of the genes or proteins is altered by the aberrations,

wherein assigning the profile-specific score to each gene or protein of the first set of genes or proteins comprises the following steps:

- computing an aberration-specific functional impact score for each aberration taking into account the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the aberration-specific functional impact score correspond to an activating and inactivating impact, respectively, and

- computing a protein- or gene-specific direct functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with the gene or protein, or

- computing a protein- or gene-specific indirect functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with another

set of genes or proteins, the another set including the gene or protein and a set of genes or proteins that are upstream of the gene or protein according to a molecular interaction network, the molecular interaction network being a directed graph, consisting of nodes each of which represents a gene or a protein, and of edges each of which connects two nodes,

and wherein the profile-specific score is set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these,

assigning, for every patient, a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being a sum, a weighted sum or a product of its indication-specific score and its profile-specific score,

retrieving, for the treatment, associated target genes or proteins from a treatment database, the treatment database storing treatments, in particular drugs and drug combinations, together with information on the genes or proteins that are the targets of the drugs, which are also in the first set of genes or proteins,

assigning a score to every patient from the set of patients by calculating a sum or a weighted sum over the patient-specific scores of the associated target genes or proteins, and

generating an ordered list of the patients, wherein the order of the patients is determined by their score.

4. The method of any of the preceding claims, wherein the profile-specific score is set to zero if the gene or protein has no aberrations.
5. The method of any of the preceding claims, wherein assigning the score to a treatment or patient further takes into account a directionality of the impact of the treatment on its associated target genes or proteins by adding the profile-specific scores of target genes or proteins that are inhibited or antagonized by the treatment and by subtracting the profile-specific scores of target genes or proteins that are activated or stimulated or agonized by the treatment.

6. The method of any of the preceding claims, wherein assigning the score to a treatment or patient involves the step that patient-specific scores are further aggregated over genes or proteins downstream of their associated target genes or proteins according to the molecular interaction network by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with another set of genes or proteins, the another set including the gene or protein and a set of genes or proteins that are downstream of the gene or protein according to the molecular interaction network.
7. System for generating a personalized treatment guideline for a patient and/or for selecting patients for a clinical trial of a treatment, comprising a processing unit with:

an interface for providing a first set of genes or proteins to the system, the first set of genes or proteins being part of the human microbiome,

an interface for providing identifications of disease indications of patients to the system,

a module for assigning, for an identification of a disease indication of a patient, an indication-specific score to each gene or protein of the first set of genes or proteins, the indication-specific score representing an evidence of the gene or protein being associated with the disease indication of the patient,

wherein the indication-specific score of each gene or protein is computed by first assigning numerical values to attributes indicating degrees of potential relationship to the indication and then by forming a weighted sum of the numerical values or of any monotonic transformation of the numerical values over the attributes, the attributes comprising at least one of the following list:

  - a) The gene or protein is a drug target,
  - b) The gene or protein is a biomarker or part of a biomarker,
  - c) The gene or protein is disease-associated,
  - d) The gene or protein is an oncogene or a product of an oncogene, respectively,
  - e) The gene or protein is a tumor suppressor,
  - f) The gene or protein has a cancer pathway association,
  - g) The gene or protein is gene ontology annotated for cancer relevant processes,

- h) The gene or protein is part of a cancer-associated gene fusion,
- i) The gene or protein has a tractable domain,
- j) The gene or protein is embryonic lethal, and
- k) The gene or protein is highly mutated in the specific indication;

an interface for providing molecular profiles of patients to the system, wherein a molecular profile is a set of molecular measurements of DNA or RNA extracted from a patient, indexed by a second set of genes or proteins, the second set of genes or proteins being equal to the first set of genes or proteins or being a subset of the first set of genes or proteins,

a module for assigning, for a molecular profile of a patient, a profile-specific score to each gene or protein of the first set of genes or proteins, wherein the module for assigning the profile-specific score is adapted to inferring, from the molecular profile, for each gene or protein of the first set of genes or proteins a possibly empty set of aberrations as the differences of the molecular profile to a reference profile from a control sample of the patient or from other individuals,

wherein assigning the profile-specific score to each gene or protein of the first set of genes or proteins comprises the following steps:

computing an aberration-specific functional impact score for each aberration taking into account the direction of the impact of an aberration on the function of a gene or protein, wherein positive and negative signs of the aberration-specific functional impact score correspond to an activating and inactivating impact, respectively, and

computing a protein- or gene-specific direct functional impact score for each protein or gene by computing a weighted sum, a quantile, a median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with the gene or protein, or

computing a protein- or gene-specific indirect functional impact score for each protein or gene by computing a weighted sum, a quantile, a

median, a maximum or a geometric mean of the aberration-specific functional impact scores over the aberrations associated with another set of genes or proteins, the another set including the gene or protein and a set of genes or proteins that are upstream of the gene or protein according to a molecular interaction network, the molecular interaction network being a directed graph, consisting of nodes each of which represents a gene or a protein, and of edges each of which connects two nodes,

and wherein the profile-specific score is set equal to the direct functional impact score, equal to the indirect functional impact score or equal to a weighted sum of these,

a module for assigning a patient-specific score to each gene or protein of the first set of genes or proteins, the patient-specific score of each gene or protein being a sum, a weighted sum or a product of its indication-specific score and its profile-specific score,

an interface for providing targeted treatments, together with associated target genes or proteins, as a user's input or by selecting treatments that are approved or under investigation for the indication of the patient or for related indications from a treatment database, the treatment database storing treatments, in particular drugs and drug combinations, together with information on the genes or proteins that are the targets of the drugs, to the system, and

a module for assigning a score to patients and to treatments by calculating a sum or a weighted sum over the patient-specific scores of the associated target genes or proteins, wherein the module for assigning the score to patients and to treatments is adapted to generating ordered lists of patients or treatments, wherein the order of the patients or treatments is determined by their score.