

METHOD AND SYSTEM FOR GENERATING A METABOLIC DIGITAL TWIN FOR CLINICAL DECISION SUPPORT

TECHNICAL FIELD

5

The present disclosure relates to methods and systems for generating metabolic digital twins for use in clinical decision support.

BACKGROUND

10

Chronic metabolic disease is the result of changes of metabolite fluxes through biomolecular pathways and gene networks accumulated over long periods of time. Obtaining clinically actionable insights in relation to these changes, based on readily measured physiological parameters, remains a challenge. At present, there is no
15 computationally efficient way to bridge the gap between health state characterization obtained through measurement of clinical and laboratory parameters on one hand, and the pathological disturbance of biomolecular processes on the other hand.

20

Traditional enzymatic kinetics models, such as those based on the Michaelis-Menten (M-M) and Briggs-Haldane equations, have been successful in describing dynamics of isolated biochemical reactions. However, metabolic modelling approaches that use these models are not scalable, because they require experimental estimation of kinetic parameters and enzyme amounts and this quickly becomes computationally expensive as the number of reactions increases. Further, these models are unable to account for non-linear effects.

25

To address the above issues, an approach called Metabolic Flux Analysis (MFA) has been proposed. MFA represents a complex biochemical system as a composition of metabolic fluxes connected by stoichiometric relations and the mass balance relations, rather than as a set of enzymatic reactions. Unlike enzymatic kinetics models, MFA uses a set of
30 constraints on possible flux values as inputs to separate the states that the biochemical system can reach from those that it cannot. Another component of MFA is the formal optimization target of the entire metabolic network. Typically, the optimization target is maximal production of a certain metabolite or a metabolite combination. The set of fluxes that maximize or minimize the target is computed as the best fit. Thus the best fit fluxes
35 describe the optimal state of the entire metabolic network.

Despite its potential benefits, practical application of MFA in clinical settings is currently limited to cases where the metabolic network is large enough to include diverse pathways

that can incorporate complex information characterizing a health state. Another limiting requirement of MFA is the need for a large number of metabolite concentrations to be measured at each state, proportional to the complexity of the phenotype. High throughput metabolite concentration measurement is not commonly available in clinics at present. At the same time, many biochemical physiological parameters, such as urine albumin or blood pressure, currently measured in clinics, are not used in MFA. Finally, unlike in experimental settings, disease health state changes can occur at different rates in different patients, such that comparability of health state dynamics in different patients is questionable. If such comparisons cannot be made, then there is no common interpretation of the associated metabolic flux dynamics across patients, and the clinical value of the entire analysis is greatly reduced.

It would be desirable to overcome or alleviate one or more of the above difficulties, or at least to provide a useful alternative.

15

SUMMARY

The present disclosure relates to a method of generating a metabolic digital twin of a subject for clinical decision support in relation to a medical condition, the method comprising:

receiving data indicative of an extended metabolic map comprising nodes representing a plurality of metabolites and one or more non-metabolic parameters, and edges representing relationships between them;

determining an extended stoichiometry matrix at least partly from the extended metabolic map, wherein coefficients in the extended stoichiometry matrix define quantitative relationships between abundances of the metabolites and/or values of the non-metabolic parameters;

receiving data indicative of measurements of a sample of the subject, wherein the measurements comprise measurements of one or more of available metabolite concentrations for one or more of the plurality of metabolites, and measurements of one or more of the non-metabolic parameters; and

optimizing, subject to one or more constraints, an objective function that depends on a product of the stoichiometry matrix and a flux vector, each component of the flux vector corresponding to an edge of the extended metabolic map, wherein the one or more constraints are based on the measurements;

wherein the metabolic digital twin comprises data indicative of a best-fit flux vector obtained from said optimizing, and wherein components of the best-fit flux vector are

indicative of metabolite-metabolite fluxes and metabolite-physiological fluxes for the subject.

The present disclosure also relates to a method for identifying one or more biomarkers associated with a disease phenotype, comprising:

5 obtaining, for a plurality of subjects each having one or more status indicators corresponding to one or more respective disease phenotypes, data indicative of a plurality of respective metabolic digital twins, wherein each metabolic digital twin is generated by a method as disclosed herein; and

10 determining an indication of statistically significant association of respective components of the flux vector with the one or more status indicators;

wherein the one or more biomarkers comprise one or more components of the flux vector having statistically significant association with a corresponding disease phenotype.

15 The present disclosure further relates to a method for identifying metabolic flux patterns associated with a disease phenotype, comprising:

obtaining, for a plurality of subjects each having one or more status indicators corresponding to one or more respective disease phenotypes, population data indicative of respective flux vectors of a plurality of respective metabolic digital twins, wherein each
20 metabolic digital twin is generated by a method as disclosed herein;

clustering the population data to generate a plurality of clusters;

performing one or more pair-wise tests using respective first and second clusters of said plurality of clusters, and the one or more status indicators; and

25 for each statistically significant test of said one or more pair-wise tests, a difference between a median flux vector of subjects in the respective first cluster and a median flux vector of subjects in the respective second cluster.

The present disclosure further relates to a method of predicting progression of a medical condition in a subject, comprising:

30 generating a metabolic digital twin of the subject by a method as disclosed herein; and

outputting a comparison of one or more components of the best-fit flux vector for the metabolic digital twin to one or more corresponding components of flux vectors of metabolic digital twins for a reference population of subjects;

35 wherein said comparison is indicative of progression of the medical condition in the subject relative to the reference population.

The present disclosure also relates to a system comprising:

storage; and
at least one processor in communication with the storage;
wherein the storage comprises machine-readable instructions for causing the at least one processor to carry out a method as disclosed herein.

5

The present disclosure further relates to non-transitory computer-readable storage having stored thereon instructions for causing at least one processor to carry out a method as disclosed herein.

10 BRIEF DESCRIPTION OF THE DRAWINGS

Some embodiments of methods and systems for generating metabolic models, in accordance with present teachings will now be described, by way of non-limiting example only, with reference to the accompanying drawings in which:

15

Figure 1 is a block diagram of an example system for generating and making use of metabolic digital twins for clinical decision support, showing an example of data flow between the system components;

20

Figure 2 is a flow diagram of an example of a method of generating a metabolic digital twin for clinical decision support;

Figure 3 is a schematic overview of an example method of generating a metabolic digital twin from a patient's biomarker data;

25

Figure 4 is a schematic depiction of an example of a method of constructing a metabolic map based on available clinical data inputs;

Figure 5 shows an example workflow for generating and using a metabolic digital twin for clinical decision support for a subject;

30

Figures 6(a) and 6(b) illustrate an example metabolic map and an example simplified metabolic map;

35

Figure 7 is a glycolysis metabolic map used in a simulation study according to an example, and representing the biochemical reactions that occur during glycolysis in muscle tissue, from glycogenolysis to lactate production;

Figure 8 shows the computational simulation results for a glycolysis example for a simulation time of 60 minutes, with the evaluated system state representing muscle at rest and the initial conditions for the simulation being provided in Table 1;

- 5 Figure 9 shows computation simulation results for the same glycolysis example as Figure 8, for a simulation time of 30 seconds;

Figure 10 shows evidence of the linearity of the parameters in the generalized flux coordinate system observed in the glycolysis model under a wide range of perturbations;

10

Figure 11 is a simplified metabolic map used in a Type 2 diabetes study according to an example;

- 15 Figure 12 shows the results of using embodiments of the present method to predict development of (a) retinopathy and (b) nephropathy in type 2 diabetes within 3 years from the time point of the patient data input; and

- 20 Figure 13 shows the results of using embodiments of the present method to predict development of coronary artery disease in type 2 diabetes within 3 years from the time point of the patient data input.

DETAILED DESCRIPTION

- 25 Embodiments of the present disclosure make health states and their progression comparable across subjects by replacing the time variable with an extent variable quantifying the degree of progression between two extreme health states. For example, in the context of metabolic syndrome, change in BMI may be chosen as the extent variable. By using a non-temporal extent variable to characterise progression of disease, it becomes possible to more directly compare subjects, and thus to draw inferences and make
30 predictions regarding disease progression.

- Referring to Figure 1, a system 100 for generating and making use of metabolic digital twins comprises a digital twin generation component 110 that receives input data from one or more data sources. The input data may comprise patient data comprising metabolic,
35 physiological and/or clinical data relating to one or more subjects, and may for example be at least partly retrieved from a patient database 120. The patient data may comprise measurements of one or more physiological parameters and/or one or more metabolite concentrations for the one or more subjects. The input data may also comprise disease-

specific data from a knowledge base 122. The disease-specific data may comprise metabolic map data relating to a metabolic map of one or more metabolic and/or physiological processes relevant to a disease. The input data may further comprise data from other external data sources 128, such as one or more scientific literature databases (e.g. PubMed), one or more repositories of data relating to biochemical compounds and reactions (such as KEGG or MetaCyc), or one or more model repositories (such as BioModels, SBML models for KEGG, CellML repository, PANTHER pathways, SigPath, SIGMOID, and the like). Such model repositories may include entire stoichiometry matrices which may be readily adapted for use with embodiments of the present disclosure.

The digital twin generation component 110 is configured to obtain the input data, and perform an optimization process, using the measurements of the one or more physiological parameters and/or one or more metabolite concentrations, to determine a best-fit flux vector for each of the one or more subjects. For a given subject, a best-fit flux vector represents a flux profile for the subject in the metabolic map, with otherwise unobservable quantities being able to be inferred based on actual measurements of a limited number of parameters, and can be considered to be (or to form part of) a metabolic digital twin of the subject. As used herein, a metabolic digital twin refers to at least a list of flux magnitudes (also referred to herein as a vector of flux values), noting that a "flux" may refer to either a metabolic or non-metabolic flux as will be explained further. The process for generating digital twins will be described in further detail below.

The digital twin generation process may be performed for multiple subjects, and the respective digital twins then stored in a digital twin database 124. The stored digital twins may subsequently be retrieved for further analyses. To this end, the system 100 may comprise a population-based analysis component 112 for performing population-based analyses using the digital twin data 124, and a biomarker/pathway inference component 114 for determining one or more biomarkers and/or one or more metabolic pathways associated with disease phenotypes, using the outputs of the population-based analysis component 112 and/or the digital twin data 124. The results of such analyses may be stored in a biomarker database 126, and may subsequently be used to provide diagnostic or prognostic decision support to a clinician or other user. The system 100 may comprise a clinician report component 116 for this, and other, purposes.

The system 100 may comprise one or more computer processors forming part of one or more computing systems, and each of the components 110, 112, 114, and 116 may be implemented in the form of computer-readable instructions stored on transitory and/or

non-transitory storage and executable by the one or more processors to carry out the processes described herein. For example, the computer-readable instructions may be stored on one or more solid state or magnetic storage media for access by the one or more processors to cause execution of the processes described herein. In some embodiments, 5 databases 120, 122, 126, and/or 128 may also be stored on such media as part of system 100. In other embodiments, one or more of these databases may be located remotely from the system 100 and accessible via a local area or wide area network. In some embodiments, each of the components 110, 112, 114, 116 may be located at physically separate computing devices.

10

Turning now to Figure 2, an example method 200 of generating a metabolic digital twin will be described. The method 200 may be performed partly or entirely by the digital twin generation component 110 of system 100, for example.

15

The method 200 begins at step 202 by receiving data indicative of a metabolic map comprising nodes representing a plurality of metabolites and one or more physiological parameters, and edges representing relationships between them. In metabolic maps used in embodiments of the present disclosure, the one or more physiological parameters can be considered to be exchange fluxes.

20

Each node in the metabolic map identifies a different metabolite or physiological parameter, and each edge represents a flux. Where an edge joins two metabolite nodes, it represents a stoichiometric ratio between the abundances of the metabolites. Where an edge joins a metabolite node with a non-metabolite (physiological parameter) node, it represents a coefficient of a regression relationship between the metabolite abundance and the value of the physiological parameter.

25

The data indicative of the metabolic map can be obtained in various ways. For example, the metabolic map may be defined using Systems Biology Markup Language (SBML) and/or another framework that is suitable for representing graphs, including directed graphs or directed acyclic graphs (DAGs). The SBML or other representation may be stored in knowledge base 122 for retrieval by the digital twin generation component 110.

30

A metabolic map may be defined at least partly based on information available from scientific literature, or from some other source of prior knowledge regarding metabolic processes relevant to the disease under consideration. For example, one or more SBML models of one or more relevant metabolic processes may be obtained from a repository such as the BioModels repository. Such metabolic models typically only represent

35

stoichiometric relationships between metabolites, and not between metabolites and exchange fluxes. Accordingly, in the presently described embodiments, a modified metabolic map may be defined that incorporates exchange fluxes.

- 5 A metabolic system characterized by a stoichiometric matrix exchanges matter and information with the environment. Each such exchange process can be represented as a flux, referred to as an exchange flux, which quantifies the rate of incoming or outgoing matter or information.
- 10 Systemic incoming and outgoing exchange fluxes are not subject to mass conservation laws in the same way as regular fluxes, as they represent links to unconstrained resources. Therefore, together with stoichiometrically-constrained metabolic fluxes, they form an "extended" stoichiometry matrix.
- 15 In some embodiments, the modified metabolic map may be simplified by grouping individual metabolites into pools, such that the metabolic system is modelled with a more coarse-grain network that is simpler to be parameterized and computationally analyzed.

For example, free fatty acids (FFA) represent a group of compounds endogenously synthesized in the liver and adipose tissue from carbohydrates via the fatty acids synthesis pathway and are catabolized via beta-oxidation. The main exogenous source of FFA is ingested triglycerides. Since all individual FFA types share the same incoming and outgoing metabolic fluxes, from the point of view of the remaining system they are indistinguishable. Therefore, in any metabolic map that includes only a subset of the above

20 fluxes, any particular type of FFA would be represented by a node topologically identical to any other FFA type. Thus, a single metabolic entity FFA would adequately represent the multitude of all FFA species. The respective stoichiometric coefficients of the single metabolic FFA entity then would represent the sum of fluxes of each FFA type weighted by their molar contribution to acetyl-CoA, carbohydrates and triglycerides.

30 Similarly to metabolites, the topological argument also applies to metabolic fluxes. A group of metabolic fluxes can be reduced to a single pooled flux if the each flux of this group is topologically identical to any other of the same group. Consider M metabolites connected with the remaining system via only one incoming and one outgoing flux, their interaction with the system is identical to that of a single metabolite connected with the incoming and

35 outgoing fluxes. This situation is illustrated in Figure 6(a), in which metabolites M1 (602), M2 (604) and M3 (606) receive a single incoming flux from metabolites IN1 (608) and IN2 (610), and have a single outgoing flux (that is combined with IN3 (612) and IN4 (614) to

produce OUT (618)). Such a map can be simplified by grouping nodes M1 (602), M2 (604), and M3 (606) together to form a pooled node MX (616), as shown in Figure 6(b). The mass accumulated across all M metabolites is identical to the difference between the incoming and the outgoing flux magnitudes. At the same time, if there are N parallel fluxes
5 connecting two metabolites, their interaction with the system is identical to that of a single flux whose magnitude is a sum of all N pooled fluxes. Consider a non-branching pathway of N fluxes connecting N-1 metabolites. If all the N-1 metabolites and N-2 fluxes internal to the pathway are not available for observation, there are only two fluxes interacting with the remaining system: the first and the N-th. Thus such a pathway can also be represented
10 as a single metabolite whose mass accumulates from the difference of the incoming and outgoing fluxes.

Thus, in a metabolic graph, any connected subset of parallel or branching fluxes and metabolites, whose values are not available for observation or deducible, can be
15 represented as a single pooled node interacting with the remaining system via a smaller set of incoming and outgoing fluxes.

Once the metabolic map data has been obtained, an extended stoichiometry matrix for the metabolic map may be generated, at step 202. The stoichiometry matrix is "extended"
20 in the sense that it captures relationships not just between metabolite abundances (which must obey mass conservation laws), but also between metabolite abundances and values of physiological (non-metabolic) parameters. In some embodiments, the metabolic map may be analyzed to automatically determine the relationships between its edges and nodes, and to determine the appropriate stoichiometric coefficients based on those
25 relationships. The stoichiometric coefficients may be determined based on mass conservation laws and/or based on empirical data. In other embodiments, the stoichiometric coefficients may be retrieved from a database, such as knowledge base 122. In some embodiments, the stoichiometric coefficients may be stored as metadata associated with the metabolic map, and the metabolic map data may comprise both the
30 metabolic map itself and the associated metadata.

Next, at step 206, data relating to measurements of a sample of the subject may be obtained. The measurements include respective values of one or more physiological parameters represented in the metabolic map, and may optionally also include
35 measurements of one or more metabolites represented in the metabolic map. The measurements may be obtained by retrieving them from patient database 120, or by receiving user input indicative of such measurements. For example, a clinician user

interacting with the digital twin generation component via a user interface may enter values of one or more physiological parameters that have been measured for the patient.

Next, at step 208, the method comprises optimizing an objective function to obtain a set of best-fit fluxes based on the extended stoichiometry matrix and the measurements of the subject.

When considering a metabolic system undergoing a long-term smooth change along a trajectory between states A and B, this gradual change of the system's state is commonly quantified by the variable expressing the number of molecular transformations of one particular type that has to occur while the system transforms from state A to state B. This variable is termed the reaction extent ξ . Real-life chemical systems consist of a large, but finite number of discrete molecules, and their minimal unit of the extent ξ is a single molecular transformation. However, in most real-world chemical and biochemical applications, the number of molecules is so large that it is impractical to trace reaction extent by individual molecules. Therefore, thermodynamic laws apply to the metabolic system. Its evolution can be analyzed as a smooth function of ξ (instead of as a function of time) along with thermodynamic variables.

Accordingly, fluxes and metabolite concentration changes can be expressed in the differential form:

$$\frac{d[X_i]}{d\xi} = \sum_j S_{i,j} v_j(\xi) \quad (1)$$

Thus, a connection between the extent ξ , time t and metabolite concentrations X^A and X^B (at states A and B, respectively) can be calculated as follows:

$$X^B = X^A + \int_A^B \xi d\xi = X^A + \int_A^B S v(\xi) d\xi = X^A + \int_{t_A}^{t_B} \xi \frac{\partial \xi}{\partial t} dt \quad (2)$$

When changes of metabolite concentrations and metabolic fluxes are fully defined as functions of ξ , any two systems with identical ξ are characterized with identical metabolite concentrations, fluxes and evolution path. Thus, instead of analyzing the sequence of states along the trajectory in time (evolution in time), it is possible to analyze a representative ensemble of states (evolution of states) to calculate changes of system parameters. Such systems are termed ergodic.

At steady state, the concentration of metabolites in a network is constant and the net flux generating each metabolite equals the net flux consuming it. In conventional MFA this

condition is achieved when the time derivative of the metabolite concentration is zero. When the system reaches a steady state, no further evolution of the system is expected, unless it is driven externally. Notably, the extended formalism of embodiments of the present disclosure using the reaction extent ξ naturally allows handling of both steady and
 5 non-steady state conditions.

In the clinical setting, it is frequently more convenient to measure macroscopic physiological parameters instead of metabolite concentrations. For example, pulse wave velocity (PWV) is easily measured with a cuff at the upper arm and leg. At the same time,
 10 measuring the related plasma concentrations of oxidized low-density lipid cholesterol (ox-LDL) requires substantial laboratory effort.

The better the linear correlation between the non-metabolic parameter, e.g. PWV, and the related metabolite concentration, e.g. ox-LDL, the more justified is an inclusion of an
 15 additional edge between the two characteristic nodes (here: the edge between PWV and ox-LDL) into the metabolic network and the description with a generalized flux between them. The slope of this regression enters the stoichiometric matrix as S_{ij} , where i is the index of the dependent node, e.g. ox-LDL, and j is the index of the edge leading to the independent node, e.g. PWV.

20 In embodiments of the present disclosure, the stoichiometry matrix describes the relationship between the rates of change of quantities represented by nodes of the metabolic map as follows:

$$\frac{dX_i}{d\xi} = \sum_j S_{ij} \frac{dX_j}{d\xi} \quad (3)$$

25 The differential at the right side of the equation can be termed the generalized flux v_j . For the edge in the reverse direction, S_{ji} is the reciprocal value of the slope. .

Once non-metabolic nodes are included into the molecular network, it becomes possible
 30 to incorporate nodes quantifying disease progression into the same network. Diseases are characterized by development of syndromes that have measurable parameters. For example, endothelial dysfunctions progression can be quantified by the reactive hyperemia index (RHI) that can be included as a node related to the total and low density lipid cholesterol plasma concentrations. In this framework, the reaction extent ξ , which is
 35 conventionally interpreted as a concentration change, can instead be generalized as a state progress variable ξ , for example the extent of disease progression measured by quantifying a syndrome.

In the traditional MFA approach, due to the direct interpretation of metabolite concentrations and the stoichiometry matrix, it is only possible to evaluate only variables that are stoichiometric, i.e. that are bound by mass conservation laws. This rule does not strictly apply to the incoming and outgoing fluxes connecting the biochemical system with the environment. Within the traditional MFA framework they are termed exchange fluxes. Since they represent unlimited sources and sinks of matter for the biochemical system, exchange fluxes are not subject to the same mass conservation laws as regular fluxes. From the point of view external to the system (i.e. the environment), the system transforms the incoming exchange fluxes to the outgoing exchange fluxes, subject to a particular state of the system.

Since the process extent defining state is an internal variable of the system, the state of the system cannot be inferred from the exchange fluxes alone. Therefore, the environment receives outgoing exchange fluxes as the responses of the system in an unknown state to the incoming exchange fluxes. This description is also applicable to measurements of medical parameters directly or indirectly dependent on the metabolic state of the subject. This provides a way to extend the MFA methodology to medical applications by extending the flux list with quantifiable non-stoichiometric variables as exchange fluxes.

Defining the stoichiometric coefficient connecting a physiological variable with another variable presents another challenge that is hard to address with the traditional MFA. One reason is wide variability of the time scales of measurable physiological changes. Another reason is the impossibility to apply to such fluxes the principle of consuming and producing mass, since physiological variables often cannot be interpreted as masses.

The solution for any two variables (i and j) can be found by applying the progress extent approach defined in the following equation to evaluate the local states (A and B). The stoichiometric coefficient S_{ik} can be derived empirically from points A and B. As an example, if the measured observations (X_i and X_j) of variables i and j are defined at points A and B, the coefficient S_{ik} may be quantified as a change in X_i in the range, relative to the respective change in X_j : $S_{ij} = (X_i^B - X_i^A) / (X_j^B - X_j^A)$. In a general case, X_i and X_j are required to be continuous in the region between points A and B, and the value of the coefficient S_{ij} is obtained as the directional derivative at the point where the measurement is done along the direction of the progression extent ξ at that point (Equation K below). When the relationship between X_i and X_j is linear along ξ between A and B, a linearity condition applies, and the coefficient S_{ij} is represented by a constant (Equations E-G below).

The vector of observables X of the metabolic system at a given state represents the metabolite concentrations and physiological variables measurement taken at that state. The difference between the observations at two states $\Delta X = X^B - X^A$ characterizes the progression extent ξ . The best fit set of generalized fluxes v satisfying Equation (1) can be obtained by minimizing the squared difference between the observed state variable change ΔX and the best fit estimate of it, Sv , which is the expected change of the state variable given the generalized fluxes v .

In certain embodiments, X^A is the vector of measurements for the subject for whom the digital twin is being generated, who is in a state of progression A of the medical condition. X^B is a vector of measurements from either another individual, or from a "population average" at state B . The extent variable ξ is a metric of progression of the subject at state A towards state B . If individual (or population) B represents an extreme case of progression (severe state of a medical condition), the value of ξ calculated for subject A will provide an indication as to how far subject A is from developing the severe state of the disease.

The best fit solution for fluxes might result in a deviation from the observed state variables due to conflicting experimental measurement errors as well as due to approximately inferred components of the extended stoichiometric matrix. Therefore, the flux solution can be allowed to result in a deviation from the observed state variables within an empirically set vicinity ϵ .

$$\begin{cases} v = \operatorname{argmin}(v^T S^T S v - S^T \Delta X) \\ Sv \leq (1 + \epsilon) \Delta X \\ Sv > (1 - \epsilon) \Delta X \\ v \leq v_{\max} \\ v > v_{\min} \\ \epsilon = 0.2 \end{cases} \quad (4)$$

The solution to the optimization problem defined in Equation (4) can be obtained using the Goldfarb-Idnani quadratic programming procedure, for example using the Quadprog v.0.1.8 software library. The calculated generalized metabolic fluxes v are the best fit solution of Equation (4). These generalized metabolic fluxes v are then output as the metabolic digital twin of the subject.

A schematic overview of an example method of generating a metabolic digital twin from a patient's biomarker data input obtained at a single time point is shown in Figure 3. As shown in this example, the method 200 (represented in Figure 3 as "generalized flux

analysis") takes as input a vector $\underline{x}$ of measurements of the patient, where the vector comprises measurements of available (i.e. measurable) metabolite concentrations and of one or more non-metabolic (physiological) parameters. An extended stoichiometry matrix S is also input to the process 200. Further, the process 200 takes as input a set of constraints for the fluxes and/or for one or more unobserved parameters (e.g., metabolite concentrations that are present in the stoichiometry matrix S but are not available in the measurements X). A vector Y of reference measurements, for the same parameters as measured in the vector X , are also input to the process 200 and these may be obtained from a reference population such as an "average" diabetes patient (e.g., each parameter in Y is averaged over the reference population). A choice of optimization procedure, such as quadratic minimization, may also be input to process 200 (if not selected by default). The process 200 is executed as disclosed above to obtain a metabolic digital twin represented as a vector v of fluxes.

In some embodiments, a graphical representation of the metabolic digital twin may be generated. For example, as shown at bottom right of Figure 3, the metabolic flux magnitudes and directions realized in the output of the process 200 (the digital twin model) can be represented by thickness, directions and colors of the arrows overlaid on the input metabolic graph.

Figure 4 shows a schematic depiction of an example of a method of constructing a metabolic map based on available clinical data inputs. At step 402 a list of observable quantities is obtained. The observable quantities may comprise measurable metabolite concentrations and non-metabolic parameters such as physiological or vascular parameters. Next, at step 404, these quantities are mapped in the metabolic context of the target health states. At step 406 the observables are connected with fluxes via observed and unobserved species to obtain the metabolic map and extended stoichiometry matrix. The obtained metabolic map/extended stoichiometry matrix are then used as inputs to the digital twin generation process 200 as before, together with observed input data that is obtained (step 408) for a subject. Note that in Figure 4, the empty white boxes in the diagram represent unobservable biomarkers, while colored arrows represent the generalized fluxes (the components of the flux vector for the digital twin).

Turning now to Figure 5, an example workflow for generating and using a metabolic digital twin for clinical decision support for a subject is shown. Input data for the workflow includes subject-specific data including a measurement vector (biomarker input) X for the subject 510, and a set of K non-flux parameters for the subject 512. Input data also includes a reference vector Y (which may be from an individual or derived from a reference

population each member of which is characterised by the same progression state of the medical condition under consideration) 502, a constraints matrix G 504, an extended stoichiometry matrix S 506, and a choice of optimization function f_{opt} 508. Digital twin generation process 200 then obtains, via the optimization process described above, a
 5 metabolic digital twin comprising a flux vector v_{opt} .

Non-limiting examples of non-flux parameters are the subject's age, ethnicity and gender.

The optimal generalized fluxes (v_{opt}) can, together with the K non-flux parameters, be
 10 used as inputs to a classifier 520 such as a binary regression function. The classifier 520 may be trained using a plurality of metabolic digital twins each having known phenotype, and the trained classifier 520 then being used to generate a phenotype prediction for the subject using the obtained v_{opt} for the subject. For example, for a binary regression classifier, the prediction is one of two possible outcomes (health states).

15 The present invention now discusses mathematical basis of embodiments of the extended flux balance analysis. We first discuss quantification of system state change in the units of extent.

20 The inter-connectivity of metabolites within a network of biochemical reactions is given by reaction equations defining the stoichiometric conversion of substrates into products for every reaction. Enzymatic reactions, as well as the transport of metabolites across system boundaries constitute fluxes, which serve to dissipate and generate metabolites. In following the law of conservation of mass, material balances describing the activity of a
 25 particular reactant through each reaction can be written where the difference between the rate of production and consumption of a particular metabolite is equivalent to the change in concentration of that metabolite over time. Here time is chosen as the variable representing the production/consumption process extent, while using other extent variables is permitted. This yields the following MFA equation for every metabolite in a
 30 system:

$$\frac{d[X_i]}{dt} = \sum_j N_{i,j} v_j \quad (5)$$

where v_j are the fluxes that produce and consume the metabolite in the system and the stoichiometric coefficient $N_{i,j}$ stands for the number of moles of metabolite X_i formed in reaction j . It is negative if X_i is a substrate of this reaction.

35 At steady state (homeostasis), the concentration of metabolites in a network is constant and the activity of those fluxes that generate a metabolite must equal to the activity of

the fluxes that consume the metabolite. Since the time constants associated with growth of the organism are much larger than those associated with individual reaction kinetics, it is reasonable to consider the metabolic system being in a steady state when investigating aspects of metabolism related to growth. This reduces the above system of equations to a system of homogeneous linear equations, which in matrix notation is

$$N \cdot v = 0 \quad (6)$$

The size of the stoichiometric matrix N is $m \times n$, where m corresponds to the number of metabolites and n is the number of reactions or fluxes taking place within the network. The vector v refers to the magnitude of each flux. Thus the combination of magnitudes of all n fluxes define the metabolic phenotype of the biological system at a given time point or a steady state.

The stoichiometric matrix links the vector of time derivatives of the metabolite concentrations with the vector of reaction rates.

Equation (5) is applicable to describe progression of the metabolic system between states A and B under the assumption that fluxes v_j are time-dependent in the progression scale ($v_j = f(t)$). Thus the fluxes and the time are considered smooth variables in the interval between A and B .

$$\begin{cases} A, B \in Xv(t): \forall t, \exists X(t) \in [A, B] \\ t: t \geq 0, t_x < t_y, \forall t_x \neq t_y \\ \forall v(t) \leftrightarrow X(t) \in [A, B] \exists \frac{dx}{dt} \end{cases} \quad (7)$$

In a more general case, when progression is not assumed smooth in time, let the flux v_j represent a function on measure μ on states A and B , which belong to the domain of the metabolic process in question, quantifying the progression. This measure quantifies the change between states A and B and thus the metabolic process as its extent ξ . The extent ξ quantifies the number of metabolite transformations occurring during the system's transition from state A to state B , frequently normalized per mole by the Avogadro's constant N_A . Chemical kinetics defines reaction rate as the extent per second. This provides an equivalent formulation:

$$\begin{cases} A, B \in X \\ \xi(B - A) = \frac{B - A}{N_A} \\ v(\xi): \forall \xi \exists X(\xi) \in [A, B] \end{cases} \quad (8)$$

Thus, a connection between the extent ξ , time t and metabolite concentrations X^A and X^B (at states A and B , respectively) can be defined as follows:

$$X^B = X_A + \int_A^B \xi d\xi = X^A + \int_A^B Nv(\xi)d\xi = X^A + \int_{t_A}^{t_B} \xi \frac{\partial \xi}{\partial t} dt \quad (9)$$

Equation (9) shows that the extent ξ and fluxes $v(\xi)$ defined on the the extent are state functions in the thermodynamic sense, i.e. their values at any given state do not depend on the path the system took to arrive to the state. Thus, if we consider an ensemble of independently measured systems progressing between states A and B , all members of the ensemble are mathematically identical, independent of the individual's time history of progression and have identical expected future state on the scale of ξ :

$$X(\xi + \delta\xi) = X_B + Nv(\xi_B + \delta\xi) \quad (10)$$

10

Let us consider progression of an individual member of the ensemble along a timeline. If the progression reaches a steady state, no further changes of the state will be observed for any subsequent duration of time. From Equation (9) it also follows that any other individual system reaching that state will also remain unchanged.

15

Altogether, such individual systems will form their ensemble characterizing the steady state, which in its turn will be an attractor. This also means that, similarly to thermodynamic state variables, ξ and $v(\xi)$ are ergodic, with the average across the members of the ensemble with respect to their individual progression time points being identical to the average across the states in the limits:

20

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n-1} Nv(\xi)^i = \int_0^\infty Nv(t)dt \quad (11)$$

Moreover, Equation (9) applies to any state of the system, regardless of it being steady or not, as long as it is defined on ξ . Therefore, at each point of the scale ξ , the respective flux $v(\xi)$ generalizes an ensemble of all individual systems with any prior time progression trajectories. This property satisfies the definition of a strong ergodic system, where for each point state of progression there can be found an ensemble of systems, the state average of which is equivalent to the average across all the timelines of the ensemble. Therefore, for each point state of an individual's progression in ξ with fluxes $v(\xi)$ there can be found an ensemble of other individuals representing the same state without losing the same point state if and only if their fluxes are identical to $v(\xi)$.

30

In an approximation, this can be expressed as follows:

$$\forall \delta > 0 \exists \epsilon > 0: \| \hat{v} - v(\xi + \delta\xi) \| < \epsilon \xi + \delta\xi \quad (12)$$

35

Similar to Equation (5), fluxes and metabolite concentration changes can be also expressed in the differential form:

$$\frac{d[X_i]}{d\xi} = \sum_j N_{i,j} v_j(\xi) \quad (13)$$

In Equation (13), the stoichiometry matrix N represents a table of constant coefficients, and therefore Equation (13) expresses a linear dependence of metabolite level changes on the extent variable ξ .

5

In Equation (13), metabolite level changes $dX/d\xi$ along the scale of ξ are expressed as a linear transformation of flux magnitude $v(\xi)$ since the stoichiometry matrix of the system is taken as a constant. If, at the same time, the fluxes are linear functions of ξ , the entire expression (13) describes a linear dependency between any two metabolic states. This linear dependency may arise in one of the following conditions.

10

The first condition is linearity bounded by constraints, that is,

$$\forall \xi \exists \{c_k = \text{const}\}_{k=1:K}: v(\xi) = \sum_{k=1}^K c_k v_k \xi \quad (14)$$

The second condition is local linearity, that is,

$$\exists(\delta, [\xi - \delta, \xi + \delta], \{c_k = \text{const}\}_{k=1:K}): v(\xi) = \sum_{k=1}^K c_k v_k \xi \quad (15)$$

15

The third condition relates to steady state in ξ . There exist points on ξ , where the left side of Equation (13) approaches zero:

$$\forall \delta > 0 \exists \xi_k: \left| \frac{dX}{d\xi} \right| = |Nv(\xi_k)| < \delta \quad (16)$$

20

The fourth condition relates to linear map between two steady states. For any two steady state points described in Equation (16) the linear path between them corresponds to the minimal possible perturbation of the system allowing transition between the two points:

$$\forall \delta > 0 \exists (\xi_p, \xi_q, \{c_k = \text{const}\}_{k=1:K}): |Nv(\xi_p)| < \delta, |Nv(\xi_q)| < \delta, \sum_{k=1}^K c_k (v(\xi_p) - v(\xi_q)) \xi \quad (17)$$

25

Condition 3 (Equation (16)) describes a steady state assumption. It implies that the system reaches such state when no significant changes of metabolite concentrations are observed, and the system ceases to progress. Cases describing this condition are often considered in the vicinity of a single steady state point. Using progress extent derivative instead of time derivative extends the application of steady state analysis to case 3 (Equation (17)).

30

For each linearization condition (Equations (14) to (16)), it is possible to find the values of $v(\xi_k), \{c_k\}$ and $dX/d\xi$ by a linear or quadratic optimization technique.

- In the quadratic optimization method, the best fit is obtained by minimizing the deviation from a defined progress extent ξ constrained within the limited deviations δ and ϵ from the expected flux value $\hat{v}$ and metabolite deviation from the reference point $\hat{N}v$

$$\begin{cases} \text{minimize: } \frac{1}{2} v^T N^T N v + \hat{N} v^T N v \\ \text{subject: } [Nv; -Nv] \leq [\hat{N}v + \epsilon; -(\hat{N}v - \epsilon)] \\ \text{subject: } [v; -v] \leq [\hat{v} + \delta; -(\hat{v} - \delta)] \end{cases} \quad (18)$$

- The present invention now discusses derivative along the extent variable generalizes stoichiometric coefficients for metabolic and non-metabolic variables.

From Equation (13) the value of a stoichiometric coefficient can be found:

$$N_{ik} = \frac{\partial x_i}{\partial \xi v_k} \quad (19)$$

- It may appear that the ratio of the stoichiometric coefficients N_{ik} and N_{jk} of the same flux k is independent of the extent ξ :

$$\frac{N_{ik}}{N_{jk}} = \frac{\partial x_i}{v_k \partial \xi} \frac{v_k \partial \xi}{\partial x_j} = \frac{\partial x_i}{\partial x_j} \quad (20)$$

- Indeed, if metabolites i and k are related with mass conservation law, the two differentials cancel each other, because along the extent ξ any change in the magnitude of v_k will proportionally change all metabolite concentrations connected with this flux.

$$\frac{N_{ik}}{N_{jk}}(\xi) = \frac{\partial x_i}{v_k \partial \xi} \frac{v_k \partial \xi}{\partial x_j} = \frac{\partial \xi \partial x_i}{\xi \partial v_k \partial \xi} \frac{\xi \partial v_k \partial \xi}{\partial \xi \partial x_j} = \frac{\partial x_i \partial v_k}{\partial v_k \partial x_j} \quad (21)$$

- However, Equation (21) contains an implicit differentiation of x_i and x_j by ξ . In a general case, when conservation laws between the quantities are not implied, the differentials in the inner expression should be taken into account. This would make N_{ik} and N_{jk} dependent on the $v_k(\xi)$ at a particular state in which these coefficients are defined. Therefore, Equation (21) can be written in the form of derivative along the curve ξ at a given state P :

$$\frac{N_{ik}(\xi)}{N_{jk}(\xi)} \Big|_P = \frac{\partial x_i(\xi) \partial \xi}{\xi(P) \partial v_k(\xi)} \frac{\xi(P) \partial v_k(\xi)}{\partial x_j(\xi) \partial \xi} \Big|_P = \frac{\partial x_i \partial \xi}{\partial v_k \partial x_j} \frac{\partial v_k}{\partial \xi} \Big|_P = \frac{\partial x_i \partial \xi}{\partial \xi \partial x_j} \quad (22)$$

Consider the case where the system has been observed in states A to state B along ξ , and two variables i and j were measured at both states.

$$\begin{cases} \frac{x_i(B) - x_i(A)}{\xi(B) - \xi(A)} = \frac{\Delta_{AB} x_i}{\Delta_{AB} \xi} \\ \frac{x_j(B) - x_j(A)}{\xi(B) - \xi(A)} = \frac{\Delta_{AB} x_j}{\Delta_{AB} \xi} \end{cases} \quad (23)$$

If x_i and x_j change linearly along ξ in the interval between A and B ,

$$\forall (x_i, x_j, \xi) \in [A, B]: \begin{cases} \frac{\Delta_{AB} x_i}{\Delta_{AB} \xi} = \frac{\partial x_i}{\partial \xi} \\ \frac{\Delta_{AB} x_j}{\Delta_{AB} \xi} = \frac{\partial x_j}{\partial \xi} \end{cases} \quad (24)$$

- 5 So, for two variables (i and j) not connected with mass conservation (e.g. two physiological variables or a physiological variable and a metabolite concentration), a pair of stoichiometry coefficients can be found as a solution of the following system of equations:

$$\begin{cases} \left(\frac{N_{ik}}{N_{jk}}(\xi) \right) \Big|_A^B = \frac{\partial x_j}{\partial x_i}(\xi) \Big|_A^B = \frac{\Delta_{AB} x_i}{\Delta_{AB} x_j} \\ \left(\frac{N_{ik}}{N_{jk}}(\xi) \right) \Big|_C^D = \frac{\partial x_j}{\partial x_i}(\xi) \Big|_C^D = \frac{\Delta_{CD} x_i}{\Delta_{CD} x_j} \end{cases} \quad (25)$$

10

Equation (25) suggests a method to find the generalized coefficients N_{ik} and N_{jk} from observations. When one coefficient, e.g. N_{jk} is known, the other can be found from measurements of two variables (x_i and x_j) at two points (A and B). When both are unknown, measurements at four points (A , B , C and D) would be required.

15

Practically, it means that it should be possible to find the coefficients in the stoichiometry matrix relating two physiological variables, as well as any physiological and any metabolic variable.

- 20 The present disclosure now illustrates a first example regarding computational modelling and analysis of glycolysis.

To illustrate key features of the extended MFA framework of embodiments of the present disclosure, and to compare the results produced thereby with those of the steady-state enzyme kinetics approach, we analyzed the simple example of glycolysis in muscle tissue. The simulated metabolic pathway (shown in Figure 7) involves 18 metabolites (Table 1) and 12 reversible reactions from glycogenolysis to pyruvate and lactate production (Table 2). Table 1 shows simulation results of glycolysis in the muscle. Table 2 shows simulation results of human glycolysis in the muscle. We carried out a set of computational simulations by applying the M-M and the MFA modeling methods.

ID	Metabolite	Initial concentration, Mole/L
GLY	Glycogen	0.112
G1P	Glucose 1-phosphate	5.0e-05
G6P	Glucose 6-phosphate	7.5e-04
F6P	Fructose 6-phosphate	2.0e-04
FDP	Fructose 1,6-bisphosphate	7.0e-5
GAP	Glyceraldehyde 3-phosphate	3.0e-5
DHAP	Dihydroxyacetone phosphate	7.0e-5
DPG	1,3-Bisphosphoglycerate	6.5e-05
P3G	3-Phosphoglycerate	5.0e-05
P2G	2-Phosphoglycerate	5.0e-06
PEP	Phosphoenolpyruvate	1.0e-05
PYR	Pyruvate	1.4e-04
LAC	Lactate	2.1e-3 [*] / 6.0e-3 ^{**}
NAD	Nicotinamide adenine dinucleotide oxidized	5.0e-04
NADH	Nicotinamide adenine dinucleotide reduced	5.0e-07
ADP	Adenosine diphosphate	7.8e-6 [*] / 4.1e-4 ^b
ATP	Adenosine triphosphate	8.2e-3 [*] / 7.8e-3 ^b
Ph	Inorganic phosphate	6.4e-6 [*] / 0.03 ^b

Table 1

Flux ID	Reaction	Enzyme
v ₁	$Ph + GLY \rightarrow G1P$	Glycogen phosphorylase
v ₂	$G1P \rightarrow G6P$	Phosphoglucomutase
v ₃	$G6P \rightarrow F6P$	Glucose-6-phosphate isomerase
v ₄	$F6P + ATP \rightarrow FDP + ADP$	Phosphofructokinase-1
v ₅	$FDP \leftrightarrow DHAP + GAP$	Fructose-bisphosphate aldolase
v ₆	$GAP \leftrightarrow DHAP$	Triose-phosphate isomerase
v ₇	$GAP + NAD + Ph \leftrightarrow DPG + NADH$	Glyceraldehyde-3-phosphate dehydrogenase
v ₈	$DPG + ADP \leftrightarrow P3G + ATP$	Phosphoglycerate kinase
v ₉	$P3G \leftrightarrow P2G$	Phosphoglycerate mutase
v ₁₀	$P2G \leftrightarrow PEP$	Phosphopyruvate hydratase
v ₁₁	$PEP + ADP \rightarrow PYR + ATP$	Pyruvate kinase
v ₁₂	$PYR + NADH \leftrightarrow LAC + NAD$	Lactate dehydrogenase

5

Table 2

For the M-M kinetic study, we utilized the SBML model of glycolysis in the muscle from the Biomodels repository (ID MODEL6623617994). Correctness of the model was checked using the JWS Online server. The model was imported into the Python v.3.6.9 programming environment, using the library libsbml v.5.19. Numerical integration was performed using the Scipy v.0.19 ODEINT solver.

The initial and expected end-point metabolite concentrations in the muscle at rest and under heavy load were taken from Table 3 and 4 of the original paper, with the exception

15

of lactate concentration. The original model assumed identical concentration of lactate in the muscle at resat and under load, equal 1.3 mMol/L. This does not correspond to observed accumulation of lactate in the muscle under load. Therefore, we set the lactate concentration at load to 6 mMol/L. To check the correctness of the custom Python code, the results of a 60 minutes simulation were compared and found reproducible with those generated by the JWS Online server under the same model conditions. Most importantly, we fully reproduced the results of when using their original parameterization. Further, a series of simulations with modified values of the V_{gly} parameter in the range between 10 to 100 % of its initial value was produced, 200 min of simulated time each.

10

The input for the MFA included the stoichiometry matrix (Table 3, which shows stoichiometry matrix of the muscle glycolysis model), the minimal metabolite concentrations matrix (Table 4, which show simulation results of glycolysis in the muscle) and the flux limits matrix (Table 5, which shows Flux constraints matrix used in the MFA simulation of glycolysis in the muscle). Metabolite concentrations X were observed at the state of the muscle at rest (here: state A) and under load (here: state B). The extent variable ξ was declared to represent the change in the muscle from the rest state to loaded state observed as a change in the pyruvate concentration. The solution for the fluxes was calculated with Equation (4).

20

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}
DHAP	0	0	0	0	1	1	0	0	0	0	0	0
DPG	0	0	0	0	0	0	1	-1	0	0	0	0
FSP	0	0	1	-1	0	0	0	0	0	0	0	0
FDP	0	0	0	1	-1	0	0	0	0	0	0	0
G1P	1	-1	0	0	0	0	0	0	0	0	0	0
G6P	0	1	-1	0	0	0	0	0	0	0	0	0
GAP	0	0	0	0	1	-1	-1	0	0	0	0	0
NAD	0	0	0	0	0	0	-1	0	0	0	0	1
NADH	0	0	0	0	0	0	1	0	0	0	0	-1
P2G	0	0	0	0	0	0	0	0	1	-1	0	0
P3G	0	0	0	0	0	0	0	1	-1	0	0	0
PEP	0	0	0	0	0	0	0	0	0	1	-1	0
PYR	0	0	0	0	0	0	0	0	0	0	1	-1

Table 3

Metabolite ID	Concentration at rest , Mol/L	Concentration under load , Mol/L
GLY	0.112	0.112
G1P	1.8e-3	1.7e-3
G6P	2.8e-3	5.6e-3
F6P	1.3e-3	2.3e-3
FDP	6.6e-1	3.5e-2
GAP	5.1e-03	2.9e-4
DHAP	3.5e-02	5.8e-3
DPG	4.1e-6	2.4e-7
P3G	2.3e-4	5.5e-4
P2G	3.9e-5	7.1e-5
PEP	1.9e-5	1.8e-5
PYR	5.5e-6	1.8e-5
LAC	1.3e-3	6.0e-3
NAD	1.2e-4	1.2e-4
NADH	3.8e-4	3.8e-4
ADP	7.8e-6	4.1e-04
ATP	8.2e-3	7.8e-03
PH	6.4e-6	0.03

Table 4

Flux	v_{min}	v_{max}
v_1	---∞	∞
v_2	---∞	∞
v_3	---∞	∞
v_4	---∞	∞
v_5	---∞	∞
v_6	---∞	∞
v_7	---∞	∞
v_8	---∞	∞
v_9	---∞	∞
v_{10}	---∞	∞
v_{11}	---∞	∞
v_{12}	---∞	∞

Table 5

5

The glycolysis example (shown in Figure 7) was aimed at illustrating the key concepts of the extended MFA (the extent variable ξ , its states and generalized metabolic fluxes) providing a better approximation to the experimentally measured metabolic fluxes *in vivo* compared to a M-M kinetic model. The initial state of the system was characterized as a rest condition and the system evolved towards the steady state of the muscle a) at rest or b) at a heavy load.

The steady state in the M-M kinetic model was characterized with the constant glycogen supply, constant depletion of the pyruvate and lactate efflux. The simulated lactate efflux was set to achieve the steady state lactate concentration of 1.3 mMol/L at rest and 6.0 mMol/L under load.

15

In both the rest and the loaded muscle simulations, the system approached the steady state within less than 60 minutes from the initial state reported in the original paper (as shown in Figure 8 and Table 4). Consistent with the conclusions of Lambeth and Kushmerick, the steady state was observed when the fluxes through pyruvate and lactate were maximal (Table 6). Table 6 shows simulation results of human glycolysis in the muscle. The steady state fluxes at rest and under load were computed by the M-M model, The generalized fluxes $v(\xi)$ along the extent ξ quantifying the transition between the states a) at rest and b) under load. The respective ratio of the magnitudes of these fluxes was stoichiometric. For each individual flux, the ratio of its value at load and at rest was equal 43.6. However, this value did not fit the range of flux changes (less than 10-fold) measured experimentally.

Flux ID	Reaction	Flux at rest v^r , Mol/(L · s)	Flux at load v^l , Mol/(L · s)	Generalized flux $v(\xi)$, Mol/(L · Mol _{PYR})
v_1	$PR + GLY \rightarrow GAP$	1.1e-3	0.048	7.7e-3
v_2	$GAP \rightarrow G6P$	1.1e-3	0.048	6.2e-3
v_3	$G6P \rightarrow F6P$	1.1e-3	0.048	3.6e-3
v_4	$F6P + ATP \rightarrow FDP + ADP$	1.1e-3	0.048	2.5e-3
v_5	$FDP \leftrightarrow DHAP + GAP$	1.1e-3	0.048	2.5e-3
v_6	$GAP \leftrightarrow DHAP$	-1.1e-3	-0.048	-2.5e-3
v_7	$GAP + NAD^+ + PR \leftrightarrow DPG + NADH$	2.1e-3	0.095	5.1e-3
v_8	$DPG + ADP \leftrightarrow P2G + ATP$	2.1e-3	0.096	5.1e-3
v_9	$P2G \leftrightarrow P2G$	2.1e-3	0.096	4.7e-3
v_{10}	$P2G \leftrightarrow PEP$	2.1e-3	0.096	4.7e-3
v_{11}	$PEP + ADP \rightarrow PYR + ATP$	2.1e-3	0.096	4.7e-3
v_{12}	$PYR + NADH \leftrightarrow LAC + NAD$	2.1e-3	0.096	4.7e-3

Table 6

To find a better correspondence between the results of modeling and the available experimental data, we applied our extended MFA method to obtain the best fit values for the flux magnitudes in the muscle at rest and under load, taken both as states A and B (in Equation (2)) reached by simulation time $t = 60$. To apply Equation (1), we chose the extent variable ξ to represent the muscle state progression from the resting to the loaded condition. Based on the glycolysis stoichiometry (Table 3), we computed the generalized flux values $v(\xi)$ quantifying the transition between the two states, such that the flux values would best fit the observed differences between the metabolite concentrations (shown in Table 4), while staying within constraints (see Equation (4)). In the present simulation, we set all the fluxes to be potentially reversible (see Table 5).

The results shown in Table 6 indicate that the absolute magnitude of the generalized fluxes quantifying the difference between the rest and the loaded muscle states is within the range between 2.5 to 7.7 mMol/(L·mMol_{PYR}) upon the muscle mode state change, which corresponds to the ranges observed in experiments. The negative flux at triose-phosphate isomerase indicates that dihydroxyacetone phosphate (DHAP) produced from fructose 1,6-

bisphosphate (FDP) is converted to glyceraldehyde 3-phosphate (GAP) by fructose-bisphosphate aldolase.

The positive flux values decrease down the pathway, from glycogen phosphorylase ($v_1 = 7.7 \text{mMol}/(\text{L.mMol})$) to fructose-bisphosphate aldolase ($v_5 = 2.5 \text{mMol}/(\text{L.mMol})$), which may suggest that a progressive increase in intensity of exercise should increase the demand of upstream metabolites. While it is confirmed by the fact that muscle glycogen is the primary metabolite depleted during exercise, a recent study provided details indicating an increase of G6P, F6P, and P2G when the muscle changes the state from rest to load, which are the products of excess fluxes reported here.

Thus, in the case of muscle glycogenolysis, application of our extended MFA methodology allowed us to correctly identify and characterize underlying molecular and physiological processes, while quantifying them within the same order of magnitude as in experimental studies. Unlike the M-M reaction kinetics, the extended MFA framework does not assume the system states to be steady.

At the same time, the underlying assumption of this methodology is that each term in Equation (1) (and its equivalent formulation in terms of time as the progress variable) is considered linear. To test the validity of the linearity assumption, we simulated a series of perturbations to the system, the results of which are presented in Figure 10. Effectively, we ran 100 independent simulations, each of which demonstrated a perturbed system after reaching the steady state. Each perturbation was a small change in the value of the parameter V_{gly} , which determines the turnover rate of hexokinase. The effect of the perturbation was quantified in the extent of pyruvate production $\xi(\text{PYR})$. The observation that each perturbation has a representation as a single smooth path was in accordance with Equations (1) and (2). In the range of extent values $\xi > 0.1\max(\xi)$, the system's progress path demonstrates linearity for every metabolite.

Next, to demonstrate that our analysis extends to the cases far from steady state conditions, we carried out a similar analysis on the time scale of 0.5 minutes. Despite the fact that the system had not reached the steady state with respect of most of the parameters (Figure 8), the respective representation of the simulated system in the progress extent coordinate was very similar (Figure 9). The differences were observed in relatively small local deviations from the linear progression paths at around 40% and 90% of its maximal value. These deviations did not exceed 3% for any of the metabolites and thus the linearity of the progression line was overall observed, thus suggesting that the

extended MFA is applicable to the simulated glycolysis system in its initial state, and is also robust against perturbations of kinetic parameters.

Figure 10 shows the progress extent profile of the glycolysis simulation under perturbation, away from the steady state. The hexokinase rate perturbations were selected as the process defining the changes in the system state. The perturbations simulated by varying the turnover rate V_{gly} in the range between 10 to 100% with step of 0.01%, each step in a separate simulation. Each simulation ran for 0.5 minutes of the system's time to exclude convergence to the steady state. Pyruvate concentration PYR was selected as the variable quantifying the progress extent ξ . The values of the metabolite concentrations and ξ are shown normalized by their respective maximal values. For each metabolite the plot demonstrates a single continuous path, which is also linear for $\xi > 10\%$.

The present disclosure now illustrates a second example regarding computational modeling and analysis of clinical data from diabetes patients.

To evaluate the applicability of the proposed methods to clinical data, we studied a multi-ethnic cohort of 163 patients with type 2 diabetes visiting a tertiary centre in Singapore. We examined data with regards to the development of complications of diabetic retinopathy over a period of up to 3 years after enrolment into the cohort study.

Male and female participants were represented in equal proportion (Table 7, which shows baseline clinical and biochemical characteristics of the observational cohort study). In the cohort, the median age was 56. Median duration of type 2 diabetes, hyperlipidemia and hypertension was 10, 7 and 6.5 years, respectively. Cataracts were detected in 68 patients (41.7 %), while 52 patients (31.9 %) developed diabetic retinopathy.

Variable	Measurement
Males (%)	83(50)
Ethnicity: n(%)	
Chinese	103(62.55)
Malays	24(14.46)
Indians	39(23.89)
Age years, Mean(SD)	56.34(10.07)
T2DM duration years, Median(IQR)	10(4-17)
Hypertension duration years, Median(IQR)	6.5(0-13)
Hyperlipidemia duration years, Median (IQR)	7(2-13)
BMI kg/cm ² , Mean(SD)	27.27(4.05)
Systolic BP, mm Hg, Mean(SD)	134.27(14.86)
Diastolic BP, mm Hg, Mean(SD)	70.04(8.35)
Fasting glucose mmol/L, mean(SD)	8.03(2.29)
HbA1c %, mean(SD)	7.93(1.17)
Total Cholesterol mmol/L, Mean(SD)	4.28(0.92)
HDL-Cholesterol mmol/L, Mean(SD)	1.12(0.38)
LDL-Cholesterol mmol/L, Mean(SD)	2.59(0.78)
Triglycerides mmol/L, Mean(SD)	1.04(1.32)
Urea nitrogen mmol/L, Mean(SD)	7.50(2.08)
Ferritin mmol/L, Mean(SD)	190.62(88.15)
Transferrin mmol/L, Mean(SD)	2.46(0.38)

Table 7

The summary biochemical characteristics of the patients are provided in Table 7. Similar to the earlier study, vascular functions of the patients were assessed. Table 8 shows ophthalmic complications. Table 9 provides the summary regarding vascular function measurement. The patients were characterized by measuring the concentrations of C-reactive protein (CRP), reactive oxygen molecules (ROM) and oxidized LDL (ox-LDL). Arterial stiffness was quantified by the pulse wave velocity (PWV), and the endothelial function was assessed with reactive hyperemia index (RHI). Table 10 shows a list of exchange generalized fluxes representing dependencies between the metabolic fluxes and physiological variables.

Variable	Measurement
Cataracts, n (%)	68 (41.7)
Diabetic retinopathy, n (%)	52 (31.9)

Table 8

Variable	Measurement
CRP, mg/L Median (IQR)	1.45(0.7-3.8)
ROM, Median(IQR)	271(241-313)
MAP, μ M Median(IQR)	2225(2051-2388)
Ox-LDL, IU/L Mean(Sd)	55.93(21.85)
CIMT, mm Mean(SD)	0.55(0.13)
LnRHI Mean(SD)	0.67(0.35)
Pulse Wave Velocity m/cm	8.35(1.74)

Table 9

Flux	Pathway	Description
<i>Food</i> → <i>Glucose</i>	Digestion	Conversion of simple carbohydrates ingested from food to glucose in the blood
<i>Food</i> → <i>Protein</i>	Digestion	Conversion of proteins ingested from food to protein in the blood
<i>Food</i> → <i>FFA</i>	Digestion	Conversion of fats ingested from food to free fatty acids in the blood
<i>Food</i> → <i>LDL</i>	Digestion	Conversion of fats ingested from food to low-density lipoproteins in the blood
<i>Liver</i> → <i>Liver</i> → <i>FA</i>	Lipid metabolism	Production of fatty acids in the liver
<i>Bileacid</i> → <i>Bile</i>	Lipid metabolism, heme metabolism	Bile production
<i>Respiration</i> → <i>Erythrocytes</i>	Respiration and oxidation	Adaptive erythrocyte production
<i>Respiration</i> → <i>Bicarbonate</i>	Respiration and oxidation	Serum bicarbonate (dissolved carbon dioxide) production by the tissues
<i>Endothelium</i> → <i>NO</i>	Vasodilation and vasoconstriction	Nitric oxide production in the endothelium
<i>Liver</i> → <i>Liver</i> → <i>cholesterol</i>	Lipid metabolism	Production of cholesterol in the liver
<i>RMT</i> → <i>CIMT</i> → <i>avg</i> → <i>R</i>	Endothelial process	Carotid intima-media thickness increase driven by the endothelium
<i>Ox</i> → <i>LDL</i> → <i>CIMT</i> → <i>avg</i> → <i>R</i>	Endothelial process	Carotid intima-media thickness increase driven by lipid oxidation
<i>Ox</i> → <i>LDL</i> → <i>Pulse</i> → <i>wave</i> → <i>velocity</i>	Endothelial process	Pulse wave velocity increase driven by lipid oxidation-driven arterial stiffness
<i>RAP</i> → <i>Deactivation</i>	Respiration and oxidation	Deactivation of reactive oxygen molecules by antioxidants
<i>ROM</i> → <i>Deactivation</i>	Respiration and oxidation	Deactivation of reactive oxygen molecules by antioxidants
<i>Inflammation</i> → <i>Ferritin</i>	Inflammation	Increase of ferritin level in inflammation

Table 10

A metabolic map was designed to include the metabolites measured in the study, to quantify the fluxes through the major biochemical and physiological pathways implicated in diabetes (Figure 11, Table 11). In Figure 11, the map is a graphical representation of the extended stoichiometry matrix in Table 11, which shows the extended stoichiometry matrix representing metabolic and physiological pathways driving diabetes complications. Nodes represent metabolites and physiological parameters. Edges represent generalized fluxes. The fluxes statistically associated ($P < 0.01$ by the Wilcoxon-Mann-Whitney test) with proliferative retinopathy and cataract are highlighted with red and brown, respectively.

For each patient a separate model was constructed and initiated with all available parameter values measured in the patient's blood sample. The initial values of the starting parameters were imputed as population averages (Table 12). Metabolite concentrations were constrained within $\pm 20\%$ of the patient's initial values and within the physiological ranges for every unavailable parameter, obtained from reference literature. Fluxes were constrained with reversibility conditions, based on reference literature. For each patient

the best fit flux and metabolite concentration vectors were obtained by the quadratic programming procedure (Equation (4)).

[illegible]

Table 11

In particular, Table 11 shows the extended stoichiometry matrix representing metabolic and physiological pathways driving diabetes complications. Table 12 shows literature-based values in the states of non-diabetic (state A) and late diabetic (state B) states of diabetes progression.

5

Metabolite / phys. variable	Non-diabetic	Diabetic	Maximal
TC	0.2	1.2	6
HDL	1.1	1.5	4
VLDL	0.1	0.5	0.8
LDL	1.1	3	6
Cv-LDL	0.1	50	200
Pulse-wave-velocity	2	8.3	30
lnRHR	0.01	0.66	3
CMT-avg-L	0.2	0.62	2
CMT-avg-R	0.2	0.62	2
ROM	1	200	1000
BAP	1	2000	5000
nat.RP	0	1.8	10
FFA	0.7	1.2	1.5
Fat-lipids	0.5	2.1	2.52
Liver-FA	0.2	0.61	0.732
Liver-cholesterol	0.2	0.458	0.551
Protein	0.9	1.09	1.308
Glucose	3.8	6	7.2
HbA1C	1	5	20
Bilirubin	0.08	0.8	0.98
Creat	0.002	0.07	0.1
Ur-LpH	4	6	9
Ur-prost	0	5	20
Ur-Glc	0	0.5	10
Ur-ket	0	0.2	10
BMI	18	18	40
Haptoglobin	3	200	1000
Hb	1	20	100
Ferritin	1	300	3000
Erythrocytes	1	50	100
Endothelium	1	50	100
NO	23	24	25
Haematocrit	1	40	90
Bicarbonate	10	23	40
Iron	0	0.68	0.5
Ferritin-Iron	1	300	2000
ESR	0	5	60
Liver	0	1	100
ALT	0	35	100

Table 12

Each component of the metabolite and flux vectors was then evaluated as an independent predictor of the patient classification by one of the following phenotypes representing diabetic complications detected by the sample collections time point: proliferative retinopathy and cataract. Statistical associations between each vector component and each phenotype was assessed by using the Wilcoxon-Mann-Whitney non-parametric test with the null hypothesis of the vector components being equal between two groups of patients corresponding to two phenotypic (disease) states. Multiple hypothesis testing bias was controlled via the false discovery rate assessment. The false discovery rate was

calculated according to the Benjamini-Yekutieli procedure, and the P-value with FDR not exceeding 15% were reported.

- Figure 12 shows the results of using the proposed methodology to predict development of retinopathy (a) and nephropathy (b) in type 2 diabetes patients within 3 years from the time point of the patient data input. Biomarkers used as inputs were: BMI, HbA1C, RG, TC, TG, HDL, LDL, Creat, ACR, Ox-LDL, PWV, lnRHI, CIMT-avg, hsCRP, ROM, BAP, Haemoglobin, Haptoglobin, Ferritin.
- Figure 13 shows the results of using the proposed methodology to predict development of coronary artery disease in type 2 diabetes patients within 3 years from the time point of the patient data input. Biomarkers used as inputs: BMI, HbA1C, FG, TC, TG, HDL, LDL, ACR, Haemoglobin, Haematocrit.
- The present disclosure now discusses application of the extended MFA framework to diabetes patients data reveals metabolic and physiological mechanisms associated with diabetic retinopathy progression.

The glycolysis model analyzed in the first example represented the case of an isolated pathway, where all key metabolite concentrations were available for the model construction and the extent ξ has a clear interpretation as a metabolite production process. However, real-world clinical scenarios provide additional challenges. In patients, changes in the entire metabolism are quantified by measuring a sparse set of parameters, only a fraction of which are metabolite concentrations, while the rest of quantitative information is obtained from measuring physiological parameters, such as blood pressure or blood vessel calcification scores. At the same time, the state of a patient's health is characterized by a set of symptoms and diagnoses made by the doctor after a holistic assessment of all available observables.

Here, we analyzed type 2 diabetes patients to demonstrate how the above challenges can be addressed in the extended MFA framework. We selected diabetes progression as the health state progress variable ξ . The initial point of diabetes type 2 progression (state A) was selected to represent a healthy individual with demographic characteristics similar to the studied population (see Table 12). The end-point state (state B) represented advanced type 2 diabetes, where common complications, such as hypertension and nephropathy are fully manifest. Each state was characterized by measured values of key metabolite concentrations and quantitative physiological parameters of vascular health (Tables 7, 9 and 12). On the progress scale ξ stretching between points A and B, we considered an

individual patient's health history as a smooth change of metabolite concentrations and physiological characteristics marked by developing diabetic complications.

To obtain the extended MFA model of health state evolution, we started with constructing the stoichiometry matrix from the empirically defined graph, where nodes represented the measured metabolites and the edges were the generalized fluxes between them. Then, we extended the stoichiometric matrix by integrating the measured physiological parameters in it. Below we illustrate this approach by finding a stoichiometric coefficient connecting the metabolic variable oxidized low density lipoprotein (ox-LDL) with the physiological variable pulse wave velocity (PWV).

A long-term study of patients with rheumatoid arthritis provided information on ox-LDL and PWV changes as a result of 6 weeks treatment with a daily 20 mg simvastatin and 10 mg ezetimibe. The study reports a decrease of PWV by 0.71 m/s and ox-LDL by 19.7 U/L as a result of simvastatin treatment. With this initial information, we can use Equation (L) to estimate the coefficients N_{ik} and N_{jk} , where i and j denote the indices of PWV and ox-LDL, respectively, k denotes the index of the flux v_k from ox-LDL to PWV and points A and B are taken as the start and the end of the treatment period. We estimate the average change per week of treatment in both parameters as $\Delta_{AB}x_i = -0.118$ and $\Delta_{AB}x_j = -3.283$.

Assuming that v_k is an exchange flux, we can set $N_{jk} = 1$ to obtain $N_{ik} = 0.006$ per week of treatment (i.e. per unit of the extent variable). Importantly, when ezetimibe is used as a treatment, applying the same method would provide a different value for the coefficient, i.e. $N_{ik} = 0.02$. This is not surprising, since the estimates depend on the choice of the extent variables ξ . Indeed, two drugs have different mechanisms of action: simvastatin inhibits endogenous cholesterol production, while ezetimibe inhibits absorption of cholesterol in the small intestine. Each mechanism results in a different direction and extent magnitude of metabolic changes.

In an independent study of diabetes type 2 patients the authors estimated the effect of a 4-weeks long atorvastatin treatment on ox-LDL. The value re-scaled per week is estimated as $\Delta x_i = -3.372$, which is close to the value of -3.283 obtained in the equivalent dose simvastatin study.

In a separate study, effects of atorvastatin therapy on PWV were evaluated in patients with hypertension and hypercholesterolemia in the duration of 26 weeks. The reported change in PWV re-scaled per week produces $\Delta x_i = -0.065$, which is lower than the value

obtained in the simvastatin study (-0.118). The difference may be due to different basal level of arterial stiffness in patients with hypertension and hypercholesterolemia compared to the patients with rheumatoid arthritis.

- 5 Population data incorporates numerous sources of variation, such as patient's medical history, age, gender and ethnicity. In addition, time, methods and conditions of metabolite concentrations measurements may be dissimilar across patients. Moreover, these factors may differ for the same patient across multiple data collection points. Unlike experimental studies, clinical data often contains missing values and even errors in the records.
- 10 For each patient standard biochemical measurements were done at two time points, the baseline and the follow-up, separated by 178 ± 13.6 days. The clinical characteristics of the patients are summarized in Tables 7, 9 and 8.
- 15 Using the methodology presented here, we built two personalized flux profiles per patient: at the baseline and the follow-up time points. We aimed to find whether at the baseline time point information provided by fluxes can correlate with observed clinical characteristics.
- 20 As the first step of the analysis, in a cross-sectional study we evaluated the capabilities of the fluxes to describe patient phenotypes. We considered ophthalmic complications of diabetes an interesting group of phenotypes to explore metabolic and physiological pathways, using the proposed methodology. The results are presented in Tables 14 and 15 (see also Figure 11).
- 25 Proliferative retinopathy was one of the ophthalmic complications for which the extended MFA was more informative than direct observations of metabolite concentrations. Pulse wave velocity (PWV) and reactive hyperemia index (RHI) were significantly increased in diabetes patients with proliferative retinopathy. Similar observations were reported earlier.
- 30 Another flux of significance was conversion of high-density lipoprotein cholesterol (HDL) into low-density lipoproteins (LDL) localized in the liver ($P=9.7e-6$). Unlike many other tissues producing cholesterol locally, for e.g. the retina, the cholesterol produced by the liver is transported via the bloodstream. LDL and the HDL/LDL ratios are known as significant factors of retinopathy progression. The flux quantifying the effect of oxidized LDL (ox-LDL) on PWV was also significant ($P=2.8e-3$). LDL oxidation and lipid oxidation in general are important mediators implicated in retinopathy. Iron and ferritin play an
- 35

important role in oxidation reactions affecting diabetic retinopathy progression, in particular, by producing ox-LDL, and our results support that ($P=1.9e-4$).

5 Notably, we also found that the ferritin and bilirubin-bile fluxes were also associated with cataract development ($P=3.5e-3$ and $P=2.3e-3$, respectively), which was not not detected on the level of individual metabolites. Recently, an evidence was found that blood bilirubin might be a compound protecting retina from degradation in diabetes patients.

10 We found that the rate of protein consumption and protein-dependent decrease of urine pH are the parameter significantly associated with proliferative retinopathy ($P=2.4e-5$). Studies report equivocal effects of protein consumption on diabetic retinopathy development. At the same time, there is a consensus with respect to the role of high protein consumption in diabetic microvascular changes, which can also be observed in associations with diabetic nephropathy and has been reacted in clinical recommendations.

15 Urine pH is considered an independent negative prognostic and progression indicator of type 2 diabetes. At the same time, ammonium ions concentration are considered a factor significantly affecting urine pH of diabetes patients.

20 Thus we observed that with respect to specific clinical phenotypes, statistical results obtained with metabolic and physiological flux models are not contradicting the results obtained with metabolic and physiological variables alone. Moreover, flux models in certain cases have the potential to provide more biomarkers characterizing the disease and can improve statistical power. The ability of flux models to describe additional details of long-term dynamics is complementary to the descriptive power of metabolites alone. All these

25 conclusions confirm the applicability of the computational framework provided here to address the practical needs of integrative biochemical, physiological and clinical data analysis.

30 It will be appreciated that many further modifications and permutations of various aspects of the described embodiments are possible. Accordingly, the described aspects are intended to embrace all such alterations, modifications, and variations that fall within the spirit and scope of the appended claims.

35 Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

5

CLAIMS

1. A method of generating a metabolic digital twin of a subject for clinical decision support in relation to a medical condition, the method comprising:
- 5 receiving data indicative of an extended metabolic map comprising nodes representing a plurality of metabolites and one or more non-metabolic parameters, and edges representing relationships between them;
- determining an extended stoichiometry matrix at least partly from the extended metabolic map, wherein coefficients in the extended stoichiometry matrix define quantitative relationships between abundances of the metabolites and/or
- 10 values of the non-metabolic parameters;
- receiving data indicative of measurements of a sample of the subject, wherein the measurements comprise measurements of one or more of available metabolite concentrations for one or more of the plurality of metabolites, and
- 15 measurements of one or more of the non-metabolic parameters; and
- optimizing, subject to one or more constraints, an objective function that depends on a product of the stoichiometry matrix and a flux vector, each component of the flux vector corresponding to an edge of the extended metabolic map, wherein the one or more constraints are based on the measurements;
- 20 wherein the metabolic digital twin comprises data indicative of a best-fit flux vector obtained from said optimizing, and wherein components of the best-fit flux vector are indicative of metabolite-metabolite fluxes and metabolite-physiological fluxes for the subject.

- 25 2. A method according to claim 1, wherein said optimizing is quadratic optimization.

3. A method according to claim 2, wherein said optimizing is performed via:

$$\left\{ \begin{array}{l} v = \operatorname{argmin}(v^T S^T S v - S^T \Delta X) \\ Sv \leq (1 + \epsilon) \Delta X \\ Sv > (1 - \epsilon) \Delta X \\ v \leq v_{\max} \\ v > v_{\min} \\ \epsilon = 0.2 \end{array} \right.$$

- 30 where S is the extended stoichiometry matrix, v is the flux vector, and ΔX is a difference between the measurements and a reference set of measurements, wherein the subject has a first state of progression of the medical condition, and the reference set of measurements is obtained for a second state of progression of the medical condition that is different from the first state.

4. A method according to claim 3, wherein the reference set of measurements is obtained from another subject or from a population of subjects characterised by the second state of progression.
5. A method according to any one of claims 1 to 4, wherein for any unmeasured metabolites or non-metabolic parameters, initial values of the corresponding components of the flux vector are based on respective population averages.
6. A method according to any one of claims 1 to 5, wherein the metabolic map is a simplified metabolic map in which at least one node represents a plurality of metabolites.
7. A method according to any one of claims 1 to 6, wherein said outputting causes display of the best-fit flux vector relative to a reference curve that is indicative of evolution of the best-fit flux vector as a function of an extent variable that quantifies evolution of the medical condition from the first state to the second state.
8. A method according to any one of claims 1 to 7, comprising outputting a comparison of one or more components of the best-fit flux vector to reference values of the one or more components obtained from a reference population of subjects.
9. A method according to any one of claims 1 to 8, wherein the medical condition comprises one or more diabetic complications.
10. A method according to claim 9, wherein the one or more medical conditions comprise one or more ophthalmic complications and/or one or more cardiovascular complications.
11. A method according to claim 10, wherein the one or more medical conditions comprise diabetic retinopathy, diabetic neuropathy, and/or coronary artery disease.
12. A method for identifying one or more biomarkers associated with a disease phenotype, comprising:
 - obtaining, for a plurality of subjects each having one or more status indicators corresponding to one or more respective disease phenotypes, data indicative of a plurality of respective metabolic digital twins, wherein each metabolic digital twin is generated by a method according to any one of claims 1 to 11; and

determining an indication of statistically significant association of respective components of the flux vector with the one or more status indicators;

wherein the one or more biomarkers comprise one or more components of the flux vector having statistically significant association with a corresponding disease phenotype.

- 5
13. A method for identifying metabolic flux patterns associated with a disease phenotype, comprising:
- 10 obtaining, for a plurality of subjects each having one or more status indicators corresponding to one or more respective disease phenotypes, population data indicative of respective flux vectors of a plurality of respective metabolic digital twins, wherein each metabolic digital twin is generated by a method according to any one of claims 1 to 11;
- 15 clustering the population data to generate a plurality of clusters;
- performing one or more pair-wise tests using respective first and second clusters of said plurality of clusters, and the one or more status indicators; and
- for each statistically significant test of said one or more pair-wise tests, a difference between a median flux vector of subjects in the respective first cluster and a median flux vector of subjects in the respective second cluster.
- 20
14. A method of predicting progression of a medical condition in a subject, comprising:
- generating a metabolic digital twin of the subject by a method according to any one of claims 1 to 11; and
- 25 outputting a comparison of one or more components of the best-fit flux vector for the metabolic digital twin to one or more corresponding components of flux vectors of metabolic digital twins for a reference population of subjects;
- wherein said comparison is indicative of progression of the medical condition in the subject relative to the reference population.
- 30 15. A method according to claim 14, wherein the metabolic digital twins for the reference population are generated by a method according to any one of claims 1 to 11.
16. A system for generating a metabolic digital twin of a subject for clinical decision support in relation to a medical condition, the system comprising:
- 35 storage; and
- at least one processor in communication with the storage;

wherein the storage comprises machine-readable instructions for causing the at least one processor to carry out a method according to any one of claims 1 to 11.

- 5 17. A system for identifying one or more biomarkers associated with a disease phenotype, comprising:
 storage; and
 at least one processor in communication with the storage;
 wherein the storage comprises machine-readable instructions for causing
10 the at least one processor to carry out a method according to claim 12.
18. A system for identifying metabolic flux patterns associated with a disease phenotype, comprising:
 storage; and
15 at least one processor in communication with the storage;
 wherein the storage comprises machine-readable instructions for causing the at least one processor to carry out a method according to claim 13.
19. A system for predicting progression of a medical condition in a subject, comprising:
20 storage; and
 at least one processor in communication with the storage;
 wherein the storage comprises machine-readable instructions for causing the at least one processor to carry out a method according to claim 14.
- 25 20. Non-transitory computer-readable storage having stored thereon instructions for causing at least one processor to carry out a method according to any one of claims 1-14.