

SYSTEMS AND METHODS FOR CLINICAL DECISION SUPPORT

FIELD OF THE DISCLOSURE

5 [0001] The present disclosure relates to methods and systems for prioritizing possible treatment options based on mined biomedical data.

BACKGROUND OF THE DISCLOSURE

10 [0002] A large number of publications exist regarding human disease etiology and progression, many of which offer information of the relation between diseases, biomarkers, treatments, and outcomes. Many publications further discuss various molecular entities such as proteins, small molecules such as metabolites, nutrients, drugs, transporters, enzymes, complexes, and/or pathways. 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 easily and quickly perform analyses, and is difficult for
15 clinicians to identify and/or judge personalized patient treatment options.

BRIEF SUMMARY OF THE DISCLOSURE

20 [0003] Systems and methods are described herein for the prioritization of possible treatment options based on biomarkers, such as, but not limited to, tumor and germline-based genomic variants. The system and methods may thereby identify treatment options tailored to a patient's genetic data, in particular to his/her clinical, molecular, and/or genetic condition. Furthermore, the system and method provides a means for prioritizing the possible treatment options based on the extraction and contextualization of clinical and molecular knowledge. The system gathers and/or accesses biomarker information and transforms the information into
25 prioritized, clinically actionable options identified for a specific patient case.

[0004] The present disclosure provides a method for prioritization of patient treatment options based on an analysis of biomarkers, comprising:

retrieving the results of measurements of a set of one or more biomarkers of the patient,

5 identifying a plurality of treatments associated with any of the set of one or more biomarkers via identifying rules relating the treatments to biomarkers in a treatment information database, the rules corresponding to one or several observations of an outcome after a treatment, applied either to a human or to a model system like a cell line,

generating a score for each of the identified plurality of treatments, and
10 ordering the identified plurality of treatments according to the generated score to provide a treatment option or treatment contraindication prioritization for the patient,

characterized in that for each of the one or more biomarkers the score is computed by setting values for

- 15 a) whether rules are associated with
- response to the treatment,
 - resistance to the treatment, and
 - risk of adverse effects from the treatment.

[0004A] The present disclosure further provides a curated database comprising
20 information prepared for being used for generating a score for a treatment to be prioritized when performing a method for providing a prioritization of patient treatment options as described above, in particular comprising one or more of the following:

- data extracted from other databases,
- 25 - information extracted from text documents via text mining,
- measurement data or processed measurement data from appropriate assays,
- manually entered data,
- sign-offs and/or changes by curators.

[0004B] The present disclosure still further provides clinical decision support device comprising a curated database as described above and a data mining module.

[0004C] Further aspects and embodiments of the present disclosure are provided
5 below.

[0004D] In a first aspect, the present disclosure is directed to a method for prioritization of patient treatment options based on an analysis of biomarker information. The method includes retrieving the results of measurements of a set

of one or more biomarkers of the patient. The method further includes identifying a plurality of treatments associated with any of the set of one or more biomarkers in a database. The method also includes generating a score for each of the identified plurality of treatments, the score being based on

- 5 a) whether predictive rules for each of the one or more biomarkers are associated with response to the treatment, resistance to the treatment, or risk of adverse effects from the treatment.

The method further includes ordering at least a portion of the identified plurality of treatments according to the generated score to provide a treatment option or
10 treatment contraindication prioritization for the patient.

[0005] In order to identify promising treatments for a patient, it is useful to check whether any of the biomarkers, e.g. mutations found in the patient, e.g. in the patient's tumor or in the patient's normal control tissue, are indicative of the patient's outcome under any treatment. The inventors observed repeatedly that in
15 practice some biomarker (for instance, a specific mutation) has been observed to be linked with both response to and resistance against some treatment, often in the same indication. Such apparently contradictory information can lead to classifying the treatment as both likely efficacious and likely inefficacious. The invention offers a principled resolution to this dilemma.

20 **[0006]** "Identifying a plurality of treatments" requires the identification of at least two treatments associated with any of the set of one or more biomarkers. As a matter of fact, the treatments are identified via identifying predictive rules relating the treatments to biomarkers. By sorting the identified plurality of treatments according to their scores, the treatment prioritization can be easily communicated
25 and made accessible to the users of the method, in particular to the patient himself or to a treating physician.

[0007] In preferred embodiments of the method, at least two discordant predictive rules for a treatment are identified. These predictive rules may even refer to the same biomarker. Discordant predictive rules may be outright contradictory, such
30 that one is associated with response to the treatment and the other one resistance to the treatment, or one with response to the treatment and the other one with risk

of adverse effects from the treatment. The discordance may also be of a less fundamental kind, for instance such that one predictive rule predicts a moderate increase of the response rate as compared to a placebo, whereas another rule predicts a larger increase. Such discordances may pose a problem to prioritizing treatments, since separate consideration of each of such two predictive rules may lead to a different ordering of treatments.

[0008] In some embodiments of the method, a value for a predictive rule for the biomarker being associated with response to the medication is set equal to 1; being associated with resistance to the medication is set equal to -1; and being associated with a risk of adverse effects from the medication is set equal to a defined negative value between 0 and -1, in particular equal to -0.2, -0.4, -0.6, or -0.8.

[0009] In other embodiments, the value for a predictive rule for the biomarker being associated with response is set to 1 and the value for the biomarker being associated with resistance is set to -1, and the value for risk is set to between 0 and -1 for minor side effects and to less than -1, e.g. to -3 or -5 for severe side effects.

[0010] In preferred embodiments of the method, generating the score for each of the identified plurality of treatments further comprises, for each generated score, generating a sub-score for each biomarker and aggregating the sub-scores to generate the score. In some of these embodiments, aggregating the sub-scores comprises adding each sub-score for each biomarker associated with the treatment, or multiplying each sub-score for each biomarker associated with the treatment, or adding functions, such as logarithms, of each sub-score, or multiplying functions, such as exponentials, of each sub-score. Thus, by numerically quantifying the evidence in either direction by one or more sub-scores and balancing it in the process of aggregation, a principled solution to the problem with apparently contradictory information is given. The sub-scores can be aggregated over different pieces of evidence, possibly even linked to different mutations or other biomarkers, such that the final score reflects the balance within a continuum from complete response to total resistance, or an even higher-dimensional space of possible outcomes.

[0011] In preferred embodiments of the method, generating the score for each of the identified plurality of treatments further comprises generating two or more separate scores for each of the identified plurality of treatments, said separate scores corresponding to categories that relate to responsiveness, resistance, and/or risk, or, in other words, two or more separate scores may be computed for two or more categories of response, resistance, and risk. Separate scores are preferably computed by aggregating (separate) sub-scores. There may be one separate score for response and a separate score for resistance and risk. There may also be one separate score for response, one separate score for resistance, and a third separate score for risk in general. The term risk may also be refined by considering one or more specific types of risk. These different types of risks may be classified according to a medical ontology for side effects, for instance MedDRA. Risk-related separate scores may also be used to model groups of adverse events of different severity; for instance there may be a separate score for life-threatening side effects and/or a separate score for adverse events that are merely annoying. Thus, there may be one separate score for risk representing risk in general, or there may be one or more separate scores that correspond to different types of risk, in particular even thousands of separate scores that relate to different diseases caused as different side effects.

[0012] Generating the score then further comprises aggregating the two or more separate scores. Thus, by numerically quantifying the evidence in either direction by one or more separate scores and balancing it in the process of aggregation, a principled solution to the problem of taking into account different categories of possible outcomes is presented.

[0013] Scoring and ordering at least a portion of the identified plurality of treatments may comprise weighting the two or more separate scores according to treatment risk/benefit preferences, and aggregating the weighted two or more separate scores to determine the order of treatments. In some embodiments, the method thus allows assessment of the impact of different choices of trade-offs between treatment risks and treatment benefits on the prioritization of treatments. Aggregating the weighted two or more separate scores may in particular comprise adding each weighted separate score, or multiplying each weighted separate

score, or adding functions, such as logarithms, of each weighted separate score, or multiplying functions, such as exponentials, of each weighted separate score.

[0014] In some embodiments of the method, the score, separate score, sub-score, or separate sub-score is computed by applying a monotonic functional to the value; specifically, this functional may be the identity function.

[0015] In further embodiments, the method includes specifying, by a user, treatment risk/benefit preferences, in particular via input devices such as a touch-screen, a keyboard, or a mouse. These preferences may be represented by weights of (different types of) adverse events, ideally in relation to efficacy. Consequently, the method may thus provide treatment prioritization taking into account specific types of risk, e.g. adverse events, which the patient would not accept during treatment, for example due to specific patient co-morbidities. Ordering at least a portion of the identified plurality of treatments then comprises weighting the two or more separate scores according to the treatment risk/benefit preferences. The user of the method may thus allow for a down-ranking of such treatments which the patient would not accept during treatment.

[0016] In some embodiments of the method, the score is further based on

b) effect sizes of the predictive rules for each of the one or more biomarkers.

[0017] Effect sizes may simply refer to numbers of cases in which the respective outcome (e.g. response, or some adverse event) has been observed. In further embodiments, the effect sizes of the predictive rules for each of the one or more biomarkers comprise a measure of likelihood of response or resistance, a measure of a hazard ratio, a measure of likelihood odds of response or resistance, or a measure of a quotient of hazard ratios. Similarly, a measure of a log likelihood of response within a retrospectively observed cohort may be used as quantification of the therapeutic effect of a treatment. As an example, a useful effect size could also be a logarithm of a hazard ratio relating overall survival under treatment to overall survival with placebo in two arms of a clinical trial as estimated by a proportional hazards model.

[0018] In preferred embodiments of the method, the score, the sub-scores, the separate scores, or the separate sub-scores are computed from the two or more values for the attributes a) and b) of the predictive rules by multiplying monotonic functions of those values. In preferred embodiments, separate sub-scores are
5 computed by multiplying the values. Values for further attributes of predictive rules, as described below, in particular the values labeled c) - i), may be treated analogously.

[0019] Some embodiments of the method further comprise retrieving an identification of an indication of a patient. Preferably, the measurements of the set
10 of one or more biomarkers of the patient are well chosen with regard to the indication of the patient. In a preferred embodiment, the score is further based on c) whether the predictive rules for each of the one or more biomarkers have been observed in the indication of the patient, in related indications, or in unrelated indications.

[0020] In further embodiments, the score comprises a value for whether the predictive rules for each of the one or more biomarkers have been observed in the indication of the patient, in related indications, or in unrelated indications. Specifically, a value for a predictive rule for the biomarker being validated in the indication of the patient may be set equal to 1; for the biomarker being validated in
20 a related indication equal to a defined positive number between 0 and 1, in particular equal to 0.2, 0.4, 0.6, or 0.8; and for the predictive rule for the biomarker being validated in an unrelated indication equal to a defined non-negative number less than the value of a related indication, in particular 0, 0.01, 0.02, 0.05, or 0.1.

[0021] In preferred embodiments, the value for the predictive rule for the biomarker
25 being detected in a related indication may be derived from a medical classification system. The indication of the patient and the indication for which the predictive rule for the biomarker exists may be identified according to disease ontologies, for instance ICD-10, MeSH, or MedDRA. For certain classes of indications there may also be specialized ontologies that may offer advantages like more precise
30 categorization of the indication. For example, in oncology it may be beneficial to use ICD-O-3 and/or the TNM staging system. Many ontologies (e.g. MeSH) are

organized as directed acyclic graphs (DAGs), which allow to derive different types of relatedness between indications. Specifically, for any two indications A and B, that are nodes in such a DAG, it may be useful to distinguish between the cases A equals B; A specializes B (directly), for A being a descendent (son) of B; A
5 generalizes B, for A being an ancestor (parent) of B; A is closely related (level-1-related) to B, for A and B sharing a parent; A is level-2-related to B, for A and B sharing a grandparent; possibly relationships of uncle/nephew-type; possibly further, more distant relationships; and A being dissimilar ("unrelated") to B, otherwise. Each type of relatedness may be associated with a different value,
10 typically with non-negative values that increase as the relationship gets closer.

[0022] In further embodiments of the method, the score is further based on

d) clinical validation levels of the predictive rules for each of the one or more biomarkers.

[0023] In a further embodiment, the score comprises a value for the clinical
15 validation levels of the predictive rules for each of the one or more biomarkers. Preferably, the clinical validation level is one out of the following list: endorsed, clinical, pre-clinical, inferred. A biomarker or a predictive rule may be classified as endorsed if it is approved by an appropriate authority, for instance by the European Medicines Agency (EMA), by the U.S. Food and Drug Administration
20 (FDA), or by an expert committee. A biomarker or a predictive rule may be classified as inferred if it is derived by modifying another predictive rule, in particular by replacing a protein-related biomarker while keeping its other characteristics, for instance by considering a homologous protein. Specifically, a value for the endorsed validation level may be set equal to 1; a value for the
25 clinical validation level equal to a defined number less than 1, preferably between 0.5 and 1, most preferably equal to 0.8; a value for the pre-clinical validation level equal to a defined number less than or equal to 0.2; and a value for the inferred validation level equal to a defined number less than or equal to 0.5. In further
30 embodiments, the validation level and the corresponding value may take into account whether a study, trial, or investigation was clinical or non-clinical; observational or interventional; randomized or non-randomized; prospective or retrospective; n-of-one or not; and/or possibly further attributes.

[0024] In further embodiments of the method, the information about the predictive rules for the set of one or more biomarkers with respect to the identified plurality of treatments is stored in one or more mined databases. Specifically, the database may contain data generated by text mining. In other words, identifying the

5 predictive rules for any of the set of one or more biomarkers with respect to each of the identified plurality of treatments may implicitly comprise a step of searching for document sources comprising identifications of the biomarker having a measure of co-occurrence with identifications of a treatment above a defined threshold.

10 **[0025]** In further embodiments of the method, the information about the predictive rules for the set of one or more biomarkers with respect to the identified plurality of treatments is stored in one or more curated databases. In other words, identifying the predictive rules (identifying the plurality of treatments) for any of the set of one or more biomarkers implicitly comprises the step of curating document sources

15 comprising identifications of the biomarker and identifications of the treatment. Curating document sources may comprise extracting information automatically and/or manually by human resources, such as physicians, or other qualified personnel. In particular, curating may also refer to a two-step process comprising a first step of data mining performed by an analysis engine executed by a

20 computing device in order to pre-select relevant document sources and a second step of manually extracting the relevant information about the predictive rules for the biomarker in question with respect to the treatment in question from the pre-selected relevant document sources.

[0026] In further embodiments of the method, the score is further based on

25 e) a reliability of the involved sources of the predictive rules.

[0027] In further embodiments, the score comprises a value for the reliability of the involved sources. Specifically, the value for the reliability of the sources of the predictive rules may be a defined value from 0 to 1, wherein a higher value corresponds to a greater reliability of the involved source. Typically, a value of 1

30 would represent full trustworthiness, and a value of 0 would represent total lack of reliability. Frequently, the source of a predictive rule is a document, for instance a

scientific publication. In those cases, the value for the reliability of the source of the predictive rule may comprise a value attributed to the source of the publication, e.g. to medical journals, academic societies or universities, international organizations, public organizations, such as ministries or public research agencies. The value for the reliability of the document source may also comprise a value attributed to the authors of the document, which may also involve evaluating a reputation or counting overall citations of the authors. The value for the reliability of the document source may also comprise an impact factor of the document source, such as an impact factor or a number of citations of the document source.

10 The reliability of the source of a predictive rule may also comprise a value assigned by a curator, for instance based on a judgment of the experimental techniques used to obtain the result, or based on a judgment of the thoroughness of the investigation. Said values may be summed, multiplied or averaged in order to obtain the value for the reliability of the involved document source in question.

15 **[0028]** In further embodiments of the method, the score is further based on

f) a reliability of measurement of the biomarkers.

[0029] In further embodiments, the score comprises a value for the reliability of measurement of the biomarkers. Specifically, the value for the reliability of measurement of the biomarkers may be a defined value from 0 to 1, wherein a

20 higher value corresponds to a greater reliability of measurement for a given biomarker. Typically, a value of 1 would represent full trustworthiness, and a value of 0 would represent total lack of reliability. For instance, for genomic variants as biomarkers the value could be a posterior probability estimate of the variant being present in the sample. As an example, the somatic SNV calling software

25 JointSNVMix estimates such posterior probabilities. In some embodiments of the method, the value for the reliability of measurement of the biomarkers is composed of a value for a reliability of the detection of the biomarkers and a value representing the abundance of biomarkers in the patient, e.g. a frequency of detection of the biomarkers in the patient. For example, a posterior probability of

30 the variant being present could be multiplied with the estimated relative frequency of the variant within the sample. In a further embodiment, the value for the reliability of the detection method of the biomarker and the value representing the

abundance of a biomarker in the patient are summed, multiplied or averaged in order to build the value for the reliability of detection of the biomarker.

[0030] In some embodiments of the method, the score is further based on

g) whether the treatments recommended in a standard treatment guideline.

- 5 **[0031]** In some embodiments, it is verified whether any treatment is recommended or applicable according to a standard treatment guideline. The score then comprises a value for whether the treatments are recommended or applicable in a standard treatment guideline. In case any treatment is recommended, the treatments recommended most may receive a value of 1, further recommended
- 10 treatments may receive values above 0.1 up to 1, and any other treatments may receive a fixed value below 0.1. By setting the values such that the ratio of the lowest value for a recommended treatment over the highest value of a non-recommended treatment is high, for instance at least 10, it can be made highly likely that treatments identified in standard treatment guidelines appear in top ranked
- 15 positions. In case no recommendation applies to the patient case at hand, the value for all treatments is 1, such that the scale of the multiplicative score is not reduced due to the lack of a standard therapy; in other words, the scores of the treatments are sensitive to the degree of clinical need.

[0032] In some embodiments of the method, the score is further based on

- 20 h) an availability of the treatments.

[0033] In further embodiments, the score comprises a value for the availability of the treatments. Specifically, the value for the availability of the treatment comprises a defined value from 0 to 1, wherein a higher value corresponds to a greater availability of the associated treatment.

- 25 **[0034]** In some embodiments, the method includes retrieving an identification of a past treatment history of the patient. In some embodiments of the method, the score is further based on

i) the past treatment history of the patient.

[0035] In further embodiments, the score comprises a value for the past treatment history of the patient. The value for the past treatment history may be equal to 1 if the patient has never been subjected to the treatment in question and may be equal to a defined number less than 1, e. g. 0.1, or 0.01, if the patient has been
5 subjected to the treatment. Thus a particular treatment may be ranked down if it has already been tested at the patient.

[0036] In preferred embodiments, the value for the past treatment history of the patient depends on the attribute a) of the predictive rule and on the observed outcomes of the treatment in the past of the patient. The dependence may, in
10 particular, be in the following way: for treatments not previously applied to the patient, the value is 1; for treatments previously applied to the patient, for a response biomarker, the value is set as above; for a resistance biomarker or a risk biomarker, the value is set to a value greater than 1 if resistance or a corresponding adverse event has occurred in the patient, and to a non-negative
15 value below 1 otherwise.

[0037] In alternative embodiments, patient-specific predictive rules are introduced for past treatments of the patient that reflect the observations from the patient history. Technically speaking, the biomarker is the identity of the patient, and the predicted outcome of a re-enacted application of a past treatment would be the
20 outcome observed at the previous application.

[0038] As already stated above, further embodiments of the method comprise retrieving an identification of an indication of a patient. In some of these embodiments, the method includes identifying one or more treatments associated with the indication of the patient. The method also includes generating a score for
25 the identified treatments associated with the indication. In further embodiments, the method includes identifying one or more treatments associated with the indication of the patient and not associated with any of the set of one or more biomarkers of the patient. The method then includes generating a score for the identified treatments associated with the indication and not associated with any of
30 the set of one or more biomarkers of the patient. The treatment option or treatment contraindication prioritization may thus comprise both treatments associated with any of the set of one or more biomarkers and treatments associated with the

indication and not associated with any of the set of biomarkers of the patient. This allows for providing a prioritization of treatments that includes treatments without biomarker information, and this may also lead to the result that such a treatment may achieve the highest priority. Generating a score for the identified treatments
5 associated with the indication preferably comprises providing background knowledge on response, resistance, and/or risk. Such background knowledge may be represented in form of predictive rules that do not depend on a biomarker and may be stored in a database. In particular, such background knowledge may comprise average effect sizes with respect to representative mixtures of biomarker
10 values in patients. Hence typical background knowledge may include the efficacy and adverse reaction results of clinical studies of treatments in their respective indication, as are often cited in drug label information.

[0039] The method may further comprise outputting the treatment option prioritization to a user as a list comprising one or more of the highest ordered
15 treatments, in particular by displaying them on a screen or display or by printing them onto paper. The method may further comprise outputting the treatment contraindication prioritization to a user as a list comprising one or more of the lowest ordered treatments.

[0040] The following paragraphs comprise further embodiments of the invention
20 being part of a priority application to this application. Some of the features labeled A) – G) below may be identical to features labeled a) – i) above. However, there is no correspondence between the literals.

[0041] In another embodiment, the present disclosure is directed to a method for prioritization of patient treatment options based on multivariate analysis of
25 biomarker information. The method includes retrieving, by an analysis engine executed by a computing device, an identification of an indication of a patient and the results of measurements of a set of one or more biomarkers in the patient. The method further includes identifying, by the analysis engine, a plurality of treatments associated with the indication, the patient, or any of the set of one or
30 more biomarkers in a treatment information database. The method also includes generating, by the analysis engine, a score for each of the identified plurality of treatments, the score being based on A) clinical validation levels of predictive rules

for each of the one or more biomarkers; B) whether predictive rules for each of the one or more biomarkers are associated with response to the treatment, resistance to the treatment, or risk of adverse effects from the treatment; and C) a reliability of measurement of each of the one or more biomarkers. The method also includes
5 ordering at least a portion of the identified plurality of treatments according to the generated score to provide a treatment option or treatment contraindication prioritization for the patient.

[0042] In some embodiments of the method, generating a score for each of the identified plurality of treatments further comprises, for each generated score:

10 generating a plurality of sub-scores for a corresponding plurality of the set of biomarkers; and aggregating the plurality of sub-scores to generate the score. In further embodiments, aggregating the plurality of sub-scores comprises adding each sub-score for each biomarker associated with the treatment.

[0043] In some embodiments, the method includes two or more separate scores

15 are computed for each treatment, said separate scores corresponding to responsiveness, resistance, or risk. Ordering at least a portion of the identified plurality of treatments includes weighting the two or more separate scores according to treatment risk/benefit profile, and aggregating the weighted two or more separate scores to determine the order of treatments. In further
20 embodiments, the treatments are prioritized according to a weighted sum of their separate scores. In a still further embodiment, the method includes specifying, by a user, weights via a treatment risk/benefit profile.

[0044] In some embodiments of the method, the clinical validation level is one out of the following list: endorsed, clinical, pre-clinical, inferred. In a further

25 embodiment, a value for the endorsed validation level is set equal to 1; a value for the clinical validation level is set equal to a defined number less than 1, preferably between 0.5 and 1, most preferably equal to 0.8; a value for the pre-clinical validation level is set equal to a defined number less than or equal to 0.2; and a value for the inferred validation level is set equal to a defined number less than or
30 equal to 0.5.

[0045] In some embodiments of the method, the value for a predictive rule for the biomarker being associated with response to the medication is set equal to 1; being associated with resistance to the medication is set equal to -1; and being associated with a risk of adverse effects from the medication is set equal to a defined negative value between 0 and -1, in particular equal to -0.2, -0.4, -0.6, or -0.8.

[0046] In some embodiments, the value for whether a predictive rule for the biomarker is associated with response to the treatment, resistance to the treatment, or risk of adverse effects from the treatment is further based on a measure of an effect size of the predictive rule for the biomarker.

[0047] In many embodiments of the method, the value for the reliability of detection of the biomarker comprises a value for a reliability of the detection method of the biomarker and a value for a frequency of detection of the biomarker in the patient. In a further embodiment, the value for the reliability of the detection method of the biomarker and the value for the frequency of detection of the biomarker in the patient are multiplied or averaged in order to build the value for the reliability of detection of the biomarker. In some embodiments of the method, the sub-score is built by a product of values, preferably between 0 and 1, attributed to each feature out of A), B) and C).

[0048] In many embodiments of the method, the score is further based on D) a real-valued quantification of the effect size of the predictive rule for each of the one or more biomarkers. In a further embodiment, the effect size of the predictive rule for each of the one or more biomarkers comprises a measurement of a likelihood of response or resistance or a measurement of a hazard ratio. In a still further embodiment, the effect size of each of the predictive rule for the one or more biomarkers comprises a log likelihood of response or resistance ratio or a log hazard ratio. In a yet still further embodiment, the score comprises the real-valued quantification of the effect size multiplied by a function of values attributed to each feature out of A)-C) for each of the one or more biomarkers.

[0049] In some embodiments of the method, the score is further based on E) whether the predictive rule for each of the one or more biomarkers has been

validated in an indication of the patient, in a related indication, or in any unrelated indication. In a further embodiment, the value for the predictive rule for the biomarker being validated in an indication of the patient is set equal to 1; and wherein the value for the biomarker being validated in a related indication is set
5 equal to a defined positive value between 0 and 1, in particular equal to 0.2, 0.4, 0.6, or 0.8; and wherein the value for the predictive rule for the biomarker being validated in an unrelated indication is set equal to a defined non-negative value less than the value of a related indication, in particular 0, 0.01, 0.02, 0.05, or 0.1. In a further embodiment, the value for the predictive rule for the biomarker being
10 detected in a related indication is weighted responsive to a homology between involved proteins and/or a structural similarity of exchanged amino acids. In a still further embodiment, the sub-score comprises a product of values, preferably between 0 and 1, attributed to each feature out of A), B), C) and D); or out of A), B), C), and E); or out of A), B), C), D), and E).

15 **[0050]** In many embodiments of the method, the score is further based on F) an availability of the associated treatment. In a further embodiment, the value for the availability of the associated treatment comprises a defined value between 0 and 1, wherein a higher value corresponds to a greater availability of the associated treatment. In a still further embodiment, the sub-score comprises a product of
20 values, preferably between 0 and 1, attributed to each feature out of A), B), C) and F); or out of A), B), C), D) and F); or out of A), B), C), E), and F); or out of A), B), C), D), E), and F).

[0051] In some embodiments, the method includes retrieving an identification of a past treatment history of the patient for the indication. The score is further based
25 on G) the past treatment history of the patient. In a further embodiment, the sub-score comprises a product of values, preferably between 0 and 1, attributed to each feature out of A), B), C) and G); or out of A), B), C), D) and G); or out of A), B), C), E) and G); or out of A), B), C), F) and G); or out of A), B), C), D), E) and G); or out of A), B), C), D), F) and G); or out of A), B), C), E), F), and G); or out of A),
30 B), C), D), E), F) and G). In a still further embodiment, the method includes placing a first treatment of the identified plurality of treatments in a first order position, responsive to the first treatment being identified in a standard treatment guideline.

[0052] In another aspect, the present disclosure is directed to a curated database comprising information prepared for being used for generating a score for a treatment to be prioritized when performing one of the described methods for providing a prioritization of patient treatment options. Such information always
5 incorporates information about at least two predictive rules associated with at least one biomarker and associated with a plurality of treatments.

[0053] The curated database may comprise one or more of the following

- data extracted from other databases,
- information extracted from text documents via text mining,
- 10 - measurement data or processed measurement data from appropriate assays,
- manually entered data,
- sign-offs and/or changes by curators.

[0054] The prepared information may comprise information about the predictive rules for the biomarkers with respect to the identified plurality of treatments, and
15 also associated information, in particular information about effect sizes of predictive rules, information about indications or related indications for a treatment, information about clinical validation levels of the predictive rules for biomarkers, information about the involved sources of the predictive rules, information about side effects of treatments, information about whether a treatment is recommended
20 in a standard treatment guideline and/or about an availability of a treatment.

[0055] Preparation of information to be stored in a curated database may comprise mining of document sources, meaning searching for documents comprising identifications of a piece of information having a measure of co-occurrence with identifications of an associated piece of information above a defined threshold.
25 Mining of document sources may be performed continually over time and updated mined databases may be provided at regular intervals. Preparation of information to be stored in curated databases may further comprise extracting information manually by human resources, such as physicians, or other qualified personnel. In particular, a two-step process may be performed, comprising a first step of data

mining, performed by an analysis engine executed by a computing device, in order to pre-select relevant document sources, and a second step of extracting the relevant information manually.

5 **[0056]** The curated database may generally be one database, sometimes also referred to as a drug or treatment response database (DRDB), or a set of databases, sometimes referred to as a set of mined databases.

10 **[0057]** In still another aspect, the present disclosure is directed to a clinical decision support device comprising at least one curated database as described herein. The clinical decision support device may in particular comprise a data mining module and suitable input/output devices.

BRIEF DESCRIPTION OF THE DRAWINGS

15 **[0058]** 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:

[0059] FIG. 1A is a block diagram depicting an embodiment of a network environment comprising client device in communication with server device;

[0060] FIG. 1B is a block diagram depicting a cloud computing environment comprising client device in communication with cloud service providers;

20 **[0061]** FIGs. 1C and 1D are block diagrams depicting embodiments of computing devices useful in connection with the methods and systems described herein;

[0062] FIG. 2 is a block diagram of a clinical decision support device, according to one implementation;

25 **[0063]** FIG. 3A is an exemplary embodiment of a graphical user interface for a clinical decision support device, similar to the device of FIG. 2;

[0064] FIG. 3B is an exemplary embodiment of icons that may provide a clinical user with information regarding the validity of variants;

[0065] FIGS. 4A and 4B are flow diagrams depicting embodiments of a method for selecting and prioritizing possible treatment options for a patient;

[0066] FIG. 5 comprises table 1 as referred to in the examples section; and

[0067] FIG. 6 comprises table 3 as referred to in the examples section.

5

DETAILED DESCRIPTION

[0068] 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
10 contexts.

[0069] The purpose of the scores generated for treatments is to afterwards order them. Therefore, only the comparisons, such as ratios, between score values are relevant, and not the scale of the scores. For instance, if all scores for all treatments are multiplied by some positive number, for example by two, the
15 resulting order does not change. Without loss of generality, the numerical values and ranges referring to scores are scaled such as to achieve a maximum magnitude of one.

[0070] The signs of the scores and sub-scores represent whether the effects that they refer to are desirable or not. Specifically, positive values may be used for
20 response and negative values for resistance and risk of adverse events. If these scores are computed separately, this policy of signing sub-scores is equivalent to first computing and adding (or averaging) positive sub-scores, and afterwards signing the separate scores computed in this way. Further, in the case of computing a single score as a weighted sum of the separate scores, the signing
25 policy is also equivalent to subtracting unsigned weighted separate scores for resistance and risk from the unsigned weighted separate score for response. Therefore, whenever introduction of signs of sub-scores and/or separate scores or related entities is described at one stage, it is understood that equivalent implementations introducing the signs at a different stage but to the same effect,
30 are also referred to.

[0071] The term “biomarker” may be generally used in two different ways. In one definition, biomarkers may be simply any measurable quantities. In an alternate definition used herein, the term “biomarker” may also be used for predictive rules that are based on a biomarker. Such predictive rules may comprise a combination of a measurable quantity (e.g. a biomarker as discussed above in the first definition), a value range, an indication, a treatment option, and/or an effect on the outcome. For example, “response”, “resistance”, and “risk” may be possible qualitative descriptions of the type of effect. Accordingly, via such predictive rules, two otherwise similar cohorts of patients with a given indication may be compared, where the first cohort comprises patients with a biomarker measurement value outside a given range and the second cohort comprises patients with the biomarker measurement value inside the given range. The outcomes achieved by a given treatment in both cohorts may differ, as described by a given effect on the outcome. In some implementations, these predictive rules may be referred to variously as an “actionable biomarker”, a “predictive biomarker”, or a “theranostic biomarker”. A biomarker or measured quantity may apply to more than one predictive rule, for example related to different indications or different drugs. Accordingly, in some instances of the term “biomarker” in this disclosure, the predictive rule may rather be meant than the strict biomarker. However, for the person skilled in the art, this will be clear from the context as well.

[0072] Biomarkers may be genomic (e.g. single nucleotide variations (SNV), Copy number variations (CNV), insertions, deletions, gene fusions, polyploidy, gene expression, miRNA's), proteomic (e.g. post-translational modifications, expression), metabolites, electrolytes, physiological parameters (e.g. blood pressure), patient age, patient weight, patient ethnicity, disease indication, patient disease stage, patient disease grade, patient co-morbidities. Moreover, novel biomarkers may be inferred or predicted based on functional similarities to established biomarkers. For example, in the case of a novel variant such as BRAF V600D, it can be inferred or predicted that it too may be predictive of response to vemurafenib, by virtue of the structural similarity between an aspartate (D) and a glutamate (E). In other words, the V600D mutation should have similar functional consequences for the BRAF protein as the predictive V600E mutation and as such it can be inferred that the former may hold similar biomarker properties to the

latter. Further, an instance can be considered where over-expression of a protein/gene is predictive of drug response. Functionally this normally results in increased protein/gene activity. Thus, if a novel mutation in this gene is identified, which is predicted to activate the protein, then it can be inferred that this

5 mutationally activated protein has similar biomarker properties to the over-expressed predictive gene. By corollary, a mutation that inactivates a gene might be predicted to hold similar biomarker properties to an established biomarker that relies on under-expression or deletion of a specific gene. While such inferred biomarkers are clearly predictions, they can be validated in the laboratory before

10 deciding to treat a patient. Alternatively, if the patient need is particularly pressing, a clinician may decide to proceed with an associated treatment after ensuring that they are satisfied with the evidence supporting the prediction.

[0073] The term "drug" may be used interchangeably with the term "treatment", because the systems and methods discussed herein may be readily applied to

15 non-drug based treatments. In some of the corresponding implementations, treatments may include: targeted therapy, chemotherapy, any drug or drug combination, surgery, radiation, laser ablation, vaccination, biological therapy, immunotherapy, stem cell transplant, transfusion, transplantation (e.g. bone marrow), hyperthermia, photodynamic therapy, nutritional adjustment (e.g.

20 fasting), physical exercise.

[0074] The prioritization of treatments is aimed to primarily reflect the anticipated efficacy of the treatment and the risk of adverse events, which is often called just "risk" in the following. Statements on the efficacy are often distinguished into "response", which refers to high or increased efficacy, and "resistance", which

25 refers to low or decreased efficacy, where the comparisons refer to an implied base level, often defined by the genetic wildtype. The term "risk" is used to represent the possibility of any undesired side effect of the treatment. Any treatment may have the potential to cause a set of side effects that differ in type and severity, e.g. including headache and fatal liver injury. The Common Toxicity

30 Criteria (CTC), also referred to as the Common Terminology Criteria for Adverse Events (CTCAE), is a standardized classification of side effects used in assessing drugs for cancer therapy. It is a product of the National Cancer Institute. Currently version 4.0 is in use, released in 2009. Most US and UK drug trials base their

observations on this system, which has a range of grades from 1-5. Specific conditions and symptoms may have values or descriptive comment for each level, but the general guideline is: 1 = Mild; 2 = Moderate; 3 = Severe; 4 = Life threatening; 5 = Death. Such grades may be used for scoring the degree of risk associated with a patient treatment. In some embodiments of the presented methods, this may lead to five further separate scores to be computed, besides separate scores for resistance and response. Alternatively, side effects associated with the MedDRA ontology may also be used. MedDRA or Medical Dictionary for Regulatory Activities is a clinically validated international medical terminology used by regulatory authorities and the regulated biopharmaceutical industry during the regulatory process, from pre-marketing to post-marketing activities, and for data entry, retrieval, evaluation, and presentation. In addition, it is the adverse event classification dictionary endorsed by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). MedDRA is used in the United States, European Union, and Japan. Its use is currently mandated in Europe and Japan for safety reporting.

[0075] The term "risk/benefit preferences" refers to associated combinations of treatment efficacy with potential adverse events for a given patient. Depending on life situation and personal preferences of the patient, the optimal choice from a set of such combinations may vary. For instance, a cancer patient with a very bad prognosis, e.g. with only a few weeks to live, whatever the treatment, may prefer to not receive chemotherapy, even if this decision further shortens his/her life by a week, for the sake of not having to suffer from the side effects of the chemotherapy. Particularly useful is this concept for patients advanced in years who often do not accept risks of specific side effects.

[0076] The term "outcome" refers to any clinical measure of condition of a patient after treatment, which may hence be (partially) caused by the treatment. Examples of commonly assessed outcomes are overall survival (OS); progress-free survival (PFS); objective response (OR); response-classifications, eg into complete response, partial response, stable disease, or progressive disease; and, in the case of tumors, tumor size.

[0077] The term “effect size” refers to a numerical value that quantifies the magnitude of any given effect of the drug or the frequency of any given effect or any mathematical combination thereof, such as an addition or an average. The magnitude of an effect may in particular refer to an overall survival time, which is a very important measure of efficacy. The frequency of an effect may in particular refer to a percentage of treated patients that have been observed to experience resistance, response or to suffer from a side effect. Effect sizes are often taken relative to an appropriate baseline; for instance, a hazard ratio can be used as a quantitative description of the increase or decrease in overall survival achieved by some medication in comparison to a placebo.

[0078] The term “past treatment history of the patient” refers to a medical history of a patient, in particular the history of treatments and associated outcomes. The past treatment history may provide valuable information on response, resistance, and risk of a treatment for that patient. For example, if a former tumor already turned out to be resistant to a treatment, it may be conjectured that a recurrence will also be resistant to that treatment. One way to incorporate such information into the score is to assign high weights to predictive rules that predict resistance to this treatment and/or to assign low weights to predictive rules that predict response.

[0079] A further concept of the invention is a “clinical need” leading to understanding the clinical utility (i.e. actionability) of various biomarker validities. For example, in the absence of an obvious treatment path for a patient, biomarkers with preclinical or inferred validity may be considered clinically actionable by physicians. For example, a preclinically observed biomarker may provide sufficient evidence for assigning a late stage cancer patient to a specific phase I clinical trial. However, this same biomarker would not be considered clinically actionable in the case of a primary cancer patient where established treatment guidelines exist. Clearly the treatment guideline would take precedence over pre-clinical or inferred biomarker predictions. The clinical actionability of a biomarker can therefore be related to the validity level of the biomarker and the degree of clinical need for a patient. Where the clinical need is high e.g. no established treatment guideline, then less reliable biomarker validities may be considered to be clinically actionable by physicians. In the context where

guidelines exist, clinicians may opt to consider only endorsed or well-validated clinical biomarkers, since the degree of clinical need would be considered less. Nevertheless, clinical need is also related to the survival rates of particular cancer types. In the context of a cancer type such as glioblastoma, where guidelines exist
5 but survival rates are poor, physicians may be even more inclined to use endorsed biomarkers from other indications and/or clinical/pre-clinically validated biomarkers. For this reason, it is possible to define a level of clinical need (e.g. very high – no guideline, poor survival; high – no guideline, moderate survival or guideline but poor survival; medium – guideline but moderate survival; low –
10 guideline with good survival) and use this parameter to normalize the clinical actionability of identified biomarkers. The degree of clinical need may therefore be used as an input parameter for biomarker assessment and treatment prioritization.

[0080] For purposes of reading the description of the various embodiments below, the following descriptions of the sections of the specification and their respective
15 contents may be helpful.

[0081] Section A describes a network environment and computing environment which may be useful for practicing embodiments described herein.

[0082] Section B describes embodiments of systems and methods for prioritizing clinical treatments with a Clinical Decision Support Device.

20 A. Computing and Network Environment

[0083] Prior to discussing specific embodiments of the present solution, it may be helpful to describe aspects of the operating environment as well as associated system components (e.g., hardware elements) in connection with the methods and systems described herein. Referring to FIG. 1A, an embodiment of a network
25 environment is depicted. In brief overview, the network environment includes one or more clients 102a-102n (also generally referred to as local machine(s) 102, client(s) 102, client node(s) 102, client machine(s) 102, client computer(s) 102, client device(s) 102, endpoint(s) 102, or endpoint node(s) 102) in communication with one or more servers 106a-106n (also generally referred to as server(s) 106,
30 node 106, or remote machine(s) 106) via one or more networks 104. In some embodiments, a client 102 has the capacity to function as both a client node

seeking access to resources provided by a server and as a server providing access to hosted resources for other clients 102a-102n.

5 **[0084]** Although FIG. 1A shows a network 104 between the clients 102 and the servers 106, the clients 102 and the servers 106 may be on the same network 104. In some embodiments, there are multiple networks 104 between the clients 102 and the servers 106. In one of these embodiments, a network 104' (not shown) may be a private network and a network 104 may be a public network. In another of these embodiments, a network 104 may be a private network and a network 104' a public network. In still another of these embodiments, networks 104
10 and 104' may both be private networks.

[0085] The network 104 may be connected via wired or wireless links. Wired links may include Digital Subscriber Line (DSL), coaxial cable lines, or optical fiber lines. The wireless links may include BLUETOOTH, Wi-Fi, Worldwide Interoperability for Microwave Access (WiMAX), an infrared channel or satellite
15 band. The wireless links may also include any cellular network standards used to communicate among mobile devices, including standards that qualify as 1G, 2G, 3G, or 4G. The network standards may qualify as one or more generation of mobile telecommunication standards by fulfilling a specification or standards such as the specifications maintained by International Telecommunication Union. The
20 3G standards, for example, may correspond to the International Mobile Telecommunications-2000 (IMT-2000) specification, and the 4G standards may correspond to the International Mobile Telecommunications Advanced (IMT-Advanced) specification. Examples of cellular network standards include AMPS, GSM, GPRS, UMTS, LTE, LTE Advanced, Mobile WiMAX, and WiMAX-Advanced.
25 Cellular network standards may use various channel access methods e.g. FDMA, TDMA, CDMA, or SDMA. In some embodiments, different types of data may be transmitted via different links and standards. In other embodiments, the same types of data may be transmitted via different links and standards.

[0086] The network 104 may be any type and/or form of network. The
30 geographical scope of the network 104 may vary widely and the network 104 can be a body area network (BAN), a personal area network (PAN), a local-area network (LAN), e.g. Intranet, a metropolitan area network (MAN), a wide area

network (WAN), or the Internet. The topology of the network 104 may be of any form and may include, e.g., any of the following: point-to-point, bus, star, ring, mesh, or tree. The network 104 may be an overlay network which is virtual and sits on top of one or more layers of other networks 104'. The network 104 may be of any such network topology as known to those ordinarily skilled in the art capable of supporting the operations described herein. The network 104 may utilize different techniques and layers or stacks of protocols, including, e.g., the Ethernet protocol, the internet protocol suite (TCP/IP), the ATM (Asynchronous Transfer Mode) technique, the SONET (Synchronous Optical Networking) protocol, or the SDH (Synchronous Digital Hierarchy) protocol. The TCP/IP internet protocol suite may include application layer, transport layer, internet layer (including, e.g., IPv6), or the link layer. The network 104 may be a type of a broadcast network, a telecommunications network, a data communication network, or a computer network.

[0087] In some embodiments, the system may include multiple, logically-grouped servers 106. In one of these embodiments, the logical group of servers may be referred to as a server farm 38 or a machine farm 38. In another of these embodiments, the servers 106 may be geographically dispersed. In other embodiments, a machine farm 38 may be administered as a single entity. In still other embodiments, the machine farm 38 includes a plurality of machine farms 38. The servers 106 within each machine farm 38 can be heterogeneous – one or more of the servers 106 or machines 106 can operate according to one type of operating system platform (e.g., WINDOWS NT, manufactured by Microsoft Corp. of Redmond, Washington), while one or more of the other servers 106 can operate on according to another type of operating system platform (e.g., Unix, Linux, or Mac OS X).

[0088] In one embodiment, servers 106 in the machine farm 38 may be stored in high-density rack systems, along with associated storage systems, and located in an enterprise data center. In this embodiment, consolidating the servers 106 in this way may improve system manageability, data security, the physical security of the system, and system performance by locating servers 106 and high performance storage systems on localized high performance networks. Centralizing the servers

106 and storage systems and coupling them with advanced system management tools allows more efficient use of server resources.

[0089] The servers 106 of each machine farm 38 do not need to be physically proximate to another server 106 in the same machine farm 38. Thus, the group of servers 106 logically grouped as a machine farm 38 may be interconnected using a wide-area network (WAN) connection or a metropolitan-area network (MAN) connection. For example, a machine farm 38 may include servers 106 physically located in different continents or different regions of a continent, country, state, city, campus, or room. Data transmission speeds between servers 106 in the machine farm 38 can be increased if the servers 106 are connected using a local-area network (LAN) connection or some form of direct connection. Additionally, a heterogeneous machine farm 38 may include one or more servers 106 operating according to a type of operating system, while one or more other servers 106 execute one or more types of hypervisors rather than operating systems. In these embodiments, hypervisors may be used to emulate virtual hardware, partition physical hardware, virtualize physical hardware, and execute virtual machines that provide access to computing environments, allowing multiple operating systems to run concurrently on a host computer. Native hypervisors may run directly on the host computer. Hypervisors may include VMware ESX/ESXi, manufactured by VMware, Inc., of Palo Alto, California; the Xen hypervisor, an open source product whose development is overseen by Citrix Systems, Inc.; the HYPER-V hypervisors provided by Microsoft or others. Hosted hypervisors may run within an operating system on a second software level. Examples of hosted hypervisors may include VMware Workstation and VIRTUALBOX.

[0090] Management of the machine farm 38 may be de-centralized. For example, one or more servers 106 may comprise components, subsystems and modules to support one or more management services for the machine farm 38. In one of these embodiments, one or more servers 106 provide functionality for management of dynamic data, including techniques for handling failover, data replication, and increasing the robustness of the machine farm 38. Each server 106 may communicate with a persistent store and, in some embodiments, with a dynamic store.

[0091] Server 106 may be a file server, application server, web server, proxy server, appliance, network appliance, gateway, gateway server, virtualization server, deployment server, SSL VPN server, or firewall. In one embodiment, the server 106 may be referred to as a remote machine or a node. In another
5 embodiment, a plurality of nodes 290 may be in the path between any two communicating servers.

[0092] Referring to Fig. 1B, a cloud computing environment is depicted. A cloud computing environment may provide client 102 with one or more resources provided by a network environment. The cloud computing environment may
10 include one or more clients 102a-102n, in communication with the cloud 108 over one or more networks 104. Clients 102 may include, e.g., thick clients, thin clients, and zero clients. A thick client may provide at least some functionality even when disconnected from the cloud 108 or servers 106. A thin client or a zero client may depend on the connection to the cloud 108 or server 106 to provide functionality. A
15 zero client may depend on the cloud 108 or other networks 104 or servers 106 to retrieve operating system data for the client device. The cloud 108 may include back end platforms, e.g., servers 106, storage, server farms or data centers.

[0093] The cloud 108 may be public, private, or hybrid. Public clouds may include public servers 106 that are maintained by third parties to the clients 102 or the
20 owners of the clients. The servers 106 may be located off-site in remote geographical locations as disclosed above or otherwise. Public clouds may be connected to the servers 106 over a public network. Private clouds may include private servers 106 that are physically maintained by clients 102 or owners of clients. Private clouds may be connected to the servers 106 over a private network
25 104. Hybrid clouds 108 may include both the private and public networks 104 and servers 106.

[0094] The cloud 108 may also include a cloud based delivery, e.g. Software as a Service (SaaS) 110, Platform as a Service (PaaS) 112, and Infrastructure as a Service (IaaS) 114. IaaS may refer to a user renting the use of infrastructure
30 resources that are needed during a specified time period. IaaS providers may offer storage, networking, servers or virtualization resources from large pools, allowing the users to quickly scale up by accessing more resources as needed. Examples

of IaaS include AMAZON WEB SERVICES provided by Amazon.com, Inc., of Seattle, Washington, RACKSPACE CLOUD provided by Rackspace US, Inc., of San Antonio, Texas, Google Compute Engine provided by Google Inc. of Mountain View, California, or RIGHTSCALE provided by RightScale, Inc., of Santa Barbara, California. PaaS providers may offer functionality provided by IaaS, including, e.g., storage, networking, servers or virtualization, as well as additional resources such as, e.g., the operating system, middleware, or runtime resources. Examples of PaaS include WINDOWS AZURE provided by Microsoft Corporation of Redmond, Washington, Google App Engine provided by Google Inc., and HEROKU provided by Heroku, Inc. of San Francisco, California. SaaS providers may offer the resources that PaaS provides, including storage, networking, servers, virtualization, operating system, middleware, or runtime resources. In some embodiments, SaaS providers may offer additional resources including, e.g., data and application resources. Examples of SaaS include GOOGLE APPS provided by Google Inc., SALESFORCE provided by Salesforce.com Inc. of San Francisco, California, or OFFICE 365 provided by Microsoft Corporation. Examples of SaaS may also include data storage providers, e.g. DROPBOX provided by Dropbox, Inc. of San Francisco, California, Microsoft SKYDRIVE provided by Microsoft Corporation, Google Drive provided by Google Inc., or Apple ICLOUD provided by Apple Inc. of Cupertino, California.

[0095] Clients 102 may access IaaS resources with one or more IaaS standards, including, e.g., Amazon Elastic Compute Cloud (EC2), Open Cloud Computing Interface (OCCI), Cloud Infrastructure Management Interface (CIMI), or OpenStack standards. Some IaaS standards may allow clients access to resources over HTTP, and may use Representational State Transfer (REST) protocol or Simple Object Access Protocol (SOAP). Clients 102 may access PaaS resources with different PaaS interfaces. Some PaaS interfaces use HTTP packages, standard Java APIs, JavaMail API, Java Data Objects (JDO), Java Persistence API (JPA), Python APIs, web integration APIs for different programming languages including, e.g., Rack for Ruby, WSGI for Python, or PSGI for Perl, or other APIs that may be built on REST, HTTP, XML, or other protocols. Clients 102 may access SaaS resources through the use of web-based user interfaces, provided by a web browser (e.g. GOOGLE CHROME, Microsoft

INTERNET EXPLORER, or Mozilla Firefox provided by Mozilla Foundation of Mountain View, California). Clients 102 may also access SaaS resources through smartphone or tablet applications, including ,e.g., Salesforce Sales Cloud, or Google Drive app. Clients 102 may also access SaaS resources through the client
5 operating system, including, e.g., Windows file system for DROPBOX.

[0096] In some embodiments, access to IaaS, PaaS, or SaaS resources may be authenticated. For example, a server or authentication server may authenticate a user via security certificates, HTTPS, or API keys. API keys may include various encryption standards such as, e.g., Advanced Encryption Standard (AES). Data
10 resources may be sent over Transport Layer Security (TLS) or Secure Sockets Layer (SSL).

[0097] The client 102 and server 106 may be deployed as and/or executed on any type and form of computing device, e.g. a computer, network device or appliance capable of communicating on any type and form of network and performing the
15 operations described herein. FIGs. 1C and 1D depict block diagrams of a computing device 100 useful for practicing an embodiment of the client 102 or a server 106. As shown in FIGs. 1C and 1D, each computing device 100 includes a central processing unit 121, and a main memory unit 122. As shown in FIG. 1C, a computing device 100 may include a storage device 128, an installation device
20 116, a network interface 118, an I/O controller 123, display devices 124a-124n, a keyboard 126 and a pointing device 127, e.g. a mouse. The storage device 128 may include, without limitation, an operating system, software, and a software of Clinical Decision Support Device 120. As shown in FIG. 1D, each computing device 100 may also include additional optional elements, e.g. a memory port 103,
25 a bridge 170, one or more input/output devices 130a-130n (generally referred to using reference numeral 130), and a cache memory 140 in communication with the central processing unit 121.

[0098] The central processing unit 121 is any logic circuitry that responds to and processes instructions fetched from the main memory unit 122. In many
30 embodiments, the central processing unit 121 is provided by a microprocessor unit, e.g.: those manufactured by Intel Corporation of Mountain View, California; those manufactured by Motorola Corporation of Schaumburg, Illinois; the ARM

processor and TEGRA system on a chip (SoC) manufactured by Nvidia of Santa Clara, California; the POWER7 processor, those manufactured by International Business Machines of White Plains, New York; or those manufactured by Advanced Micro Devices of Sunnyvale, California. The computing device 100 may
5 be based on any of these processors, or any other processor capable of operating as described herein. The central processing unit 121 may utilize instruction level parallelism, thread level parallelism, different levels of cache, and multi-core processors. A multi-core processor may include two or more processing units on a single computing component. Examples of a multi-core processors include the
10 AMD PHENOM IIX2, INTEL CORE i5 and INTEL CORE i7.

[0099] Main memory unit 122 may include one or more memory chips capable of storing data and allowing any storage location to be directly accessed by the microprocessor 121. Main memory unit 122 may be volatile and faster than storage 128 memory. Main memory units 122 may be Dynamic random access
15 memory (DRAM) or any variants, including static random access memory (SRAM), Burst SRAM or SynchBurst SRAM (BSRAM), Fast Page Mode DRAM (FPM DRAM), Enhanced DRAM (EDRAM), Extended Data Output RAM (EDO RAM), Extended Data Output DRAM (EDO DRAM), Burst Extended Data Output DRAM (BEDO DRAM), Single Data Rate Synchronous DRAM (SDR SDRAM), Double
20 Data Rate SDRAM (DDR SDRAM), Direct Rambus DRAM (DRDRAM), or Extreme Data Rate DRAM (XDR DRAM). In some embodiments, the main memory 122 or the storage 128 may be non-volatile; e.g., non-volatile read access memory (NVRAM), flash memory non-volatile static RAM (nvSRAM), Ferroelectric RAM (FeRAM), Magnetoresistive RAM (MRAM), Phase-change memory (PRAM),
25 conductive-bridging RAM (CBRAM), Silicon-Oxide-Nitride-Oxide-Silicon (SONOS), Resistive RAM (RRAM), Racetrack, Nano-RAM (NRAM), or Millipede memory. The main memory 122 may be based on any of the above described memory chips, or any other available memory chips capable of operating as described herein. In the embodiment shown in FIG. 1C, the processor 121 communicates
30 with main memory 122 via a system bus 150 (described in more detail below). FIG. 1D depicts an embodiment of a computing device 100 in which the processor communicates directly with main memory 122 via a memory port 103. For example, in FIG. 1D the main memory 122 may be DRDRAM.

[00100] FIG. 1D depicts an embodiment in which the main processor 121 communicates directly with cache memory 140 via a secondary bus, sometimes referred to as a backside bus. In other embodiments, the main processor 121 communicates with cache memory 140 using the system bus 150. Cache memory
5 140 typically has a faster response time than main memory 122 and is typically provided by SRAM, BSRAM, or EDRAM. In the embodiment shown in FIG. 1D, the processor 121 communicates with various I/O devices 130 via a local system bus 150. Various buses may be used to connect the central processing unit 121 to any of the I/O devices 130, including a PCI bus, a PCI-X bus, or a PCI-Express bus, or
10 a NuBus. For embodiments in which the I/O device is a video display 124, the processor 121 may use an Advanced Graphics Port (AGP) to communicate with the display 124 or the I/O controller 123 for the display 124. FIG. 1D depicts an embodiment of a computer 100 in which the main processor 121 communicates directly with I/O device 130b or other processors 121' via HYPERTRANSPORT,
15 RAPIDIO, or INFINIBAND communications technology. FIG. 1D also depicts an embodiment in which local busses and direct communication are mixed: the processor 121 communicates with I/O device 130a using a local interconnect bus while communicating with I/O device 130b directly.

[00101] A wide variety of I/O devices 130a-130n may be present in the
20 computing device 100. Input devices may include keyboards, mice, trackpads, trackballs, touchpads, touch mice, multi-touch touchpads and touch mice, microphones, multi-array microphones, drawing tablets, cameras, single-lens reflex camera (SLR), digital SLR (DSLR), CMOS sensors, accelerometers, infrared optical sensors, pressure sensors, magnetometer sensors, angular rate
25 sensors, depth sensors, proximity sensors, ambient light sensors, gyroscopic sensors, or other sensors. Output devices may include video displays, graphical displays, speakers, headphones, inkjet printers, laser printers, and 3D printers.

[00102] Devices 130a-130n may include a combination of multiple input or output devices, including, e.g., Microsoft KINECT, Nintendo Wiimote for the Wii,
30 Nintendo Wii U GAMEPAD, or Apple IPHONE. Some devices 130a-130n allow gesture recognition inputs through combining some of the inputs and outputs. Some devices 130a-130n provides for facial recognition which may be utilized as an input for different purposes including authentication and other commands.

Some devices 130a-130n provides for voice recognition and inputs, including, e.g., Microsoft KINECT, SIRI for IPHONE by Apple, Google Now or Google Voice Search.

[00103] Additional devices 130a-130n have both input and output capabilities, including, e.g., haptic feedback devices, touchscreen displays, or multi-touch displays. Touchscreen, multi-touch displays, touchpads, touch mice, or other touch sensing devices may use different technologies to sense touch, including, e.g., capacitive, surface capacitive, projected capacitive touch (PCT), in-cell capacitive, resistive, infrared, waveguide, dispersive signal touch (DST), in-cell optical, surface acoustic wave (SAW), bending wave touch (BWT), or force-based sensing technologies. Some multi-touch devices may allow two or more contact points with the surface, allowing advanced functionality including, e.g., pinch, spread, rotate, scroll, or other gestures. Some touchscreen devices, including, e.g., Microsoft PIXELSENSE or Multi-Touch Collaboration Wall, may have larger surfaces, such as on a table-top or on a wall, and may also interact with other electronic devices. Some I/O devices 130a-130n, display devices 124a-124n or group of devices may be augment reality devices. The I/O devices may be controlled by an I/O controller 123 as shown in FIG. 1C. The I/O controller may control one or more I/O devices, such as, e.g., a keyboard 126 and a pointing device 127, e.g., a mouse or optical pen. Furthermore, an I/O device may also provide storage and/or an installation medium 116 for the computing device 100. In still other embodiments, the computing device 100 may provide USB connections (not shown) to receive handheld USB storage devices. In further embodiments, an I/O device 130 may be a bridge between the system bus 150 and an external communication bus, e.g. a USB bus, a SCSI bus, a FireWire bus, an Ethernet bus, a Gigabit Ethernet bus, a Fibre Channel bus, or a Thunderbolt bus.

[00104] In some embodiments, display devices 124a-124n may be connected to I/O controller 123. Display devices may include, e.g., liquid crystal displays (LCD), thin film transistor LCD (TFT-LCD), blue phase LCD, electronic papers (e-ink) displays, flexile displays, light emitting diode displays (LED), digital light processing (DLP) displays, liquid crystal on silicon (LCOS) displays, organic light-emitting diode (OLED) displays, active-matrix organic light-emitting diode

(AMOLED) displays, liquid crystal laser displays, time-multiplexed optical shutter (TMOS) displays, or 3D displays. Examples of 3D displays may use, e.g.

stereoscopy, polarization filters, active shutters, or autostereoscopy. Display devices 124a-124n may also be a head-mounted display (HMD). In some

5 embodiments, display devices 124a-124n or the corresponding I/O controllers 123 may be controlled through or have hardware support for OPENGGL or DIRECTX API or other graphics libraries.

[00105] In some embodiments, the computing device 100 may include or connect to multiple display devices 124a-124n, which each may be of the same or
10 different type and/or form. As such, any of the I/O devices 130a-130n and/or the I/O controller 123 may include any type and/or form of suitable hardware, software, or combination of hardware and software to support, enable or provide for the connection and use of multiple display devices 124a-124n by the computing device 100. For example, the computing device 100 may include any
15 type and/or form of video adapter, video card, driver, and/or library to interface, communicate, connect or otherwise use the display devices 124a-124n. In one embodiment, a video adapter may include multiple connectors to interface to multiple display devices 124a-124n. In other embodiments, the computing device 100 may include multiple video adapters, with each video adapter connected to
20 one or more of the display devices 124a-124n. In some embodiments, any portion of the operating system of the computing device 100 may be configured for using multiple displays 124a-124n. In other embodiments, one or more of the display devices 124a-124n may be provided by one or more other computing devices 100a or 100b connected to the computing device 100, via the network 104. In
25 some embodiments software may be designed and constructed to use another computer's display device as a second display device 124a for the computing device 100. For example, in one embodiment, an Apple iPad may connect to a computing device 100 and use the display of the device 100 as an additional display screen that may be used as an extended desktop. One ordinarily skilled in
30 the art will recognize and appreciate the various ways and embodiments that a computing device 100 may be configured to have multiple display devices 124a-124n.

[00106] Referring again to FIG. 1C, the computing device 100 may comprise a storage device 128 (e.g. one or more hard disk drives or redundant arrays of independent disks) for storing an operating system or other related software, and for storing application software programs such as any program related to the software 120 for the experiment tracker system. Examples of storage device 128 include, e.g., hard disk drive (HDD); optical drive including CD drive, DVD drive, or BLU-RAY drive; solid-state drive (SSD); USB flash drive; or any other device suitable for storing data. Some storage devices may include multiple volatile and non-volatile memories, including, e.g., solid state hybrid drives that combine hard disks with solid state cache. Some storage device 128 may be non-volatile, mutable, or read-only. Some storage device 128 may be internal and connect to the computing device 100 via a bus 150. Some storage device 128 may be external and connect to the computing device 100 via a I/O device 130 that provides an external bus. Some storage device 128 may connect to the computing device 100 via the network interface 118 over a network 104, including, e.g., the Remote Disk for MACBOOK AIR by Apple. Some client devices 100 may not require a non-volatile storage device 128 and may be thin clients or zero clients 102. Some storage device 128 may also be used as a installation device 116, and may be suitable for installing software and programs. Additionally, the operating system and the software can be run from a bootable medium, for example, a bootable CD, e.g. KNOPPIX, a bootable CD for GNU/Linux that is available as a GNU/Linux distribution from knoppix.net.

[00107] Client device 100 may also install software or application from an application distribution platform. Examples of application distribution platforms include the App Store for iOS provided by Apple, Inc., the Mac App Store provided by Apple, Inc., GOOGLE PLAY for Android OS provided by Google Inc., Chrome Webstore for CHROME OS provided by Google Inc., and Amazon Appstore for Android OS and KINDLE FIRE provided by Amazon.com, Inc. An application distribution platform may facilitate installation of software on a client device 102. An application distribution platform may include a repository of applications on a server 106 or a cloud 108, which the clients 102a-102n may access over a network 104. An application distribution platform may include application developed and provided by various developers. A user of a client device 102 may

select, purchase and/or download an application via the application distribution platform.

[00108] Furthermore, the computing device 100 may include a network interface 118 to interface to the network 104 through a variety of connections including, but not limited to, standard telephone lines LAN or WAN links (*e.g.*, 802.11, T1, T3, Gigabit Ethernet, Infiniband), broadband connections (*e.g.*, ISDN, Frame Relay, ATM, Gigabit Ethernet, Ethernet-over-SONET, ADSL, VDSL, BPON, GPON, fiber optical including FiOS), wireless connections, or some combination of any or all of the above. Connections can be established using a variety of communication protocols (*e.g.*, TCP/IP, Ethernet, ARCNET, SONET, SDH, Fiber Distributed Data Interface (FDDI), IEEE 802.11a/b/g/n/ac CDMA, GSM, WiMax and direct asynchronous connections). In one embodiment, the computing device 100 communicates with other computing devices 100' via any type and/or form of gateway or tunneling protocol *e.g.* Secure Socket Layer (SSL) or Transport Layer Security (TLS), or the Citrix Gateway Protocol manufactured by Citrix Systems, Inc. of Ft. Lauderdale, Florida. The network interface 118 may comprise a built-in network adapter, network interface card, PCMCIA network card, EXPRESSCARD network card, card bus network adapter, wireless network adapter, USB network adapter, modem or any other device suitable for interfacing the computing device 100 to any type of network capable of communication and performing the operations described herein.

[00109] A computing device 100 of the sort depicted in FIGs. 1B and 1C may operate under the control of an operating system, which controls scheduling of tasks and access to system resources. The computing device 100 can be running any operating system such as any of the versions of the MICROSOFT WINDOWS operating systems, the different releases of the Unix and Linux operating systems, any version of the MAC OS for Macintosh computers, any embedded operating system, any real-time operating system, any open source operating system, any proprietary operating system, any operating systems for mobile computing devices, or any other operating system capable of running on the computing device and performing the operations described herein. Typical operating systems include, but are not limited to: WINDOWS 2000, WINDOWS Server 2012, WINDOWS CE, WINDOWS Phone, WINDOWS XP, WINDOWS VISTA, and

WINDOWS 7, WINDOWS RT, and WINDOWS 8 all of which are manufactured by Microsoft Corporation of Redmond, Washington; MAC OS and iOS, manufactured by Apple, Inc. of Cupertino, California; and Linux, a freely-available operating system, e.g. Linux Mint distribution ("distro") or Ubuntu, distributed by Canonical
5 Ltd. of London, United Kingdom; or Unix or other Unix-like derivative operating systems; and Android, designed by Google, of Mountain View, California, among others. Some operating systems, including, e.g., the CHROME OS by Google, may be used on zero clients or thin clients, including, e.g., CHROMEBOOKS.

[00110] The computer system 100 can be any workstation, telephone,
10 desktop computer, laptop or notebook computer, netbook, ULTRABOOK, tablet, server, handheld computer, mobile telephone, smartphone or other portable telecommunications device, media playing device, a gaming system, mobile computing device, or any other type and/or form of computing, telecommunications or media device that is capable of communication. The
15 computer system 100 has sufficient processor power and memory capacity to perform the operations described herein. In some embodiments, the computing device 100 may have different processors, operating systems, and input devices consistent with the device. The Samsung GALAXY smartphones, e.g., operate under the control of Android operating system developed by Google, Inc. GALAXY
20 smartphones receive input via a touch interface.

[00111] In some embodiments, the computing device 100 is a gaming system. For example, the computer system 100 may comprise a PLAYSTATION 3, or PERSONAL PLAYSTATION PORTABLE (PSP), or a PLAYSTATION VITA device manufactured by the Sony Corporation of Tokyo, Japan, a NINTENDO DS,
25 NINTENDO 3DS, NINTENDO WII, or a NINTENDO WII U device manufactured by Nintendo Co., Ltd., of Kyoto, Japan, an XBOX 360 device manufactured by the Microsoft Corporation of Redmond, Washington.

[00112] In some embodiments, the computing device 100 is a digital audio player such as the Apple IPOD, IPOD Touch, and IPOD NANO lines of devices,
30 manufactured by Apple Computer of Cupertino, California. Some digital audio players may have other functionality, including, e.g., a gaming system or any functionality made available by an application from a digital application distribution

platform. For example, the IPOD Touch may access the Apple App Store. In some embodiments, the computing device 100 is a portable media player or digital audio player supporting file formats including, but not limited to, MP3, WAV, M4A/AAC, WMA Protected AAC, AIFF, Audible audiobook, Apple Lossless audio file formats and .mov, .m4v, and .mp4 MPEG-4 (H.264/MPEG-4 AVC) video file formats.

[00113] In some embodiments, the computing device 100 is a tablet e.g. the IPAD line of devices by Apple; GALAXY TAB family of devices by Samsung; or KINDLE FIRE, by Amazon.com, Inc. of Seattle, Washington. In other embodiments, the computing device 100 is a eBook reader, e.g. the KINDLE family of devices by Amazon.com, or NOOK family of devices by Barnes & Noble, Inc. of New York City, New York.

[00114] In some embodiments, the communications device 102 includes a combination of devices, e.g. a smartphone combined with a digital audio player or portable media player. For example, one of these embodiments is a smartphone, e.g. the IPHONE family of smartphones manufactured by Apple, Inc.; a Samsung GALAXY family of smartphones manufactured by Samsung, Inc; or a Motorola DROID family of smartphones. In yet another embodiment, the communications device 102 is a laptop or desktop computer equipped with a web browser and a microphone and speaker system, e.g. a telephony headset. In these embodiments, the communications devices 102 are web-enabled and can receive and initiate phone calls. In some embodiments, a laptop or desktop computer is also equipped with a webcam or other video capture device that enables video chat and video call.

[00115] In some embodiments, the status of one or more machines 102, 106 in the network 104 is monitored, generally as part of network management. In one of these embodiments, the status of a machine may include an identification of load information (e.g., the number of processes on the machine, CPU and memory utilization), of port information (e.g., the number of available communication ports and the port addresses), or of session status (e.g., the duration and type of processes, and whether a process is active or idle). In another of these embodiments, this information may be identified by a plurality of metrics, and the plurality of metrics can be applied at least in part towards decisions in load

distribution, network traffic management, and network failure recovery as well as any aspects of operations of the present solution described herein. Aspects of the operating environments and components described above will become apparent in the context of the systems and methods disclosed herein.

5 B. Clinical Decision Support Device

[00116] The rapid advancement of high-throughput technologies available for generating large-scale molecular-level measurements in human populations has led to an increased interest in the discovery and validation of molecular biomarkers in clinical research. Biomarkers are generally defined as any
10 “biological characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”. Various types of biomarkers may include genomic biomarkers (e.g. single nucleotide variations (SNV), copy number variations (CNV), insertions, deletions, gene fusions, polyploidy, gene expression,
15 miRNA’s), proteomic biomarkers (e.g. post-translational modifications, expression), metabolites, electrolytes, physiological parameters (e.g. blood pressure), patient age, patient weight, or patient co-morbidities. Uses of biomarkers for clinical decision making is quite varied and includes identification of predictive and prognostic factors for disease management, surrogate endpoints for
20 monitoring clinical response to an intervention and early detection of disease.

[00117] In oncology, for example, tumor biomarkers may be used to determine a patient’s current clinical status and to predict disease drivers and mechanisms that might be modified by specific therapeutic interventions. These so-called “predictive biomarkers” usually represent somatic changes that have
25 emerged during the process of carcinogenesis and can be detected in cancer or healthy tissue, in secretions, and circulating in blood. Predictive biomarkers can also be inherited germ-line variations that can also predict differential uptake, distribution, metabolism, and thus response to a drug, such as is possible for certain chemotherapeutic agents.

[00118] While the discovery of potential predictive biomarkers continues at
30 pace, very few have actually made it to a point of endorsement for routine clinical

use. This is primarily due to the fact that the transition from the discovery of a potential predictive biomarker to one with endorsed clinical utility is long, expensive and holds much in common with the requirements of drug development. As a result, there exists a large “portfolio” of poorly validated biomarkers, ranging from those that might be very helpful in personalizing cancer care (but whose clinical use is currently limited), to those with unproven clinical utility that are being used to manage care, (sometimes to the detriment of patient outcomes).

[00119] Biomarker information may be captured from various sources, including published literature, drug manufacturer information, FDA adverse event reports, or other sources, such as bioassay databases (e. g. <http://www.cancerrxgene.org>) as well as from the patient or from records of past patients. Once the prevailing attributes of such information have been captured in a database, they can be used to assess the clinical actionability of any genomic variation identified in a specific patient case. Such assessment considers the clinical validation level of a biomarker (i.e. whether it has been observed clinically, pre-clinically or computationally inferred) in combination with other factors such as:

The availability of the associated treatment in the patient disease;

Whether the biomarker predicts response, resistance or risk;

Whether it was observed in the current patients disease or not; and

Genomic parameters that tell us about the reliability of the variant called and the degree to which it occurs within the disease.

[00120] Summed together, this scoring schema can in turn be used to uncover and prioritize the most reliable biomarker information present in a patient genome, which in turn can prioritize the next best treatment options. Such treatment options may include targeted therapy, chemotherapy, surgery, radiation, laser ablation, vaccination, biological therapy, Immunotherapy, Stem Cell Transplant, Transfusion, Transplantation (e.g. bone marrow), Hyperthermia, Photodynamic therapy, nutritional adjustment (e.g. fasting), or physical exercise.

[00121] In brief overview, in some implementations, one or more steps are performed to convert tumor-sequencing data or data about other indications into clinically actionable information, including:

- 5 Alignment of the sequence reads coming from tumor and germline to reference genome;
- Identifying genomic variants associated with germline and tumor independently;
- Assessing the difference between germline and tumor variations to determine tumor specific variants;
- Mapping these variants to the proteome to identify coding variants;
- 10 Comparison of the variants against a drug response database (DRDB) to ascertain if they may be previously described predictive biomarkers. This process allows users to assess how current biomedical knowledge surrounding predictive biomarkers relates to the mutations identified in the patient tumor;
- Assessing the functional effects of these variations by applying a functional impact methodology. This “functional impact scoring” evaluates and predicts the
15 functional effects of genetic variations at the single protein- and molecular network level;
- Aggregating, integrating and collating the complete and up-to-date canon of relevant biomedical knowledge on protein function and biological context, disease
20 mechanisms and drug mode of action in an indication specific manner; and
- Utilizing this combined information for the prioritization of drugs or other treatments.

[00122] These approaches allow in-depth analysis and clinical interpretation of cancer genomes, supporting physicians in the demanding task of cancer drug
25 prioritization for their patients on the basis of a genomic tumor profile. It is imaginable that via genome-wide detection and prioritization of tumor-specific gene sequence variants from single patient case samples, treatment prioritization may be based on integrating all indication-relevant information with confidence or knowledge scores, which can be applied to proteins/genes. This method allows

directly assessing the treatment-relevance of all patient variants in protein coding genes.

Detection of tumor-specific genomic variants

[00123] In some implementations, the system allows for a fully automated workflow/pipeline for the detection of tumor-specific (somatic) non-synonymous single nucleotide variants (SNVs) in tumor-normal paired exome sequencing data sets. Variant detection occurs in a two-step process: (1) *sequence alignment* and (2) *variant calling*. The first step involves the global optimal alignment of sequence reads to the most current assembly of the human genome. Sequence reads can be de-duplicated prior to this step in some implementations.

[00124] In the second step the alignments from the tumor-normal paired sequence data set are used to call genomic sequence variants. Based on pre-set cut-off values for variant calling metrics, the set of detected tumor-specific genomic variants may be further processed for prioritization via 'functional impact scoring' or scoring of importance of a variant to an indication or tumor, discussed in more detail below.

Prioritizing variants according to variant call properties

[00125] In the course of detection, variants are annotated with technical parameters that reflect the quality of the genomic call such as allele frequency and genome probability. These parameters can be used complementary to functional impact scoring or importance scoring to prioritize variants. This also includes the classification of variants into 'missense' (causing an amino acid exchange) and 'nonsense' (introducing a premature STOP codon).

Relating genomic variants to reference genes/ proteins

[00126] In some embodiments, the system may map all detected genomic variants unambiguously to reference proteins. This allows the prioritization of genomic variants based on any protein-centric information, e.g. the collective cancer or indication-relevant attributes of the affected protein. In addition, the mapping supports the precise association of genomic variants with sequence-position-based structural-functional features and annotations of proteins. This

association may be used to determine the known or predicted impact of the precise mutation on the generic biological activity/function of the protein (referred to 'functional impact scoring').

Prioritizing cancer-relevant genes/ proteins in an indication specific manner.

5 **[00127]** In some implementations, the system collates clinically relevant information for the complete human proteome across various knowledge domains and, in some embodiments, is specifically tailored to oncology. In addition to capturing oncology-wide knowledge across all cancer types, for example, it may use specific indication information from the patient under analysis. The collated
10 information is used to compute a score for each protein, which directly reflects its importance for cancer in general and the cancer type under consideration. Similar steps may be applied for other indications or subtypes. The method enables the reliable prioritization of genomic mutations in proteins that are key to the particular cancer or indication including drug targets, disease drivers, oncogenes, tumor
15 suppressors and other molecular entity types.

[00128] Importantly, this relevancy score rates the cancer or indication-relevance of a protein independent of the occurrence of a genomic variant in the protein in a concrete patient case. Thus, if a previously unknown genomic variant is detected in a tumor and cannot be assessed for its potential effect on the basis
20 of the molecular nature of the exact mutation (by 'functional impact scoring', discussed in more detail below), it can still be rated based on the relevancy information.

[00129] In many implementations, the relevancy information is associated with a protein or gene and thus, if used for variant prioritization, may be equally
25 applicable or transferred to all variants of the respective protein or gene.

Functional impact scoring

[00130] Functional impact scoring (FIS) serves to prioritize variants based on their effect on protein function. The FIS of a variant is determined independently of the cancer or indication relevancy of the associated protein or gene. In contrast to
30 the relevancy score, which may be globally assigned to a protein/gene in some

embodiments, the functional impact score may be uniquely associated with the specific variant of a protein or gene (including position and exchange). Distinct variants of the same protein or gene may have different functional impact scores, while in many implementations, a protein or gene may have just a single relevancy score.

[00131] The functional impact score allows measurement of the effect of the detected genomic variant on the function or activity of the associated protein. Complementary to the relevancy score, which allows prioritizing variants in cancer or indication-relevant proteins or genes (irrespective of the effect of a specific genomic variant), the functional impact score allows identifying and prioritizing variants that have a high likelihood of creating net effects on the biological function of proteins. The scoring method is based on categorizing variants and rating the type and amount of supporting evidence in the respective category.

Process for building the DRDB

[00132] In one implementation, each variant detected in a patient is first compared to a drug response database (DRDB) that categorizes biomarker data. In some implementations, the comparison may be restricted to matching non-synonymous SNVs (missense and nonsense mutations) detected in patient tumor exomes or similar data. The DRDB may be built via text data mining and curation.

[00133] The raw DRDB data may be provided via text data mining from various publications, including medical journals, drug manufacturer data sheets, clinical study results, or any other such data. This data may be curated via a multi-step process to ensure high-quality data. Curation may be performed by a plurality of experts, who may perform cross reviews of the data for clinical-approved (e.g. endorsed) or clinical biomarkers according to curation guidelines.

Description of information captured in the DRDB

[00134] In some implementations, the drug response database may capture some or all of the following information, which may be used in defining an evidence level that can be attributed to a particular biomarker:

The Variant – type of aberration;

The drug or treatment used;

The observation context (e.g. patients (disease, disease stage) or model system);

Effect on drug or treatment responsiveness;

Type of response;

5 Quantity of effect; and

Validation level.

The drug response database may also capture the information as links to other databases comprising parts of information.

[00135] Specifically, the database may include information about any form of
10 biomarker, among other about any form genomic aberration including single
nucleotide polymorphisms (SNV's), copy number variations (CNV's), fusion
proteins (FP's), insertions and deletions (InsDels). Each variant may also be
identified by a role, such as primary for variants in cell lines or for primary
mutations in patients and secondary for secondary mutations. The lineage of the
15 mutation may also captured, for example, whether it is a germline or somatic
mutation.

[00136] Similarly, the database may include information about the drug or
treatment associated with the biomarker (i.e. variant) being reported, as well as
information about the context in which the biomarker observation was made - for
20 example, in model systems or patients. This information may include MeSH terms
or other hierarchical classifications.

[00137] In some implementations, the database may also include indication
specific information, such as tumor stage, the extent of the cancer, size of tumor,
or presence of metastasis. For most solid tumors, for example, there are two
25 related cancer staging systems, the Overall Stage Grouping, and the TNM system,
and their classification may be included in the DRDB. Tumor stage (I-IV) may also
be included, as well as site of metastasis. Other information may include *in vitro*
model information, such as cell line identification.

[00138] Information about drug or treatment response may also be included in the DRDB, including whether the variant confers increased responsiveness to treatment, resistance to treatment, or a risk of adverse events when a treatment is applied, as well as degree of sensitivity or quantity of the effect or clinical
5 response.

[00139] In some implementations, the DRDB may also include information about validation levels, including clinically approved or endorsed variants; clinically observed variants including in prospective, retrospective, or other studies; or pre-clinically reported variants observed in *in vitro* systems.

10 **[00140]** Now referring to FIG. 2, illustrated is a block diagram of a system for disease knowledge modeling and clinical treatment discussion support. In brief overview, the Clinical Decision Support Device (CDSD) 120 mines data from knowledge sources and then recommends and prioritizes treatment options based on patient characteristics and available knowledge about the disease. In some
15 implementations, the CDSD's data mining module 210 retrieves data from at least one data source 280. The mined data is storage in one of a plurality of mined databases 220. The databases may include a genomic database 221, a disease database 222, a literature database 223, a drug response database 224, and a treatment response database 225. In some implementations, the CDSD 120
20 includes a graphical user interface (GUI) 230 that allows a clinical user 281 to input patient 282 and past patient 283 data into a patient database 240 and past patient database 250, respectively. Responsive to a request for clinical decision support, the analysis module 260 retrieves data from the plurality of databases to calculate possible treatment options. The prioritization module 261 then prioritizes
25 the treatment options based on patient specific data and data within the plurality of databases.

[00141] Still referring to FIG. 2, and in greater detail, in some implementations, the CDSD 120 may be a computing device of any type, such as a desktop computer, portable computer, smart phone, tablet computer, or any
30 other type of computing device. In some embodiments, the plurality of databases may be housed in a second computing device, such as a data server.

[00142] The CDSD 120 may include a GUI 230. The GUI may allow a clinical user 281 to input patient data, past patient data, mine data, request clinical suggestions, or any combination thereof. For example, the clinical user 281 may be a doctor that inputs a patient's data into the system. The doctor may then request possible treatment options based on the type of disease the patient has and the knowledge gathered from the document sources 280 by the data mining module.

[00143] In some embodiments, the CDSD 120 may include a data mining module 210 that mines at least one document source 280 for data. Data mining module 210 may comprise an application, service, server, daemon, routine, or other executable logic for scanning and extracting information from data sources. The document source 280 may be a repository of scientific journals, a database such as PubMed, or other such source of scientific literature. The data mining module 210 may employ computational linguistic to extract text data from the document source 280. In some implementations, the data mining module seeks to find links between genomic variants, biomarkers, diseases, and drugs. The data gathered from the mining of the document source 280 is stored in plurality of mined databases 220. In some implementations, the data mining occurs on a second device and the second device provides the CDSD 120 with mined databases 220, or in other implementations, the CDSD 120 accesses the plurality of mined databases 220 that are stored on the second device. For example, the CDSD 120 may be a client type computing device. A server may continually mine new document sources 280 and provide the client CDSD 120 with updated mined databases 220 at regular intervals. In some implementations, biomarkers may be Genomic (e.g. single nucleotide variations (SNV), Copy number variations (CNV), insertions, deletions, gene fusions, polyploidy, gene expression, miRNA's), Proteomic (e.g. post-translational modifications, expression), metabolites, electrolytes, physiological parameters (e.g. blood pressure), patient age, patient weight, patient co-morbidities. In some implementations, the term drug may be used interchangeably with the term treatment, as the systems and methods discussed herein may be readily applied to non-drug based treatments. In some of these implementations, treatments may include: targeted therapy, chemotherapy, surgery, radiation, laser ablation, vaccination, biological therapy, Immunotherapy,

Stem Cell Transplant, Transfusion, Transplantation (e.g. bone marrow), Hyperthermia, Photodynamic therapy, nutritional adjustment (e.g. fasting), physical exercise.

[00144] The CDSD 120 may include a past patient database 250. In some implementations, the past patient database 250 is a record of the current and/or past patients input by the clinical user. The record may include the disease, variant, and treatments of past patients. In some implementations, this information may supplement the mined databases 220 when clinical decisions are made. In other implementations, the past patient database 250 includes data from a plurality of CDSDs 120. For example, a CDSD 120 may save anonymized patient data to a central database. The anonymized data may be accessed by a network of CDSDs 120 when providing clinical support.

[00145] The genomic database 221 may store genomic information mined from the document sources 280. In some implementations, the genomic information may include a list of genetic variants or mutations, full or partial genetic sequences, or any such similar information. The genomic information may be associated with one or more diseases, conditions, or indications stored in a list in the disease database 222. In some embodiments, the variants are categorized as having a risk, response, or resistance when associated with a drug or treatment listed in the drug or treatment response database 224, said information being stored in the genomic database 221 and linked to the treatment response database 224, or stored in the treatment response database 224. For example, if a patient presents with a specific type of cancer and a specific genetic variant, the variant may make it such that the drug has no effect on the cancer. Similarly, some genetic variants may place the patient at high risk when consuming a particular drug compared to the general population.

[00146] In some implementations, a plurality of validations is associated with the information stored in the genomic database. The mined variant data may be clinically validated and/or have a validation context. In some implementations, the clinical validation includes a tier of validation. The tiers may include or represent that the biomarker is “endorsed by key opinion leader”, “clinically observed”, “pre-clinically observed”, or “inferred”. Key opinion leader validation may come from

sources such as the Food and Drug Administration (FDA), American Society of Clinical Oncology (ASCO), or other such organizations. This type of validation may occur when the key opinion leader has issued a report or endorsement of a correlation of a biomarker with an indication or treatment as causing a specified response or outcome. Clinically observed validation may occur when variant-disease-drug response links have been seen clinically, and the findings have been published in peer-reviewed journals or in conference abstracts, for example. Pre-clinical validation may occur when variant-disease-drug response links have been observed in pre-clinical models, such as animal models, of the disease. Inferred validation may occur when variant-disease-drug response links have been observed in computer models of the disease, or based on the similarity of a novel biomarker to a known predictive biomarker (e.g. BRAF V600D might be inferred to have a similar predictive effect to BRAF V600E). Additionally, the context of validation and its relationship to the current patient's disease may be assessed. For example, the context may be that the link was observed in the same disease as the patient currently under treatment. A second context may be that the link was observed in another disease similar to the disease of the patient. In some implementations, rankings are associated with each of the above described validation tiers and contexts. In other embodiments, other tiers may be utilized, such as "guideline" for treatments that have been published as a standard treatment guideline for an indication. In some implementations, if a standard treatment guideline exists, such treatments may be prioritized over others as a default rule.

[00147] Discussed further in relation to FIG. 4, but briefly, in some embodiments, the analysis module 260 may be utilized for generating personalized drug efficacy or risk information or identifying potential drug interactions. Analysis module 260 may comprise an application, service, server, daemon, routine, or other executable logic for analyzing biomarker and patient information, generating or aggregating a score for one or more proposed treatments based on patient-specific and biomarker information, and generating an ordered list of prioritized treatments for a patient. In some implementations, the analysis module 260 determines whether a variant or other biomarker discovered in a patient should be applied in the clinical support process. The decision to apply

the biomarker may be made by applying a plurality of predefined criteria. For example, the analysis module 260 may apply one, or any combination, of the following criteria: degree to which the variant has been clinically validated to affect the drug effect; whether clinical validation occurred in the patient disease or some other disease; availability and/or approval of medication in patient indication; 5 reliability of the variant call in the patient measurement; percentage of reads in which variant is detected in the patient sample; what type effect is the biomarker associated with (e.g. response, resistance, or risk); and measure of how strongly the a drug effect is altered on average in comparison to patients without the biomarker. In some embodiments, as shown, analysis module 260 may comprise or execute a prioritization module 261, which may comprise an application, 10 service, server, daemon, routine, or other executable logic for scoring and/or prioritizing scored treatments, as discussed above.

[00148] Now referring to FIG. 3A, illustrated is an exemplary embodiment of GUI 230 for the CDSD 120. The GUI 230 may include a number page views 300. 15 The page view 300 may include a summary section 301. In some implementations, as indicated by boxes below the summary section 301 in the example screenshot, prioritized treatment options (discussed further in relation to FIG. 4) may be displayed. The treatments may be grouped as a potentially effective therapy 302, a potentially ineffective therapy 303, and/or a potentially 20 toxic therapy 304.

[00149] Still referring to FIG. 3A, and in greater detail, in some implementations, a clinical user may be presented with a summary section 301. The summary section 301 may provide the clinical user with an overview of a 25 patient's medical record. In some implementations, the summary section 301 may list specific variants found in the patient's cells or other medical information relating to genomic information. In some implementations, the summary section links directly to third party digital medical record systems, such that the information from the disclosed system may be viewed in the patient's clinical, digital chart. In 30 many implementations, summary 301 may be generated dynamically by the analysis engine responsive to results of analysis and prioritization of treatments, patient history analysis, or other information. For example, text strings with

variables may be pre-written and may be dynamically selected and added to the summary.

[00150] As discussed above, the treatment options may be tiered as responsive, resistant, or risk. In some implementations, the responsive, resistant, and risk tiers are mapped to the GUI sections potentially effective therapy 302, potentially ineffective therapy 303, and potentially toxic therapy 304, respectively.

[00151] In some implementation, the therapy sections provide information relating to biomarker facts, validity, drug name, approved for indication, drug interactions, gene symbol, variant symbol, disease context, or any combination thereof. For example, under the potentially ineffective therapies 30 section for a particular patient, two drugs are listed Sorafenib and Nilotinib. For each drug, under validity, an icon indicates at what stage the drug has been validated. For example, the microscope icon may indicate the drug has only been validated in pre-clinical studies. Also as shown in the exemplary embodiment of page view 300, it is indicated that Sorafenib has been indicated in 8 clinical trials for this particular indication. The CDS 120 may determine this information via the data mining module's 210 analysis of the document sources 280. In some implementations, a plurality of icons may be used to quickly and effectively relay information to the clinical user. The icons may be selected to reduce confusion and aid in the ease of determining which treatment option may be the most effective for a particular patient.

[00152] For example, FIG 3B, illustrates one set of icons that may provide the user with additional information, according to one exemplary embodiment. FIG. 3B illustrates a set of icons that may provide a clinical user with information regarding the validity of variants. As illustrated in FIG. 3B, icons 354 associated with risk may be colored red, icons 355 associated with response may be colored green, and icons 356 associated with resistance may be colored grey. Additionally, in some implementation, the above described clinical validity and validation context may be represented in the icon. For example, the icon group 350 illustrates a figure fully colored. This may indicate that the clinical validity of the response type has been endorsed by a key opinion leader such as the FDA. In some implementations, each of the icon sets may include an asterisk, or other indicator,

if the validation context was the exact validation context of the patient. For example, a fully red colored figure with an asterisk, may indicate that, based on a fixed drug treatment, the variant indicates a risk to the patient, the variant has been validated in the patient's exact disease, and the validation has been
5 endorsed by an organization such as the FDA. The various figures and icons thus may be used to quickly communicate likely result of a treatment, as well as statistical or inferred confidence in the result.

[00153] In some implementations, there may be an icon for each type of clinical validity. For example, variants that have been validated through clinical
10 observation may be represented as an icon figure half colored 351; variants that have been validated through pre-clinical models may be represented as an icon of cells in a Petri dish 352; and variants that have been inferred through models may be represented as genetic map icon. Though described above as having a specific color, icon or indication, the above examples were provides only as an exemplary
15 embodiment. One of ordinarily skilled in the art will recognize and appreciate the various ways the above icons may be colored, indicated, or otherwise represented.

[00154] FIG. 4A is a flow chart illustrating a method 400 for delivering clinical decision support according to one exemplary embodiment. In general, the analysis
20 module retrieves an identification of an indication of a patient and the status of a biomarker in a patient (step 401). The analysis module then identifies a plurality of treatments associated with the biomarker or indication (step 402). Responsive to identifying the treatments, the analysis module generates a score for each of identified treatments (step 403). Then the possible treatment options are prioritized
25 and displayed to a user (step 404).

[00155] FIG. 4B is another flow chart illustrating a method 500 for delivering clinical decision support according to one exemplary embodiment. In general, the analysis module retrieves an identification of a biomarker in a patient (step 501). The analysis module then identifies a plurality of treatments associated with the
30 biomarker (step 502). Responsive to identifying the treatments, the analysis module generates a score for each of identified treatments (step 503). Then the possible treatment options are prioritized and displayed to a user (step 504).

[00156] At steps 401 or 501, the analysis module 260 retrieves patient data and biomarker data from at least one database. The patient data may indicate if the patient has or is suspected of having a specific indication (*i.e.*, disease). In some implementations, the patient data is extracted from the patient database 240
5 and converted into a form capable of being interpreted by the analysis module 260. For example, a patient may be represented as a variable or structure P having specific characteristics. The characteristics of P may a disease (or indication) I , one or more variants V_0 , and the variants may have a reliability and relative abundance in the patient's cells. Accordingly, the patient may be
10 represented as:

$$P = (I_0, V_0, \text{reliability}(V_0), \text{percentage}(V_0)).$$

[00157] Although referred to as variants V_0 , in many implementations, characteristics of other biomarkers may be utilized. Furthermore, in many implementations, the absence of a biomarker may be characterized and utilized for
15 analysis. For example, the absence of a particular protein in a patient that typically is found in other patients with the same indication may be significant, may indicate an underlying genetic or physiological difference in the patient, and may be correlated with differences in treatment outcomes that may be specific to the patient or other patients having the mutation or variation.

[00158] Similarly, the analysis module 260 may retrieve data from the mined databases 220. For example, the analysis module 260 may retrieve data characteristics of each of the biomarkers associated with an indication and/or patient. In some implementations, the data characterization of each biomarker may include at least one of: drugs used in treatment or other treatment methods
25 (D), type of effect (T) (e.g., response, resistance, risk), expected response based on the knowledge of biomarker (S), evidence level (L), such that the analysis module 260 may characterize the biomarker as:

$$B_i = (I_i, V_i, D_i, T_i, S_i, L_i).$$

[00159] In some implementations, T is equal to 1 when associated with a response and -1 when associated with a resistance. In further implementations, T
30 may be set between 0 and -1 when associated with a risk (e.g. -0.2, -0.4, -0.6, or -

0.8, or any other such value). This may allow for characterization of the severity of the risk (e.g. -0.2 for an annoying, but non-life threatening risk, and -0.8 for a potentially severe risk). In other implementations, T may be set to a greater negative value to indicate a risk, such that the system may distinguish between risk and resistance. For example, T may be set to -1 for a resistance, and may be set to a value greater than -1, such as -2, -3, -5, or any other such value to represent a risk of varying severity. In still other implementations, T may be set to a value between 0 and -1 for a risk of a minor side effect and to a value less than -1, such as -3 or -5, for a risk of a severe side effect. This may be useful in instances where a disease is life-threatening, and a minor side effect may be acceptable if there is sufficient benefit or response from the treatment.

[00160] At steps 402 or 502, the analysis module identifies a plurality of treatments associated with the indication and/or biomarker retrieved in step 401 resp. 501. The associated treatments may be stored in a treatment information database 225. In some implementations, the treatment information database 225 is one of the mined databases 220, such that the treatments have been mined, gathered and organized by the data mining module 210 from the document sources 280. For example, in some implementations, a treatment may be considered associated with an indication and/or biomarker if it is found together with said indication and/or biomarker in literature. Various thresholds may be applied, including distribution or distance between references to the indication or biomarker and treatment, number or percentage of references to the indication or biomarker and treatment within an item of literature, frequency of appearance of the combination, number of citations to papers that include the combination, or other such thresholds or rules based off such information.

[00161] At steps 403 or 503, the analysis module 260 scores each of the retrieved treatment options. Generally, in some implementations, the score is based on at least one of: a clinical validation level of the biomarker; the biomarker's association with a response to the treatment, resistance to the treatment, or the risk of adverse effects from the treatment; and the reliability of the detection of the biomarker.

[00162] In some implementations, the above scoring process determines the applicability of at least one biomarker to the patient. In some implementations, the applicability of a biomarker is determined by the patient specific data and drug treatments available to the patient. For example, the applicability A of a biomarker B_i to a fixed treatment D_k for a patient P may be represented as:

$$A(P, D_k, B_i).$$

[00163] In some implementations, the calculation of the applicability A may involve the evidence level L_i , the *similarity*(I_0, I_i) between the patient's indication I_0 and the indication I_i in which the biomarker is associated in the genomic database 221, the *availability*(I_0, D_i) of a drug or treatment D_i for the indication I_0 , the *reliability*(B_i) of the biomarker, the reliability of the variant V_0 , the percentage of patient cells with variant V_0 , the similarity between the patient variant V_0 and the variant V_i associated with the biomarker B_i or whether such variants are identical, and the similarity between the drug or treatment D_k and the drug or treatment D_i associated with the biomarker B_i or whether such drugs or treatments are identical. Accordingly, in some implementations, the applicability of a biomarker is represented as:

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{reliability}(B_i);$$

or as

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{similarity}(I_0, I_i) * \text{reliability}(B_i);$$

or as

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{similarity}(I_0, I_i) * \text{availability}(I_0, D_i) * \text{reliability}(B_i);$$

or as

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{availability}(I_0, D_i) * \text{reliability}(B_i);$$

or as

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{similarity}(I_0, I_i) * \text{availability}(I_0, D_i) * \text{reliability}(V_0) * \text{percentage}(V_0);$$

or as

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{similarity}(I_0, I_i) * \text{availability}(I_0, D_i) * \text{reliability}(V_0) * \\ \text{percentage}(V_0) * \text{identical}(V_0, V_i);$$

or as

$$5 \quad A(P, D_k, B_i) = \text{validity}(L_i) * \text{similarity}(I_0, I_i) * \text{availability}(I_0, D_i) * \text{reliability}(V_0) * \\ \text{percentage}(V_0) * \text{identical}(D_k, D_i);$$

or as

$$A(P, D_k, B_i) = \text{validity}(L_i) * \text{similarity}(I_0, I_i) * \text{availability}(I_0, D_i) * \text{reliability}(V_0) * \\ \text{percentage}(V_0) * \text{identical}(V_0, V_i) * \text{identical}(D_k, D_i).$$

- 10 **[00164]** In some implementations, the above-described variables are normalized between 0 and 1 to indicate their effect on biomarker applicability. For instance, the evidence level L may be mapped from 1 to 0, such that when the variant is "KOL endorsed" $\text{validity}(L_i)=1$, and if the variant is "inferred" $\text{validity}(L_i)=0.2$. This may imply that an inferred biomarker is considered only 20%
- 15 of the amount of evidence of an endorsed one, possibly to reflect that there is only about 20% chance that the biomarker would ever be fully confirmed and endorsed. In a further implementation, the relevance of an inferred biomarker or $\text{validity}(L_i)$ may be set to a predetermined value responsive to whether standard treatment guidelines are available for the patient. For example, as discussed above, if such
- 20 standard treatment guidelines are available, then the relevance of an inferred biomarker $\text{validity}(L_i)$ may be set equal to 0 or a similarly small value, such as 0.01. In other implementations, the standard treatment may be prioritized over other treatments as a default rule, regardless of prioritization of other treatments. If guidelines are not available and/or no other biomarkers are found, then the
- 25 relevance of the inferred biomarker $\text{validity}(L_i)$ may be increased to a predetermined level, such as 0.2, 0.3, 0.5, or any other such value. In some implementations, $\text{validity}(L_i)$ is set to 1 when "KOL endorsed"; between 1 and 0.5 when clinically validated, and less than or equal to 0.2 when pre-clinically validated. Similarly, for $\text{similarity}(I_0, I_i)$ a value of 1 may be used to encode that
- 30 two entities are identical and/or that the biomarker has been validated in an

indication of the patient; values between 0 and 1 could correspond to the likelihood of the analogous biomarker to be valid for a related indication I_0 instead of I_i , and a value of 0 would indicate that no transfer should be made between the unrelated indications (in the context of the given variant, etc). The similarity of I_0 , I_i and/or V_0 , V_i may be include molecular level similarities such as homology or expression of key proteins and/or the structure of the exchanged amino acids. Similarity values may not be a dichotomy; rather, values between 0 and 1 may be used, such as 0.2, 0.4, 0.6, 0.8, or any other such value for a related indication, and values of 0, 0.01, 0.02, 0.05, 0.1, or any other such value for an unrelated indication. Different values may be used responsive to other shared characteristics between the indications, such as whether they involve the same pathway, same organ, or other such characteristics. Likewise, similar or identical drugs or treatments D_k , D_i may be encoded with values of 1 to indicate identical treatments, or values between 0 and 1 to indicate similar but non-identical treatments (e.g. conventional external beam radiation therapy, stereotactic radiosurgery, intensity-modulated radiation therapy, etc.). In some implementations, *reliability*(B_i) is the average of a value for the reliability of the detection method of the biomarker and a value for the frequency for the detection of the biomarker in the patient. In other implementations, some, or all, of the above of the above characteristics may be combined in different ways. For example, some terms may be correlated and those terms may be averaged, e. g. geometrically or arithmetically, before determining the applicability A . In yet other implementations, the above characteristics may be summed to return an applicability A .

[00165] In many implementations, applicability A may be determined for each of a plurality of biomarkers associated with the patient, including biomarkers detected in the patient or expected biomarkers that are not detected in the patient. These thereby determined multiple applicability scores- may be aggregated over all applicable biomarkers to generate a score for a specific treatment for the patient, based on their specific physiology and/or genotype. In some implementations, the effects are grouped by the above-described type T (e.g., response, resistance, risk) for each drug or treatment, each group being provided with a separate score. In other implementations, the groupings may be summed with a weighting factor, wherein the weighting factor may be the effect size S_i . The

product of the applicability A with the effect size S_i may be referred to as a sub-score. The effect size may comprise a measurement of a likelihood of response or resistance of a treatment in patients having the indication and biomarker, or a measurement of a hazard ratio of adverse events experienced by patients having the indication and biomarker undergoing the treatment. Such measurements may be actual data values, percentages, proportional reporting ratios, log ratios, or other values or types of values. Multiplying by the effect size may be interpreted as the estimated probability of the effect of biomarker B_i to be seen in the given patient when treated with drug or treatment D_k . Accordingly, the summed effect may be represented as:

$$S_p(P, D_k, T_1) = \sum_{Ti=T1} A(P, D_k, B_i) S_i$$

[00166] In some implementations, if no biomarkers are applicable, drugs may still be prioritized based on their expected effects. This effect may be classified as the base effect S_0 of the drug or treatment on a patient with no variants. Therefore, the total effect for a patient may be:

$$S_T = S_0 + S_P$$

[00167] In some implementations, a single treatment score t is determined for each treatment based on the effect of the drug or treatment over each of the above-described types. For example:

$$t(P, D_k) = f[S(P, D_k, \text{Response}), S(P, D_k, \text{Resistance}), S(P, D_k, \text{Risk})]$$

[00168] The function f may be a function that maps the three effects to a single treatment score t . In some implementations, the weights given to each of the S terms depends on the severity of the indication and/or individual preferences regarding the trade-off between risk and benefit for taking a specific drug. In some implementations, S_{RESPONSE} and $S_{\text{RESISTANCE}}$ are inversely correlated, such that if a patient is more responsive, the patient must be, by definition, less resistant to the drug. Therefore, in some implementations, S_{RESPONSE} and $S_{\text{RESISTANCE}}$ are combined into a single S term. In some implementations, the treatments are then ranked based on the treatment score t .

[00169] At steps 404 or 504, a least a portion of the identified treatments are prioritized response to the determined scores. In some implementations, treatments with lower scores may be ordered lower than treatments with higher scores. The treatments may also be organized by their above described effect type *T*. For example, the treatments that the analysis module 260 determines the patient should respond to may be grouped separate from the treatments to which the patient may be resistant.

[00170] While the invention is particularly shown and described with 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 without departing from the spirit and scope of the invention described in this disclosure.

EXAMPLES

[00171] Example 1

[00172] In order to illustrate the method, its application to a first example case is presented in table 1 as depicted in Figure 5 and in table 2 at the end of this section.

[00173] It is assumed that the patient's primary indication is breast cancer, and that the following variant is found in her tumor:

PIK3CA.H1047R (i.e. in protein encoded by the gene PIK3CA, the 1047th amino acid is mutated from an H to an R)

[00174] Table 1 shows the applicable predictive rules that are found in a version of a drug response database (DRDB). Each of these predictive rules corresponds to one or several observations of an outcome after a treatment, applied either to a human or to a model system like a cell line. The following columns describe the predictive rules:

pmid: The PubMed ID of a publication describing the observation / predictive rule.

variants: The relevant variants reported for the patient(s) or model system(s); usually, they are primary, the single case of a secondary variant is explicitly marked as such.

BM_variants: describes to what extent the variants of the predictive rule are found in the patient.

treatment: The drug or drug combination applied to patient(s) or model system(s) [no other treatments are considered in this example].

- 5 resp_type: Either "resistance" (to indicate the lack of the intended outcome of the treatment); or either "sensitivity" (in the case of a model system) or "response" (in the case of a patient) to indicate (part of) the intended outcome.

sub_score: The sub-score computed for the predictive rule w.r.t. the given patient, which is the value assigned resp_type. The assignment was done as follows:

- 10 resistance: -1
 response: +1
 sensitivity: +0.5

- [00175] Table 2 lists all treatments that appear in at least one predictive rule (i.e. in any line of Table 1) in the column "treatment", together with the number of
15 predictive rules that are associated with it in the column "n_drdb". Higher numbers of DRDB entries may hint towards higher reliability of the corresponding scores.

[00176] The column "score" shows the scores of the treatments, which are computed by summing the sub-scores of all predictive rules associated with the corresponding treatment.

- 20 [00177] The column "rank" shows the ranks of the treatments as a numerical indication of a prioritization. There are several ties due to identical scores for two or more treatments, resulting in an ordering that is only partial.

[00178] Table 2:

treatment	n_drdb	score	rank
NVP-BEZ235	3	1.5	1.5
gabexate mesilate	3	1.5	1.5
Temsirolimus	2	1.0	3
BGT226	1	0.5	9.5
GDC-0980	4	0.5	9.5
Perifosine	1	0.5	9.5
MK2206	1	0.5	9.5

AZD6244 // NVP-BKM120	1	0.5	9.5
Everolimus // RAF265	1	0.5	9.5
Irinotecan // NVP-BEZ235	1	0.5	9.5
MK2206 // Vemurafenib (PLX-4032)	1	0.5	9.5
Refametinib	1	0.5	9.5
Refametinib // Temsirolimus	1	0.5	9.5
AKTI-1/2 (Inhibitor VIII) // Vemurafenib (PLX-4032)	1	0.5	9.5
RAF265	1	0.5	9.5
NVP-BKM120	3	0.0	16.5
PP242	3	0.0	16.5
Everolimus	7	-0.5	19
Trametinib	5	-0.5	19
AZD6244	5	-0.5	19
Saracatinib (AZD0530)	4	-1.0	23
Paclitaxel	1	-1.0	23
Dacomitinib (PF00299804)	1	-1.0	23
Lapatinib	1	-1.0	23
Panitumumab	1	-1.0	23
Vemurafenib (PLX-4032)	2	-2.0	26.5
Trastuzumab	2	-2.0	26.5
Cetuximab	12	-8.5	28

[00179] Example 2

[00180] In order to illustrate the method, its application to a second example case is presented in table 3 as depicted in Figure 6 and in table 4 at the end of this section.

[00181] It is assumed that the patient's primary indication is colorectal cancer, and that the following variants are found in his/her tumor:

gPIK3CA.E545K (i.e. in protein encoded by the gene PIK3CA, the 545th amino acid is mutated from an E to a K)

10 gKRAS.G12D (analogously)

[00182] Table 3 shows the applicable predictive rules that are found in a version of a drug response database (DRDB). Each of these predictive rules corresponds to one or several observations about response to a treatment. The following columns describe the predictive rules:

15 treatment: the drug or drug combination applied to patient(s) or model system(s)
[no other treatments are considered in this example]

variants: the relevant variants reported for the patient(s) or model system(s); usually, they are primary, the single case of a secondary variant is explicitly marked as such

5 resp_type: either "resistance", or "sensitivity" (in the case of a model system) or "response" (in patient(s))

resp_quant: "strong", "medium", or "weak"

np: the number of patients in which the reported outcome was observed

indications: the primary indications reported for the patient(s) or model system(s)

10 validation: either "endorsed" or "clinical" for patients, or "pre-clinical human" or "pre-clinical non-human" for model systems ("pre-clinical NA" if the model species is unknown)

[00183] In this example, the effect size is represented in two columns, namely "np" (which is implicitly taken to equal 1 for model systems) and "resp_quant".

15 **[00184]** Each of the rows is augmented by further columns that provide information on how the predictive rule matches to the given patient case:

BM_variants: describes to what extent the variants of the predictive rule are found in the patient

w_variants: a value for how well the variants match, here:

20 "all primary SNVs found" -> 1

"any primary SNV found" -> 0.8

"any secondary SNV found" -> 0.5

"no biomarker SNV found" -> 0

25 w_indication: a value for how well the indications match (in case of more than one indication in patient and/or predictive rule, the pair yielding the highest value is chosen), here:

	1.0	in patient indication
	0.8	in generalization of patient indication
	0.7	in specialization of patient indication
	0.5	related to patient indication
5	0.3	in any cancer type - unrelated to patient indication
	0.3	in unknown indication - unclear relation to patient indication

w_validation: a value that represents the clinical validation level of the predictive rule as follows:

	endorsed	1.0
10	clinical	0.8
	pre-clinical human	0.5
	pre-clinical non-human	0.2
	pre-clinical NA	0.2

w_strength: a value representing the resp_quant in the following way:

15	strong ->	1
	medium ->	2/3
	weak ->	1/3

w_effect: a value that represents the effect size, here computed as the product of pat_k and w_strength.

20 **[00185]** subscore: the subscore of the predictive rule for the given patient is here computed as the product of w_variants, w_indication, w_validation, and w_effect. Note that this representation permits the computation of separate scores despite being a single column, as each row is assigned to a single type of outcome by the column resp_type.

[00186] Table 4 lists all treatments that appear in at least one predictive rule (line of Table 3) in the column "treatment", together with the number of predictive rules that are associated with it in the column "n_drdb". Higher numbers of DRDB entries would hint towards higher reliability of the corresponding (separate) scores.

5 **[00187]** Based on the predictive rules and their assigned (separate) sub-scores, separate scores for each of the distinct treatments can be computed. Table 4 shows these separate scores in the columns "respond" and "resist", where the latter represents both response types (column "resp_type" in Table 3) "sensitivity" and "response". The aggregation over predictive rules that refer to the
10 same treatment is performed by summation.

[00188] The column "diff" provides scores that represent the uniformly weighted sum of the separate score "respond" and the negated separate score "resist". This may serve as a score to prioritize the treatments; however it is strongly sensitive to the number of observations (DRDB entries), such that both
15 top and the bottom of the list tend to be dominated by strongly covered treatments.

[00189] One alternative is to score and prioritize treatments according to a response rate, which ideally approximates the likelihood of the patient to respond to the treatment. The column "raw_rate" provides the naively estimate of the response rate, computed as $\text{respond} / (\text{respond} + \text{resist})$. However this estimate
20 suffers from high variance for treatments with associated little data; in the example, this shows by the large number of treatments that have a value of either 0% or of 100% in this column.

[00190] An regularized response rate per treatment, in Table 4 shown as "reg_rate", is computed as a shrinkage estimate by adding a number r_{av} to the
25 separate response score, adding $(1 - r_{\text{av}})$ to the separate resist score, and forming the corresponding regularized response ratio. Here r_{av} is an average response rate over all treatments, and is computed as the fraction of the sum of all separate scores for response over the sum of all separate scores. The regularization introduces bias, but may reduce the variance even more, such that the resulting
30 estimate is more accurate than the naive one.

- [00191]** Sometimes a conservative treatment decision is desired, i.e. the risk of selecting a treatment that appears most promising although it is less beneficial than other treatments due to statistical randomness in a small data foundation shall be low. Then it may be indicated to use a quantile (typically below the median) of the estimated response rate as score upon which to base the prioritization of treatments. Therefore, the table is augmented by several measures of statistical uncertainty of the estimated response rate. Specifically, "q05_rate" and "q25_rate" are the 5%- and 25%-quantiles of the estimated posterior distribution of the regularized response rate.
- 5
- [00192]** Any of the columns "diff", "raw_rate", "reg_rate", "q05_rate", and "q25_rate" may serve as a score to prioritize the treatments. Obviously there are further reasonable choices, including but not limited to different regularization schemes and different quantiles.
- 10

[00193] Table 4:

15

treatment	n_d rdb	respo nd	resist	diff	raw_rat e	reg_rate	q05_rat e	q25_rat e
NVP-BEZ235	7	1.63	0.00	1.63	100.00%	70.85%	23.97%	54.85%
MK2206	2	0.72	0.00	0.72	100.00%	55.57%	5.71%	29.65%
Trametinib	12	1.99	1.01	0.97	66.22%	55.53%	17.61%	38.76%
GDC-0980	2	0.68	0.00	0.68	100.00%	54.52%	5.01%	28.01%
Fluorouracil // Gefitinib // Irinotecan // Leucovorin	1	0.53	0.00	0.53	100.00%	50.07%	2.70%	21.39%
Irinotecan // NVP- BEZ235	1	0.40	0.00	0.40	100.00%	45.31%	1.21%	15.03%
Saracatinib (AZD0530)	8	1.80	1.80	0.00	50.00%	44.23%	11.46%	27.82%
BGT226	1	0.35	0.00	0.35	100.00%	43.29%	0.82%	12.61%
AG490 // NVP- BKM120	1	0.23	0.00	0.23	100.00%	37.92%	0.23%	7.17%
Docetaxel // Perifosine	1	0.19	0.00	0.19	100.00%	35.48%	0.12%	5.23%

Letrozole	1	0.19	0.00	0.19	100.00%	35.48%	0.12%	5.23%
Perifosine	1	0.19	0.00	0.19	100.00%	35.48%	0.12%	5.23%
PP242	4	0.47	0.52	-0.05	47.46%	35.35%	1.06%	10.70%
NVP-BKM120	6	0.39	0.45	-0.06	46.25%	33.87%	0.66%	8.76%
17-AAG	1	0.15	0.00	0.15	100.00%	33.43%	0.06%	3.86%
Ganetespib	1	0.15	0.00	0.15	100.00%	33.43%	0.06%	3.86%
AG490	1	0.12	0.00	0.12	100.00%	31.44%	0.03%	2.76%
C124017	1	0.12	0.00	0.12	100.00%	31.44%	0.03%	2.76%
PI-103	1	0.12	0.00	0.12	100.00%	31.44%	0.03%	2.76%
ZSTK474	1	0.12	0.00	0.12	100.00%	31.44%	0.03%	2.76%
Temsirolimus	1	0.09	0.00	0.09	100.00%	29.98%	0.02%	2.10%
Carboplatin // Paclitaxel // Reolysin	1	0.08	0.00	0.08	100.00%	29.11%	0.01%	1.76%
Everolimus // RAF265	1	0.08	0.00	0.08	100.00%	29.11%	0.01%	1.76%
Sti-571	2	0.12	0.15	-0.03	44.44%	27.91%	0.02%	2.25%
Sirolimus	1	0.06	0.00	0.06	100.00%	27.77%	0.01%	1.31%
RAF265	1	0.04	0.00	0.04	100.00%	26.38%	0.00%	0.94%
Dacarbazine	3	0.80	2.40	-1.60	25.00%	24.63%	1.80%	9.22%
Matuzumab // Paclitaxel	2	0.08	0.24	-0.16	25.00%	23.82%	0.01%	1.21%
Paclitaxel	1	0.00	0.12	-0.12	0.00%	20.99%	0.00%	0.32%
Gefitinib	5	0.16	0.77	-0.61	17.13%	20.39%	0.03%	1.75%
Cetuximab // Irinotecan	19	3.15	14.08	-10.93	18.27%	18.55%	6.17%	11.93%
Lapatinib	1	0.00	0.28	-0.28	0.00%	18.31%	0.00%	0.25%
CL-387,785	1	0.00	0.35	-0.35	0.00%	17.36%	0.00%	0.23%
Dacomitinib (PF00299804)	1	0.00	0.35	-0.35	0.00%	17.36%	0.00%	0.23%
Vemurafenib (PLX- 4032)	1	0.00	0.40	-0.40	0.00%	16.74%	0.00%	0.22%
Cetuximab	19	2.88	16.45	-13.57	14.90%	15.32%	4.69%	9.48%
Everolimus // Gefitinib	2	0.00	0.72	-0.72	0.00%	13.63%	0.00%	0.16%
AZD6244	3	0.00	0.92	-0.92	0.00%	12.21%	0.00%	0.14%
Everolimus	12	0.60	5.58	-4.97	9.76%	11.67%	0.42%	3.11%

Panitumumab	17	0.80	11.56	-10.76	6.47%	7.74%	0.47%	2.46%
Trastuzumab	3	0.00	2.22	-2.22	0.00%	7.29%	0.00%	0.07%
Irinotecan // Panitumumab	4	0.00	3.20	-3.20	0.00%	5.58%	0.00%	0.05%

Claims

1. A method for prioritization of patient treatment options based on an analysis of biomarkers, comprising:
 - 5 retrieving the results of measurements of a set of one or more biomarkers of the patient,
 - identifying a plurality of treatments associated with any of the set of one or more biomarkers via identifying rules relating the treatments to biomarkers in a treatment information database, the rules corresponding to one or
 - 10 several observations of an outcome after a treatment, applied either to a human or to a model system like a cell line,
 - generating a score for each of the identified plurality of treatments, and
 - ordering the identified plurality of treatments according to the generated score to provide a treatment option or treatment contraindication
 - 15 prioritization for the patient,

characterized in that for each of the one or more biomarkers the score is computed by setting values for

 - a) whether rules are associated with
 - response to the treatment,
 - 20 -resistance to the treatment, and
 - risk of adverse effects from the treatment.
2. The method of claim 1, wherein generating a score for each of the identified plurality of treatments further comprises, for each generated score:
 - generating a sub-score for each biomarker, and
 - 25 aggregating the sub-scores to generate the score.
3. The method of any of the preceding claims, wherein generating the score for each of the identified plurality of treatments comprises generating two or more separate scores for each of the identified plurality of treatments, said separate scores corresponding to said response to the treatment, said

resistance to the treatment, and/or said risk of adverse effects from the treatment, and aggregating the two or more separate scores.

4. The method of claim 3, comprising specifying, by a user, treatment risk/benefit preferences and wherein scoring and ordering the identified plurality of treatments comprises weighting the two or more separate scores according to the treatment risk/benefit preferences.
5. The method of any of the preceding claims, wherein the score is further computed by setting values for
 - b) effect sizes of the rules for each of the one or more biomarkers,wherein the effect sizes of the rules for each of the one or more biomarkers comprise a measure of likelihood of response or resistance, a measure of a hazard ratio, a measure of likelihood odds of response or resistance, or a measure of a quotient of hazard ratios of the treatment.
6. The method of any of the preceding claims, further comprising retrieving an identification of a disease indication of the patient, and wherein the score is further computed by setting values for
 - c) whether the rules for each of the one or more biomarkers have been established with respect to the disease indication of the patient, related disease indications, or unrelated disease indications.
7. The method of any of the preceding claims, wherein the score is further computed by setting values for
 - d) clinical validation levels of the rules for each of the one or more biomarkers,in particular wherein the clinical validation level is one out of the following list: endorsed, clinical, pre-clinical, inferred.

8. The method of any of the preceding claims, wherein the rules relating the treatments to biomarkers are stored in one or more curated treatment information databases.
- 5 9. The method of claim 8, wherein the score is further computed by setting values for
- e) a reliability of the involved document sources of the rules.
10. The method of any of the preceding claims, wherein the score is further computed by setting values for
- 10 f) a reliability of measurement of the biomarkers,
- in particular wherein a value for the reliability of measurement of the biomarkers is composed of a value for a reliability of the detection of the biomarkers and a value representing the abundance of the biomarkers in the patient.
- 15 11. The method of any of the preceding claims, wherein the score is further computed by setting values for one or more of the following:
- g) whether the treatments are recommended in a standard treatment guideline,
- h) an availability of the treatments,
- 20 i) a past treatment history of the patient.
12. The method of any of the preceding claims, further comprising retrieving an identification of a disease indication of the patient and identifying one or more treatments associated with the disease indication of the patient and not associated with any of the set of one or more biomarkers of the patient
- 25 via identifying rules relating the treatments to the disease indication in the treatment information database, and generating a score for the identified

treatments associated with the disease indication, and wherein the treatment option or treatment contraindication prioritization list comprises both treatments associated with any of the set of one or more biomarkers and treatments associated with the disease indication,

5 in particular wherein generating a score for each of the identified treatments associated with the disease indication and not associated with any of the set of one or more biomarkers of the patient comprises extracting average effect sizes with respect to representative mixtures of biomarker values in patients on response, resistance, and/or risk, wherein the effect sizes of the
10 rules for each of the one or more biomarkers comprise a measure of likelihood of response or resistance, a measure of a hazard ratio, a measure of likelihood odds of response or resistance, or a measure of a quotient of hazard ratios of the treatment.

13. The method of any of the preceding claims, further comprising outputting
15 the treatment option prioritization to a user as a list comprising one or more of the highest ordered treatments and/or comprising outputting the treatment contraindication prioritization to a user as a list comprising one or more of the lowest ordered treatments.

14. Curated database comprising information prepared for being used for
20 generating a score for a treatment to be prioritized when performing a method for providing a prioritization of patient treatment options according to any one of claims 1 to 13, in particular comprising one or more of the following:

- data extracted from other databases,

25 - information extracted from text documents via text mining,

- measurement data or processed measurement data from appropriate assays,

- manually entered data,

- sign-offs and/or changes by curators.

15. Clinical decision support device comprising a curated database according to claim 14 and a data mining module.