



(51) International Patent Classification:

G16B 40/20 (2019.01) G06N 20/10 (2019.01)
G16C 20/70 (2019.01)

(21) International Application Number:

PCT/SG2019/050314

(22) International Filing Date:

24 June 2019 (24.06.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10201805452S 25 June 2018 (25.06.2018) SG

(71) Applicant: **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 1 Fusionopolis Way, #20-10, Connexis North Tower, Singapore 138632 (SG).

(72) Inventor: **LOO, Lit Hsin**; c/o Bioinformatics Institute, 30 Biopolis Street, #07-01 Matrix, Singapore 138671 (SG).

(74) Agent: **DONALDSON & BURKINSHAW LLP**; 24 Raffles Place, #15-00 Clifford Centre, Singapore 048621 (SG).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(54) Title: A METHOD AND SYSTEM FOR DETERMINING POINT OF DEPARTURE

(57) Abstract: A method of determining a point-of-departure (POD) of a chemical and a system for determining a POD of a chemical may be provided, the method comprising (i) measuring one or more features obtained from images of one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type; (ii) measuring one or more features obtained from images of one or more control samples having been treated with a solvent control at one or a plurality of control concentrations and being obtained from a control population of the first cell type; (iii) classifying each sample of the one or more test samples at each of the plurality of test concentrations and the one or more control samples at the one or each of the plurality of control concentrations based on measured feature values; (iv) formulating a first concentration-accuracy curve based on the step of classifying each sample; and (v) deriving a first initial POD value of the chemical based on the first concentration-accuracy curve.

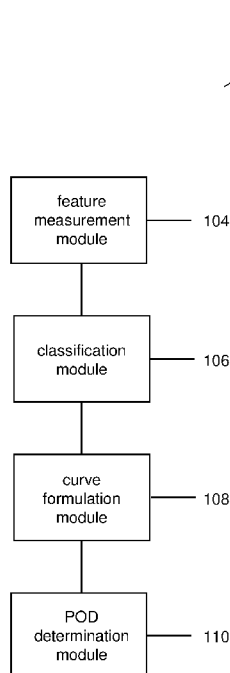


FIG. 1



A METHOD AND SYSTEM FOR DETERMINING POINT OF DEPARTURE

5

TECHNICAL FIELD

The present disclosure relates broadly to a method for determining a point of departure (POD) of a chemical and to a system for determining a point of departure (POD) of a chemical.

10

BACKGROUND

Chemical risk assessment is a process to determine and quantify the risk resulting from an exposure to a chemical, including identifying a relationship between the exposure of the chemical and its possible effects to human health. Point of departure (POD) is typically a point on a dose-response curve obtained from experimental data generally corresponding to an estimated concentration in which low or no adverse effect is observed. Therefore, POD is adverse-effect dependent, and a chemical may have multiple different PODs for different possible adverse effects or target different tissues/organs. PODs typically mark the beginning of an extrapolation to the Reference Dose (RfD) which is a standard used by regulatory agencies to regulate the acceptable exposure level of a chemical in the human population. RfD is usually lower than PODs to account for the uncertainty in the extrapolation process.

There are several quantitative definitions of POD, including the No Observable Adverse Effect Level (NOAEL), Lowest Observable Adverse Effect Level (LOAEL), Benchmark Dose (BMD), etc., which refer to different points from experimental data or on a fitted dose-response curve. For example, NOAEL is defined to be “the highest experimentally tested dose without any observable adverse response or effect”, and BMD is defined to be “the statistical lower limit on the dose corresponding to a specific increase in the response level on a fitted dose-response curve”. As an example, BMD10 is the benchmark dose responsible for a 10% increase in response. Depending on the experimental design and

outcomes, these different definitions may provide different values of PODs even on the same set of experimental data.

Traditionally, PODs are estimated using animal models (i.e. not humans). An adverse effect may be defined as a biochemical, functional, or structural change in specific tissues or organs of the animal models that may lead to detrimental effects on the growth, development, or life span of the animal models. Thus, PODs are typically estimated based on multiple *in vivo* phenotypic readouts, including pathological lesions (such as damages in specific internal organs or cell types) and/or physical changes (such as changes in the body weight) in the animals after chemical exposure.

It has been recognised that traditional phenotypic readouts from animal models may not be predictive of human responses. For example, a survey of 142 drugs approved between 2001 and 2010 found that only 19% of the pulmonary adverse drug reactions identified post-marketing could have been predicted based on pre-clinical animal studies. Further, even closely related animal species may have discrepancies in their responses. For example, a survey found that there is no concordance between mouse and rat non-carcinogenic lung lesions observed in acute and long-term rodent studies of 37 chemicals.

Furthermore, it has been recognised that animal studies may be expensive (e.g. about US\$500,000 to \$1 million per chemical) and relatively slow (e.g. about 1 to 2 years) to perform. The effects of these limitations are significant, given that there are more than 75,000 chemical species currently being manufactured or used in daily life, with additional novel chemicals being designed and synthesized regularly. It has been recognised that a significant portion of these chemicals do not have human safety information.

In view of the above, there is thus a need for a method of estimating/determining a POD of a chemical and a system for estimating/determining a POD of a chemical that seek to address or at least ameliorate one or more of the above-identified problems.

SUMMARY

In accordance with an aspect of the present disclosure, there is provided a system for determining a POD of a chemical, the system comprising: a feature measurement module arranged to measure one or more features obtained from images of (a) one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type, and (b) one or more control samples having been treated with a solvent control at one or a plurality of control concentrations and being obtained from a control population of the first cell type; a classification module arranged to classify each sample of the one or more test samples at each of the plurality of test concentrations and the one or more control samples at the one or each of the plurality of control concentrations based on feature values generated from the feature measurement module; a curve formulation module arranged to formulate a first concentration-accuracy curve based on a classification result of the classification module; and a POD determination module arranged to derive a first initial POD value of the chemical based on the first concentration-accuracy curve.

The classification module may be configured to apply a supervised classifier to classify cells of the one or more test samples treated at each of the plurality of test concentrations and the one or more control samples treated at the one or each of the plurality of control concentrations into an assigned "treated" class or an assigned "control" class based on the one or more features measured from the feature measurement module.

The classification module may be further arranged to provide a classification accuracy value based on counting a number of test samples (P) and a number of control sample (N); counting a number of correctly classified test samples (TP) and a number of correctly classified control samples (TN); and setting a balanced accuracy value based on $(TP/P + TN/N)/2$.

The supervised classifier may comprise a Support Vector Machine.

The curve formulation module may be configured to fit a plurality of statistical models to a first classification result of the classification module and to determine a preferred fit to obtain the first concentration-accuracy curve.

- 5 The plurality of statistical models may comprise at least one of a log-logistic model, a logistic model, a Weibull model, a gamma model, and a constant model.

 The POD determination module may be arranged to label the chemical as “inactive” if the preferred fit is based on a constant model and to label the chemical as
10 “active” if the preferred fit is not a constant model.

 The feature measurement module may be arranged to measure one or more features obtained from images of (a) one or more test samples having been treated with the chemical at the plurality of test concentrations and being obtained from a second test
15 population of the first cell type, and (b) one or more control samples having been treated with the solvent control at the one or plurality of control concentrations and being obtained from a second control population of the first cell type. The classification module may be arranged to classify each sample of the one or more test samples from the second test population at each of the plurality of test concentrations and the one or more control
20 samples from the second control population at the one or each of the plurality of control concentrations. The curve formulation module may be arranged to formulate a second concentration-accuracy curve based on a second classification result of the classification module. The POD determination module may be arranged to derive a second initial POD value of the chemical based on the second concentration-accuracy curve and to obtain
25 an average initial POD value based on the second initial POD value.

 The feature measurement module may be arranged to measure one or more features obtained from images of (a) one or more test samples having been treated with the chemical at the plurality of test concentrations and being obtained from a test
30 population of a different cell type, and (b) one or more control samples having been treated with the solvent control at the one or plurality of control concentrations and being obtained from a control population of the different cell type. The classification module may be arranged to classify each sample of the one or more test samples from the test

population of a different cell type at each of the plurality of test concentrations and the one or more control samples from the control population of the different cell type at the one or each of the plurality of control concentrations. The curve formulation module may be arranged to formulate a different concentration-accuracy curve that is associated with said different cell type based on a third classification result of the classification module. The POD determination module may be arranged to derive a different initial POD value that is associated with said different cell type based on the different concentration-accuracy curve.

The POD determination module may be configured to determine the POD of the chemical based on the different initial POD value.

The first cell type may comprise at least one of a lung cell, a lung-derived cell, a lung-derived cell line cell, a liver cell, a liver-derived cell, a liver-derived cell line cell, a kidney cell, a kidney-derived cell or a kidney-derived cell line cell.

In accordance with an aspect of the present disclosure, there is provided a method of determining a POD of a chemical, the method comprising (i) measuring one or more features obtained from images of one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type; (ii) measuring one or more features obtained from images of one or more control samples having been treated with a solvent control at one or a plurality of control concentrations and being obtained from a control population of the first cell type; (iii) classifying each sample of the one or more test samples at each of the plurality of test concentrations and the one or more control samples at the one or each of the plurality of control concentrations based on measured feature values; (iv) formulating a first concentration-accuracy curve based on the step of classifying each sample; and (v) deriving a first initial POD value of the chemical based on the first concentration-accuracy curve.

Classifying each sample may comprise applying a supervised classifier to classify cells of the one or more test samples treated at each of the plurality of test concentrations and the one or more control samples treated at the one or each of the plurality of control

concentrations into an assigned “treated” class or an assigned “control” class based on the measured features.

The method may further comprise obtaining a classification accuracy value based on: counting a number of test samples (P) and a number of control samples (N); counting a number of correctly classified test samples (TP) and a number of correctly classified control samples (TN); and setting a balanced accuracy value based on $(TP/P + TN/N)/2$.

Formulating a first concentration-accuracy curve may comprise fitting a plurality of statistical models to a first classification result of the classifying step and determining a preferred fit.

The plurality of statistical models may comprise at least one of a log-logistic model, a logistic model, a Weibull model, a gamma model, and a constant model.

The method may further comprise labeling the chemical as “inactive” if the preferred fit is based on a constant model and labeling the chemical as “active” if the preferred fit is not a constant model.

The method may further comprise repeating steps (i) to (iii) for a second test population and for a second control population and formulating at least a second concentration-accuracy curve based on the second test population and the second control population to derive a second initial POD value, and obtaining an average initial POD value based on the second initial POD value.

The method may further comprise repeating steps (i) to (v) for a different cell type to obtain a different initial POD value that is associated with said different cell type.

The method may further comprise determining the POD of the chemical based on the different initial POD value.

The first cell type may comprise at least one of a lung cell, a lung-derived cell, a lung-derived cell line cell, a liver cell, a liver-derived cell, a liver-derived cell line cell, a kidney cell, a kidney-derived cell or a kidney-derived cell line cell.

5 BRIEF DESCRIPTION OF THE DRAWINGS

Exemplary embodiments of the invention will be better understood and readily apparent to one of ordinary skill in the art from the following written description, by way of example only, and in conjunction with the drawings, in which:

10

FIG. 1 is a schematic diagram illustrating a system for determining a POD of a chemical in an exemplary embodiment.

15

FIG. 2 is an illustration of a method/concept utilising a supervised classifier in an exemplary embodiment.

FIG. 3 is a diagram illustrating a method/concept that may be implemented by the system in an exemplary embodiment.

20

FIG. 4 illustrates examples of accuracy-concentration curves and POD values estimated, based on an exemplary embodiment of the system or method, in respect of four chemicals (A) di(2-ethylhexyl) adipate, (B) diethyl phthalate, (C) propylparaben and (D) p,p'-DDT in HepG2 cells.

25

FIG. 5 is a comparison of POD values estimated based on an exemplary embodiment of the system or method against the POD values obtained from ToxCast assays.

30

FIG. 6 is a flowchart of a method for determining a POD of a chemical in an exemplary embodiment.

FIG. 7 is a schematic drawing of a computer system suitable for implementing an exemplary embodiment of the system or method.

DETAILED DESCRIPTION

FIG. 1 is a schematic diagram illustrating a system for determining a POD of a chemical in an exemplary embodiment.

In the exemplary embodiment, one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type may first be imaged using a microscope. One or more control samples having been treated with a solvent control at one or a plurality of control concentrations and being obtained from a control population of the first cell type may also be imaged using a microscope. The control samples may be treated at a fixed single control concentration or at a plurality of control concentrations, depending on desired implementations. The resulting images of one or more test samples and one or more control samples are provided as inputs to a feature measurement module (see below) of the system 100. For example, the images may be two- or three-dimensional matrices with 8-, 16-, or 32-bit integers, or 32- or 64-bit floating point numbers that correspond to the intensity values measured from different x, y, and z coordinates of the test or control samples.

In the exemplary embodiment, the system 100 comprises a feature measurement module 104. In the exemplary embodiment, the feature measurement module 104 is arranged to measure one or more features from the provided images of the test and control samples e.g. to generate feature values. For example, the feature values may be a two-dimensional matrix with 32- or 64-bit floating point numbers that correspond to the values of one or more features (columns of the matrix) quantified from one or more cells of the test or control samples (rows of the matrix).

In the exemplary embodiment, the system 100 further comprises a classification module 106 coupled to the feature measurement module 104. In the exemplary embodiment, the classification module 106 is arranged to classify the one or more test samples and the one or more control samples at each of the plurality of concentrations based on the feature values generated/measured from the feature measurement module 104. For example, each of the one or more test samples at each of the plurality of test concentrations and each of the one or

more control samples at the one or each of the plurality of control concentrations is classified as either being in the “treated” or “control” class. In one example, the performance of the classification module 106 is determined by measuring the accuracy of the classification results at each of the plurality of concentrations. For example, the accuracy value may be determined by counting the numbers of test samples (positives = P) and control samples (negatives = N), and those correctly classified as test samples (true positives = TP) and control samples (true negatives = TN), respectively; and set to a balanced accuracy value e.g. $((TP/P + TN/N)/2)$. In one example, a percentage value is used based on the balanced accuracy value. In one example, the classification may be performed by an unsupervised classifier or by a supervised classifier.

In the exemplary embodiment, the system 100 further comprises a curve formulation module 108 coupled to the classification module 106. In the exemplary embodiment, the curve formulation module 108 is arranged to formulate a first concentration-accuracy curve based on a classification result of the classification module 106.

In the exemplary embodiment, the system 100 further comprises a POD determination module 110 coupled to the curve formulation module 108. In the exemplary embodiment, the POD determination module 110 is arranged to derive a first initial POD value of the chemical based on the first concentration-accuracy curve.

In the exemplary embodiment, the POD determination module 110 may obtain a second initial POD value of the chemical which is based on a second concentration-accuracy curve. The second concentration-accuracy curve itself may be based on a second classification result based on one or more test samples from a second test population of the first cell type e.g. the one or more test samples having been treated at a plurality of test concentrations and one or more control samples from a second control population of the first cell type e.g. the one or more control samples having been treated at the one or a plurality of control concentrations. An average POD value may be obtained based on these initial POD values of the same cell type.

Using a different cell type, the POD determination module 110 may also obtain one or more different initial POD values. Based on the one or more different initial POD values,

another average POD value associated with the different cell type may be obtained. A final/singular POD of the chemical may be determined based on the different average POD values, e.g. obtained for the first and the different cell types.

5 FIG. 2 is an illustration of a method/concept 200 utilising a supervised classifier 206 that may be implemented at the classification module 106 of the system 100 described with reference to FIG. 1 in an exemplary embodiment. In the exemplary embodiment, the classification module 106 is configured to apply a supervised classifier 206 to classify cells of the one or more test samples treated with a chemical and the one
10 or more control samples treated with a solvent control at each of the plurality of test concentrations (for example at each of the plurality of test concentrations selected from 0 μm , 62.5 μm , 125 μm , 250 μm , 500 μm and 1000 μm). The supervised classifier 206 may classify the cells into two classes, for example an assigned "treated" class and an assigned "control" class. In the exemplary embodiment, the classification is based on the
15 one or more features 202 obtained/captured by a feature measurement module, for example a feature measurement module that is similar to the feature measurement module 104 of the system 100 described with reference to FIG. 1.

 The classification for example provides a measurement of the similarity or
20 distinguishability between cells of the test samples and of the control samples.

 In the exemplary embodiment, when the cells are fully distinguishable by the supervised classifier 206 between the cells of the one or more test samples and the cells of the one or more control samples at a particular test concentration, the classification
25 module 106 provides/computes a first classification accuracy value, for example a high first classification accuracy value 218 at the test concentration. In the exemplary embodiment, when the cells are fully indistinguishable by the supervised classifier 206 between the cells of the one or more test samples and the cells of the one or more control samples at a particular test concentration, the classification module 106
30 provides/computes a second classification accuracy value, for example a low classification accuracy value 210 at the test concentration. In the exemplary embodiment, when the cells are partially distinguishable by the supervised classifier 206 between the cells of the one or more test samples and the cells of the one or more control samples at

a particular test concentration, the classification module 106 provides intermediate classification accuracy values 212, 214, or 216 at the test concentrations.

For example, the classification module 106 may determine/compute the accuracy value by counting the numbers of test samples (positives = P) and control samples (negatives = N), and those correctly classified as test samples (true positives = TP) and control samples (true negatives = TN), respectively; and set to a balanced accuracy value e.g. $((TP/P + TN/N)/2)$. In one example, the classification module 106 may provide a percentage value based on the balanced accuracy value.

The supervised classifier 206 may be a trained classifier. In an exemplary embodiment, the supervised classifier 206 is pre-trained to classify cells into an assigned "treated" class and into an assigned "control" class. To this end, for example, a training set may first be used to train the supervised classifier 206, wherein in the training set, samples that were treated with the chemical are annotated as belonging to the "treated" class and samples that were treated with the solvent control are annotated as belonging to the "control" class. The training set can be a subset selected from the total samples, with another subset of the remaining samples making up a validation set. The supervised classifier 206 may also go through a second or further iteration of training using the same training set or a subset thereof or a revised training set based on the validation set. Cross-validation may also be performed where the total samples are repeatedly split into a training dataset and a validation dataset. These repeated partitions can be done in various ways, such as dividing the total samples into two equal datasets and using them as training/validation, and then validation/training, or repeatedly selecting a random subset from the total samples as a validation dataset. As may be appreciated, this training procedure may allow the supervised classifier 206 to distinguish between a sample that has been treated with a chemical and a sample that has been treated with a solvent control at a particular concentration e.g. based on dissimilarity or similarity in certain features between the two samples, if any.

In the exemplary embodiment, when the cells are distinguishable between the cells of the one or more test samples and the cells of the one or more control samples by the supervised classifier 206 at a particular concentration (e.g. when the cells of the one or

more test samples and the cells of the one or more control samples are dissimilar in features), a large proportion of samples may be “correctly” classified by the supervised classifier 206 e.g. the total number of test samples that are classified into the “treated” class and control samples that are classified into the “control” class by the supervised classifier 206 may be substantially higher as compared to the total number of test samples that are instead classified into the “control” class and control samples that are instead classified into the “treated” class. In such a scenario, the classification module 106 provides a first classification accuracy value (e.g. 216 or 218) which may correspond to the average proportion of samples that are “correctly” classified, or “balanced accuracy”. For example, if 90% of the test samples and 80% of the control samples are “correctly” classified by the supervised classifier 206, the classification module 106 may provide a classification accuracy value of $(90\%+80\%)/2 = 85\%$.

In the exemplary embodiment, the method/concept 200 further comprises plotting the classification accuracy values 210, 212, 214, 216 and 218 or their derivatives thereof as data points and formulating a first concentration-accuracy curve 222 by the curve formulation module 108. In the exemplary embodiment, a first initial POD value is then derived by determining a point 220 on the classification-accuracy curve 222 corresponding to a point on the curve that represents the EC_{10} value, which is the concentration that provides an accuracy value 10% between the lower and upper limits of the fitted concentration-accuracy curve. In other embodiments, the first initial POD value may be derived by determining a point on the classification-accuracy curve corresponding to a point on the curve that represents the NOAEL, LOAEL or BMD10 value.

In the exemplary embodiment, the supervised classifier 206 comprises a Support Vector Machine (SVM), for example, but not limited to a multi-dimensional SVM. In other embodiments, the supervised classifier may comprise other forms of classifiers such as, but not limited to, a random forest classifier or a neural network classifier. In addition, in yet other embodiments, an unsupervised classifier may be used in place of a supervised classifier.

FIG. 3 is a diagram illustrating a method/concept 300 that may be implemented by the system 100 described with reference to FIG. 1 for determining a POD of a chemical in an exemplary embodiment.

5 In the exemplary embodiment, at step 302, test and control samples in the form of biological samples are collected from test and control populations respectively and are treated/exposed to either a solvent control or a chemical at multiple concentrations. For example, the solvent may be dimethyl sulfoxide (DMSO) or water (H₂O). The test and/or control populations may comprise cells, tissues, organs and/or organisms. The cells,
10 tissues, organs may be obtained from animals or humans. The cells may also be obtained from *in vitro* cell lines. The test and/or control samples/populations may be of multiple different tissue origins or a single tissue origin. In various embodiments, the cell type/first cell type comprises at least one of a lung cell, a lung-derived cell, a lung-derived cell line cell, a liver cell, a liver-derived cell, a liver-derived cell line cell, a kidney cell, a kidney-
15 derived cell or a kidney-derived cell line cell. In the exemplary embodiment, the test and control samples comprise cells collected/obtained from a human cell line selected from a BEAS-2B cell line (a lung cell line), a HepG2 cell line (a liver cell line), and a HK2 cell line (a kidney cell line).

20 In the exemplary embodiment, the test and control samples/populations were stained with markers that label specific proteins, organelles, or subcellular structures inside the cells. In the exemplary embodiment, at least three markers were used for labelling. They are Hoechst (a marker for DNA), phalloidin (a marker for actin filament), and anti-γH2AX (a marker for phosphorylated histone 2AX).

25 In various embodiments therefore, the samples/populations are contacted/stained/labelled/tagged with a labelling agent/marker that is detectable by an imaging module and/or detectable by a feature measurement module, such as a feature measurement module that is similar to feature measurement module 104 of FIG. 1. Hence, the labelling agent/ marker may be fluorescent, luminescent, radioactive etc. The
30 labelling agent/marker may specifically bind to and label cell components e.g. DNA, actin filament, phosphorylated histone 2AX etc. Further, the labelling agent/marker may emit a signal which intensity is related to the concentration of the cell component to which the

agent/marker binds. In one embodiment, the signal intensity is directly proportional to the concentration of the underlying cell component. In one embodiment, the signal source is indicative of the location of the underlying cell component. As may be appreciated, the test and control samples/populations may also be contacted/stained/labelled/tagged with other suitable labelling agents/markers, other than or in addition to Hoechst, phalloidin, and anti- γ H2AX used in the exemplary embodiment. For example, a marker so-called “CellMask™” or other similar plasma membrane stains may also be used for cell segmentation purposes. The marker may facilitate cell size measurements etc

In the exemplary embodiment, the samples/cells may be imaged by microscopy (or other suitable imaging modules may also be used).

In the exemplary embodiment, at step 304, readouts/features in the form of molecular and/or phenotypic readouts/features for each sample are measured. Compare feature measurement module 104 of FIG. 1. The readouts/features may be cellular phenotypic features measured using e.g. microscopy, RNA expression levels measured using e.g. RNA sequencing, protein expression levels measured using e.g. flow cytometry or mass spectrometry, or other biological activity measurements using conventional biological activity measurement methods/apparatus. In the exemplary embodiment, a large number of molecular and phenotypic readouts/features (165 in total) as listed in Table 1 below are measured using the “cellXpress” software (v1.0, Bioinformatics Institute) from fluorescence images of the cells of the samples stained with Hoechst, phalloidin, anti- γ H2AX, and CellMask™. The features of Table 1 are understood to be provided as examples only.

Table 1

No	Feature ID	Type	Description
1	area:mask:cell_region	Morphology	Cell area
2	area:mask:dna_region	Morphology	Nuclear area
3	perimeter:mask:cell_region	Morphology	Cell perimeter length
4	perimeter:mask:dna_region	Morphology	Nuclear perimeter length
5	roundness:mask:cell_region	Morphology	Cell roundness
6	roundness:mask:dna_region	Morphology	Nuclear roundness
7	aspect_ratio:mask:cell_region	Morphology	Cell aspect ratio
8	aspect_ratio:mask:dna_region	Morphology	Nuclear aspect ratio

9	obj_number:mask:DNA_object	Morphology	Number of DNA objects
10	obj_mean_total_area:mask:DNA_object	Morphology	Mean total area of DNA objects
11	obj_stddev_total_area:mask:DNA_object	Morphology	Standard deviation in the total area of DNA objects
12	obj_number:mask:gH2AX_object	Morphology	Number of gH2AX objects
13	obj_mean_total_area:mask:gH2AX_object	Morphology	Mean total area of gH2AX objects
14	obj_stddev_total_area:mask:gH2AX_object	Morphology	Standard deviation in the total area of gH2AX objects
15	obj_number:mask:Actin_object	Morphology	Number of actin objects
16	obj_mean_total_area:mask:Actin_object	Morphology	Mean total area of actin objects
17	obj_stddev_total_area:mask:Actin_object	Morphology	Standard deviation in the total area of actin objects
18	total_intensity:gH2AX:cell_region	Intensity	Total gH2AX intensity at the whole-cell region
19	total_intensity:Actin:cell_region	Intensity	Total actin intensity at the whole-cell region
20	mean_intensity:gH2AX:cell_region	Intensity	Mean gH2AX intensity at the whole-cell region
21	mean_intensity:Actin:cell_region	Intensity	Mean actin intensity at the whole-cell region
22	cv_intensity:gH2AX:cell_region	Intensity	Coefficient of variation of gH2AX intensity at the whole-cell region
23	cv_intensity:Actin:cell_region	Intensity	Coefficient of variation of actin intensity at the whole-cell region
24	total_intensity:DNA:dna_region	Intensity	Total DNA intensity at the nuclear region
25	total_intensity:DNA:dna_chromosome	Intensity	Total DNA intensity at the chromosomal region
26	mean_intensity:DNA:dna_region	Intensity	Mean DNA intensity at the nuclear region
27	mean_intensity:DNA:dna_chromosome	Intensity	Mean DNA intensity at the chromosomal region
28	cv_intensity:DNA:dna_region	Intensity	Coefficient of variation of DNA intensity at the nuclear region
29	cv_intensity:DNA:dna_chromosome	Intensity	Coefficient of variation of DNA intensity at the chromosomal region
30	fraction_obj_intensity:DNA:dna_regionDNA_object	Intensity ratios	Fraction of total DNA object intensity at the nuclear region over the whole-cell region
31	fraction_obj_intensity:DNA:dna_chromosomeDNA_object	Intensity ratios	Fraction of total DNA object intensity at the chromosomal region over the whole-cell region
32	fraction_obj_intensity:DNA:nondna_regionDNA_object	Intensity ratios	Fraction of total DNA object intensity at the cytoplasmic region over the whole-cell region
33	total_intensity:gH2AX:dna_region	Intensity	Total gH2AX intensity at the nuclear region
34	total_intensity:gH2AX:dna_chromosome	Intensity	Total gH2AX intensity at the chromosomal region
35	mean_intensity:gH2AX:dna_region	Intensity	Mean gH2AX intensity at the nuclear region
36	mean_intensity:gH2AX:dna_chromosome	Intensity	Mean gH2AX intensity at the chromosomal region
37	cv_intensity:gH2AX:dna_region	Intensity	Coefficient of variation of gH2AX intensity at the nuclear region
38	cv_intensity:gH2AX:dna_chromosome	Intensity	Coefficient of variation of gH2AX intensity at the chromosomal region
39	total_intensity_ratio:gH2AXgH2AX:dna_region-cell_region	Intensity ratios	Ratio between total gH2AX intensity at the nuclear region over the whole-cell region
40	total_intensity_ratio:gH2AXgH2AX:dna_chromosome-cell_region	Intensity ratios	Ratio between total gH2AX intensity at the chromosomal region over the whole-cell region
41	total_intensity_ratio:gH2AXgH2AX:dna_chromosome-dna_region	Intensity ratios	Ratio between total gH2AX intensity at the chromosomal region over the nuclear region
42	fraction_obj_intensity:gH2AX:dna_regiongH2AX_object	Intensity ratios	Fraction of total gH2AX object intensity at the nuclear region over the whole-cell region
43	fraction_obj_intensity:gH2AX:dna_chromosomegH2AX_object	Intensity ratios	Fraction of total gH2AX object intensity at the chromosomal region over the whole-cell region
44	total_intensity:Actin:dna_region	Intensity	Total actin intensity at the nuclear region
45	total_intensity:Actin:dna_chromosome	Intensity	Total actin intensity at the chromosomal region
46	total_intensity:Actin:nondna_region	Intensity	Total actin intensity at the cytoplasmic region

47	total_intensity:Actin:nondna_outer	Intensity	Total actin intensity at the outer cytoplasmic region
48	total_intensity:Actin:nondna_inner	Intensity	Total actin intensity at the inner cytoplasmic region
49	total_intensity:Actin:nondna_peridna	Intensity	Total actin intensity at the pericellular region
50	mean_intensity:Actin:dna_region	Intensity	Mean actin intensity at the nuclear region
51	mean_intensity:Actin:dna_chromosome	Intensity	Mean actin intensity at the chromosomal region
52	mean_intensity:Actin:nondna_region	Intensity	Mean actin intensity at the cytoplasmic region
53	mean_intensity:Actin:nondna_outer	Intensity	Mean actin intensity at the outer cytoplasmic region
54	mean_intensity:Actin:nondna_inner	Intensity	Mean actin intensity at the inner cytoplasmic region
55	mean_intensity:Actin:nondna_peridna	Intensity	Mean actin intensity at the pericellular region
56	cv_intensity:Actin:dna_region	Intensity	Coefficient of variation of actin intensity at the nuclear region
57	cv_intensity:Actin:dna_chromosome	Intensity	Coefficient of variation of actin intensity at the chromosomal region
58	cv_intensity:Actin:nondna_region	Intensity	Coefficient of variation of actin intensity at the cytoplasmic region
59	cv_intensity:Actin:nondna_outer	Intensity	Coefficient of variation of actin intensity at the outer cytoplasmic region
60	cv_intensity:Actin:nondna_inner	Intensity	Coefficient of variation of actin intensity at the inner cytoplasmic region
61	cv_intensity:Actin:nondna_peridna	Intensity	Coefficient of variation of actin intensity at the pericellular region
62	total_intensity_ratio:Actin-Actin:dna_regioncell_region	Intensity ratios	Ratio between total actin intensity at the nuclear region over the whole-cell region
63	total_intensity_ratio:Actin-Actin:dna_chromosome-cell_region	Intensity ratios	Ratio between total actin intensity at the chromosomal region over the whole-cell region
64	total_intensity_ratio:Actin-Actin:dna_chromosome-dna_region	Intensity ratios	Ratio between total actin intensity at the chromosomal region over the nuclear region
65	total_intensity_ratio:Actin-Actin:nondna_outer-cell_region	Intensity ratios	Ratio between total actin intensity at the outer cytoplasmic region over the whole-cell region
66	total_intensity_ratio:Actin-Actin:nondna_inner-cell_region	Intensity ratios	Ratio between total actin intensity at the inner cytoplasmic region over the whole-cell region
67	total_intensity_ratio:Actin-Actin:nondna_peridna-cell_region	Intensity ratios	Ratio between total actin intensity at the pericellular region over the whole-cell region
68	fraction_obj_intensity:Actin:dna_regionActin_object	Intensity ratios	Fraction of total actin intensity at the nuclear region actin objects over the whole-cell actin objects
69	fraction_obj_intensity:Actin:dna_chromosome-Actin_object	Intensity ratios	Fraction of total actin intensity at the chromosomal region actin objects over the whole-cell actin objects
70	fraction_obj_intensity:Actin:nondna_regionActin_object	Intensity ratios	Fraction of total actin intensity at the cytoplasmic region actin objects over the whole-cell actin objects
71	fraction_obj_intensity:Actin:nondna_outerActin_object	Intensity ratios	Fraction of total actin intensity at the outer cytoplasmic region actin objects over the whole-cell actin objects
72	fraction_obj_intensity:Actin:nondna_innerActin_object	Intensity ratios	Fraction of total actin intensity at the inner cytoplasmic region actin objects over the whole-cell actin objects
73	fraction_obj_intensity:Actin:nondna_peridnaActin_object	Intensity ratios	Fraction of total actin intensity at the pericellular region actin objects over the whole-cell actin objects
74	total_intensity_ratio:gH2AX-Actin:cell_region-cell_region	Intensity ratios	Ratio between total gH2AX over actin intensities at the wholecell region
75	total_intensity_ratio:gH2AX-DNA:cell_regioncell_region	Intensity ratios	Ratio between total gH2AX over DNA intensities at the wholecell region
76	total_intensity_ratio:DNA-Actin:cell_regioncell_region	Intensity ratios	Ratio between total DNA over actin intensities at the wholecell region

77	total_intensity_ratio:gH2AX-Actin:dna_region-dna_region	Intensity ratios	Ratio between total gH2AX over actin intensities at the nuclear region
78	total_intensity_ratio:gH2AX-DNA:dna_region-dna_region	Intensity ratios	Ratio between total gH2AX over DNA intensities at the nuclear region
79	total_intensity_ratio:DNA-Actin:dna_regiondna_region	Intensity ratios	Ratio between total DNA over actin intensities at the nuclear region
80	total_intensity_ratio:gH2AX-Actin:dna_chromosome-dna_chromosome	Intensity ratios	Ratio between total gH2AX over actin intensities at the chromosomal region
81	total_intensity_ratio:gH2AX-DNA:dna_chromosome-dna_chromosome	Intensity ratios	Ratio between total gH2AX over DNA intensities at the chromosomal region
82	total_intensity_ratio:DNA-Actin:dna_chromosome-dna_chromosome	Intensity ratios	Ratio between total DNA over actin intensities at the chromosomal region
83	ccorr_normed:DNA-gH2AX:cell_region	Pixel correlations	Spatial correlation coefficient of DNA and gH2AX intensities at the whole-cell region
84	ccoeff_normed:DNA-gH2AX:cell_region	Pixel correlations	Correlation coefficient of DNA and gH2AX intensities at the whole-cell region
85	ccorr_normed:DNA-Actin:cell_region	Pixel correlations	Spatial correlation coefficient of DNA and actin intensities at the whole-cell region
86	ccoeff_normed:DNA-Actin:cell_region	Pixel correlations	Correlation coefficient of DNA and actin intensities at the whole-cell region
87	ccorr_normed:gH2AX-Actin:cell_region	Pixel correlations	Spatial correlation coefficient of gH2AX and actin intensities at the whole-cell region
88	ccoeff_normed:gH2AX-Actin:cell_region	Pixel correlations	Correlation coefficient of gH2AX and actin intensities at the whole-cell region
89	ccorr_normed:DNA-gH2AX:dna_region	Pixel correlations	Spatial correlation coefficient of DNA and gH2AX intensities at the nuclear region
90	ccoeff_normed:DNA-gH2AX:dna_region	Pixel correlations	Correlation coefficient of DNA and gH2AX intensities at the nuclear region
91	ccorr_normed:DNA-Actin:dna_region	Pixel correlations	Spatial correlation coefficient of DNA and actin intensities at the nuclear region
92	ccoeff_normed:DNA-Actin:dna_region	Pixel correlations	Correlation coefficient of DNA and actin intensities at the nuclear region
93	ccorr_normed:gH2AX-Actin:dna_region	Pixel correlations	Spatial correlation coefficient of gH2AX and actin intensities at the nuclear region
94	ccoeff_normed:gH2AX-Actin:dna_region	Pixel correlations	Correlation coefficient of gH2AX and actin intensities at the nuclear region
95	ccorr_normed:DNAgH2AX:dna_chromosome	Pixel correlations	Spatial correlation coefficient of DNA and gH2AX intensities at the chromosomal region
96	ccoeff_normed:DNAgH2AX:dna_chromosome	Pixel correlations	Correlation coefficient of DNA and gH2AX intensities at the chromosomal region
97	ccorr_normed:DNA-Actin:dna_chromosome	Pixel correlations	Spatial correlation coefficient of DNA and actin intensities at the chromosomal region
98	ccoeff_normed:DNA-Actin:dna_chromosome	Pixel correlations	Correlation coefficient of DNA and actin intensities at the chromosomal region
99	ccorr_normed:gH2AX-Actin:dna_chromosome	Pixel correlations	Spatial correlation coefficient of gH2AX and actin intensities at the chromosomal region
100	ccoeff_normed:gH2AX-Actin:dna_chromosome	Pixel correlations	Correlation coefficient of gH2AX and actin intensities at the chromosomal region
101	glcm_asm_mean:Actin:cell_region	Glcm textures	Mean angular second moment of the whole-cell actin GLCM
102	glcm_contrast_mean:Actin:cell_region	Glcm textures	Mean contrast of the whole-cell actin GLCM
103	glcm_corr_mean:Actin:cell_region	Glcm textures	Mean correlation of the whole-cell actin GLCM
104	glcm_var_mean:Actin:cell_region	Glcm textures	Mean of the variance of the whole-cell actin GLCM
105	glcm_idm_mean:Actin:cell_region	Glcm textures	Mean inverse difference moment of the whole-cell actin GLCM
106	glcm_sum_ave_mean:Actin:cell_region	Glcm textures	Mean sum average of the whole-cell actin GLCM
107	glcm_sum_var_mean:Actin:cell_region	Glcm textures	Mean of the sum variance of the whole-cell actin GLCM
108	glcm_sum_ent_mean:Actin:cell_region	Glcm textures	Mean sum entropy of the whole-cell actin GLCM

109	glcm_ent_mean:Actin:cell_region	GlcM textures	Mean entropy of the whole-cell actin GLCM
110	glcm_diff_var_mean:Actin:cell_region	GlcM textures	Mean difference variance of the whole-cell actin GLCM
111	glcm_diff_ent_mean:Actin:cell_region	GlcM textures	Mean difference entropy of the whole-cell actin GLCM
112	glcm_info_corr1_mean:Actin:cell_region	GlcM textures	Mean information measure of correlation 1 of the whole-cell actin GLCM
113	glcm_info_corr2_mean:Actin:cell_region	GlcM textures	Mean information measure of correlation 2 of the whole-cell actin GLCM
114	glcm_asm_mean:Actin:dna_region	GlcM textures	Mean angular second moment of the nuclear actin GLCM
115	glcm_contrast_mean:Actin:dna_region	GlcM textures	Mean contrast of the nuclear actin GLCM
116	glcm_corr_mean:Actin:dna_region	GlcM textures	Mean correlation of the nuclear actin GLCM
117	glcm_var_mean:Actin:dna_region	GlcM textures	Mean of the variance of the nuclear actin GLCM
118	glcm_idm_mean:Actin:dna_region	GlcM textures	Mean inverse difference moment of the nuclear actin GLCM
119	glcm_sum_ave_mean:Actin:dna_region	GlcM textures	Mean sum average of the nuclear actin GLCM
120	glcm_sum_var_mean:Actin:dna_region	GlcM textures	Mean of the sum variance of the nuclear actin GLCM
121	glcm_sum_ent_mean:Actin:dna_region	GlcM textures	Mean sum entropy of the nuclear actin GLCM
122	glcm_ent_mean:Actin:dna_region	GlcM textures	Mean entropy of the nuclear actin GLCM
123	glcm_diff_var_mean:Actin:dna_region	GlcM textures	Mean difference variance of the nuclear actin GLCM
124	glcm_diff_ent_mean:Actin:dna_region	GlcM textures	Mean difference entropy of the nuclear actin GLCM
125	glcm_info_corr1_mean:Actin:dna_region	GlcM textures	Mean information measure of correlation 1 of the nuclear actin GLCM
126	glcm_info_corr2_mean:Actin:dna_region	GlcM textures	Mean information measure of correlation 2 of the nuclear actin GLCM
127	glcm_asm_mean:Actin:nondna_region	GlcM textures	Mean angular second moment of the cytoplasmic actin GLCM
128	glcm_contrast_mean:Actin:nondna_region	GlcM textures	Mean contrast of the cytoplasmic actin GLCM
129	glcm_corr_mean:Actin:nondna_region	GlcM textures	Mean correlation of the cytoplasmic actin GLCM
130	glcm_var_mean:Actin:nondna_region	GlcM textures	Mean of the variance of the cytoplasmic actin GLCM
131	glcm_idm_mean:Actin:nondna_region	GlcM textures	Mean inverse difference moment of the cytoplasmic actin GLCM
132	glcm_sum_ave_mean:Actin:nondna_region	GlcM textures	Mean sum average of the cytoplasmic actin GLCM
133	glcm_sum_var_mean:Actin:nondna_region	GlcM textures	Mean of the sum variance of the cytoplasmic actin GLCM
134	glcm_sum_ent_mean:Actin:nondna_region	GlcM textures	Mean sum entropy of the cytoplasmic actin GLCM
135	glcm_ent_mean:Actin:nondna_region	GlcM textures	Mean entropy of the cytoplasmic actin GLCM
136	glcm_diff_var_mean:Actin:nondna_region	GlcM textures	Mean difference variance of the cytoplasmic actin GLCM
137	glcm_diff_ent_mean:Actin:nondna_region	GlcM textures	Mean difference entropy of the cytoplasmic actin GLCM
138	glcm_info_corr1_mean:Actin:nondna_region	GlcM textures	Mean information measure of correlation 1 of the cytoplasmic actin GLCM
139	glcm_info_corr2_mean:Actin:nondna_region	GlcM textures	Mean information measure of correlation 2 of the cytoplasmic actin GLCM
140	glcm_asm_mean:DNA:dna_region	GlcM textures	Mean angular second moment of the nuclear DNA GLCM
141	glcm_contrast_mean:DNA:dna_region	GlcM textures	Mean contrast of the nuclear DNA GLCM
142	glcm_corr_mean:DNA:dna_region	GlcM textures	Mean correlation of the nuclear DNA GLCM
143	glcm_var_mean:DNA:dna_region	GlcM textures	Mean of the variance of the nuclear DNA GLCM

144	glcm_idm_mean:DNA:dna_region	GlcM textures	Mean inverse difference moment of the nuclear DNA GLCM
145	glcm_sum_ave_mean:DNA:dna_region	GlcM textures	Mean sum average of the nuclear DNA GLCM
146	glcm_sum_var_mean:DNA:dna_region	GlcM textures	Mean of the sum variance of the nuclear DNA GLCM
147	glcm_sum_ent_mean:DNA:dna_region	GlcM textures	Mean sum entropy of the nuclear DNA GLCM
148	glcm_ent_mean:DNA:dna_region	GlcM textures	Mean entropy of the nuclear DNA GLCM
149	glcm_diff_var_mean:DNA:dna_region	GlcM textures	Mean difference variance of the nuclear DNA GLCM
150	glcm_diff_ent_mean:DNA:dna_region	GlcM textures	Mean difference entropy of the nuclear DNA GLCM
151	glcm_info_corr1_mean:DNA:dna_region	GlcM textures	Mean information measure of correlation 1 of the nuclear DNA GLCM
152	glcm_info_corr2_mean:DNA:dna_region	GlcM textures	Mean information measure of correlation 2 of the nuclear DNA GLCM
153	glcm_asm_mean:gH2AX:dna_region	GlcM textures	Mean angular second moment of the nuclear gH2AX GLCM
154	glcm_contrast_mean:gH2AX:dna_region	GlcM textures	Mean contrast of the nuclear gH2AX GLCM
155	glcm_corr_mean:gH2AX:dna_region	GlcM textures	Mean correlation of the nuclear gH2AX GLCM
156	glcm_var_mean:gH2AX:dna_region	GlcM textures	Mean of the variance of the nuclear gH2AX GLCM
157	glcm_idm_mean:gH2AX:dna_region	GlcM textures	Mean inverse difference moment of the nuclear gH2AX GLCM
158	glcm_sum_ave_mean:gH2AX:dna_region	GlcM textures	Mean sum average of the nuclear gH2AX GLCM
159	glcm_sum_var_mean:gH2AX:dna_region	GlcM textures	Mean of the sum variance of the nuclear gH2AX GLCM
160	glcm_sum_ent_mean:gH2AX:dna_region	GlcM textures	Mean sum entropy of the nuclear gH2AX GLCM
161	glcm_ent_mean:gH2AX:dna_region	GlcM textures	Mean entropy of the nuclear gH2AX GLCM
162	glcm_diff_var_mean:gH2AX:dna_region	GlcM textures	Mean difference variance of the nuclear gH2AX GLCM
163	glcm_diff_ent_mean:gH2AX:dna_region	GlcM textures	Mean difference entropy of the nuclear gH2AX GLCM
164	glcm_info_corr1_mean:gH2AX:dna_region	GlcM textures	Mean information measure of correlation 1 of the nuclear gH2AX GLCM
165	glcm_info_corr2_mean:gH2AX:dna_region	GlcM textures	Mean information measure of correlation 2 of the nuclear gH2AX GLCM

The features include 65 texture features (which include measuring statistics of the spatial co-occurrence patterns of markers), 36 intensity features (which include measuring the staining levels of markers at different subcellular regions), 29 intensity ratio features (which include measuring the ratios between intensity features), 18 correlation features (which include measuring the spatial correlations between two markers at the single-cell level), and 17 morphology features (which include measuring the shape properties of the nuclear and cellular regions). In various embodiments therefore, the feature measurement module is arranged to measure one or more features selected from the group consisting of: texture features, intensity features, intensity ratio features, correlation features, morphology features and combinations thereof. In various

embodiments, the feature measurement module is arranged to measure one or more features selected from a feature in Table 1. In various embodiments, the feature measurement module is arranged to measure no less than 165, no less than 160, no less than 150, no less than 140, no less than 130, no less than 120, no less than 110, no less than 100, no less than 90, no less than 80, no less than 70, no less than 60 or no less than 50 features. As may be appreciated, embodiments of the feature measurement module may also be arranged to measure other/further features that may be influenced by chemical treatment.

Further, in various embodiments, the feature measurement module is arranged to detect/measure no less than 165, no less than 160, no less than 150, no less than 140, no less than 130, no less than 120, no less than 110, no less than 100, no less than 90, no less than 80, no less than 70, no less than 60 or no less than 50 features from samples contacted/stained/labelled/tagged with for example but not limited to 4 labelling agents/markers. In various embodiments, it may also be provided that no more than 4 labelling agents/markers be used for the staining/labelling/tagging/contacting etc. Thus, embodiments of the system 100 of FIG. 1 are capable of usefully analysing a large number of readouts/features to determine a POD of a chemical based on few cellular imaging assays.

In the exemplary embodiment, at step 306, up to e.g. 1000 cells from the chemically-treated cellular population and up to e.g. 1000 cells from the control cellular population are randomly sampled. A supervised classifier, such as one similar to supervised classifier 206 described with reference to FIG. 2, is trained to classify all the cells (e.g. up to 2000 cells) into two classes, namely a "treated" class or a "control" class, based on the measured molecular and/or phenotypic readouts or based on the obtained features or measured feature values of step 304. When the control and treated cells are similar and cannot be separated by the classifier (i.e. when the control and treated cells are indistinguishable by the classifier), a classification module (compare the classification module 106 of FIG. 1) provides/computes a classification accuracy value that is of a low value (e.g. about 50-60%). On the other hand, when the control and treated cells are different and can be separated by the supervised classifier (i.e. when the control and treated cells are distinguishable by the classifier), the classification module is arranged to

provide/compute a classification accuracy value that is of a high value (e.g. about 70-100%).

In the exemplary embodiment, at step 308, a concentration-accuracy curve may be fitted based on the estimated accuracy values at all tested concentrations. For example, a curve formulation module (compare curve formulation module 108 of FIG. 1) may be used. The result is a "concentration-accuracy curve" (similar to a concentration-accuracy curve 222 described in FIG. 2), where the y-axis of the curve represents the classification accuracy or an estimated classification accuracy, which describes/indicates the magnitude of the chemical effects in any one or more of the measured molecular and/or phenotypic readouts, and the x-axis of the curve represents the chemical concentration. In the exemplary embodiment, this step also includes a procedure to fit e.g. two models. For example, an automated procedure may be used to fit both a 4-parameter log-logistic model ($Acc(x) = b + \frac{a-b}{1+\exp(c(\log(x)-\log(EC_{50})))}$) and a constant model ($Acc(x) = d$) to the data points, where $Acc(x)$ is the classification accuracy at concentration x of the chemical, EC_{50} is the effective concentration at the half accuracy-range level, a is the upper limit of classification accuracy, b is the lower limit of classification accuracy, and d is the constant accuracy level. Subsequently, the Akaike information criterion (AIC) is used to determine the best or preferred fitted model between the e.g. two models. In the exemplary embodiment, still at step 308, the chemical is labelled as "inactive" if the preferred fit is based on a constant model and the chemical is labelled as "active" if the preferred fit is not a constant model. In other words, a chemical is considered "inactive" if the constant model is the best fit; otherwise, the chemical is considered "active". Based on the best fitted concentration-accuracy curve model, an initial POD value of the chemical is determined. For example, a POD determination module (compare the POD determination module 110 of FIG. 1) may be used. In the exemplary embodiment, for an "active" chemical, the initial POD value is the EC_{10} value derived from the fitted concentration-accuracy curve; for an "inactive" chemical, the initial POD value is an arbitrary high value, which is set to 10^5 . As may be appreciated, other suitable model selection criteria may also be used. As may also be appreciated, other types of POD values may also be derived from the fitted curves, such as LOAEL, NOAEL, or BMD10. In some other exemplary embodiments, the initial POD value may be taken to be the POD of the chemical.

It may be appreciated that other statistical models may also be used for the fitting to the data points. For example, at least one of a log-logistic model, a logistic model, a Weibull model, a gamma model, and a constant model may be used.

5

In the exemplary embodiment, at step 310, imaging is performed (e.g. by an imaging module) to image (a) one or more test samples having been treated with the chemical at the plurality of test concentrations and being obtained from a second test population of the first cell type (i.e. the same cell type as the first test population described above in any of steps 302, 304, 306 or 308), and (b) one or more control samples having been treated with the solvent control at one or a plurality of control concentrations and being obtained from a second control population of the first cell type (i.e. the same cell type as the first control population described above in any of steps 302, 304, 306 or 308), to obtain one or more features of these second populations. The one or more features obtained for the one or more test samples from the second test population and the one or more control samples from the second control population are then measured (e.g. by the feature measurement module). Each sample from the second test population and the second control population is then classified based on a similarity between the one or more test samples from the second test population and the one or more control samples from the second control population at each of the plurality of concentrations e.g. by the classification module to provide a second classification result. The classification process is performed substantially similarly to step 306. A second concentration-accuracy curve based on the second classification result is then formulated e.g. by the curve formulation module. At least a second initial POD value of the chemical based on the second concentration-accuracy curve is then derived e.g. by the POD determination module. With a plurality of different initial POD values (e.g. with the second and the first initial POD values), an average initial POD value may be obtained. Thus, in the exemplary embodiment, by carrying out the steps for obtaining a second initial POD value, a biological replicate is performed, and an average POD value can be obtained.

30

In the exemplary embodiment, if the two replicates are both active or both inactive, the average POD value is set to be the mean of their POD values. Otherwise, the average

POD value is set to “NA”, or not available, meaning that a reproducible average POD value cannot be obtained.

In the exemplary embodiment, imaging is performed (e.g. by an imaging module) to image (a) one or more test samples having been treated with the chemical at the plurality of test concentrations and being obtained from a test population of a different cell type (i.e. a cell type that is not the same as that of the first test population described above in any of steps 302, 304, 306 or 308 or that of the second test population described in step 310), and (b) one or more control samples having been treated with the solvent control at one or a plurality of control concentrations and being obtained from a control population of the different cell type (i.e. a cell type that is not the same as that of the first control population described above in any of steps 302, 304, 306 or 308 or that of the second control population described in step 310), to obtain one or more features of these different cell type populations. For example, the first cell type may be a lung cell line while the different cell type may be a liver cell line. The one or more features obtained for the one or more test samples from the test population of a different cell type and the one or more control samples from the control population of a different cell type are then measured (e.g. by the feature measurement module). Each sample from the test control and control population of a different cell type is then classified based on a similarity between the one or more test samples from the test population of a different cell type and the one or more control samples from the control population of a different cell type at each of the plurality of concentrations e.g. by the classification module to provide a different classification result. The classification process is performed substantially similarly to step 306. A different concentration-accuracy curve that is associated with said different cell type based on the different classification result is then formulated e.g. by the curve formulation module. A different initial POD value of the chemical based on the different concentration-accuracy curve is then derived e.g. by the POD determination module.

In the exemplary embodiment, a second test population and a second control population may be used for the different cell type to obtain at least a second different initial POD value. An average POD value for the different cell type may then be determined. Thus, in the exemplary embodiment, by carrying out the steps for obtaining

a different initial POD value, a POD value (or an average POD value) that is associated with a cell type that is different from the first cell type is obtained.

The procedures outlined above may be repeated to obtain a further different initial
5 POD value (or a further different average POD value) associated with a further different cell type (e.g. a kidney cell line).

In the exemplary embodiment, an average POD value is obtained for each of a
lung cell (from a BEAS-2B cell line), a liver cell (from a HepG2 cell line) and a kidney cell
10 (from a HK2 cell line).

At step 312, after the average POD value is obtained (or if no biological replicates are performed, a POD value) for each cell type, the POD values are consolidated to derive a single final POD value for the chemical. In the exemplary embodiment, the minimum of
15 the three average POD values (or if no biological replicates are performed, the minimum of the three POD values) obtained is taken to be the final POD value. In the exemplary embodiment, when one of the three cell types gives NA value, the minimum of the two other average POD values (or if no biological replicates are performed, the minimum of the two POD values) obtained is taken to be the final POD value. In the exemplary
20 embodiment, when two of the three cell types give NA values and only one cell type is “active” for the chemical, the final POD value is set to be the average POD value (or if no biological replicates are performed, the POD value) of the active cell type. In the exemplary embodiment, when all three cell types are “inactive” for the chemical, the final POD value is set to NA.

In the exemplary embodiment, it is mentioned that a second population is used
e.g. for obtaining a second initial POD value. It will be appreciated that any number of different populations may be used. Also, it is possible that only one population is used
(i.e. averaging is not used).

FIG. 4 illustrates examples of accuracy-concentration curves and POD values estimated, based on an exemplary embodiment of the system or method, in respect of four chemicals (A) di(2-ethylhexyl) adipate, (B) diethyl phthalate, (C) propylparaben and

(D) p,p'-DDT in HepG2 cells. Black circles represent the data points obtained from a first experiment and white circles represent the data points obtained from a second experimental replicate. The vertical dotted lines indicate the estimated POD values from each experiment. As may be observed, (A) appeared to be "inactive" for HepG2 cells.

Embodiments of the system were used to estimate/determine the POD for a total 64 chemicals that represent 13 chemical classes with similar/related chemical structures (see table 2 below). The results for the remaining 60 chemicals are not shown.

Table 2

<u>Benzones</u> Octabenzene Benzophenone	<u>Neonicotinoids</u> Thiamethoxam Thiacloprid Imidacloprid Acetamiprid Clothianidin	<u>Parabens</u> Butylparaben Propylparaben Methylparaben	<u>Phthalates</u> Dimethyl phthalate Diallyl phthalate Diethyl phthalate Bis(2-ethylhexyl) nonanedioate Dibutyl phthalate Di(2-ethylhexyl) adipate Benzyl butyl phthalate Diethyl phthalate Dibutyl hexanedioate
<u>Carbamates</u> Bendiocarb Oxamyl Carbaryl Chlopropham Fenoxycarb	<u>Organochlorides</u> p,p'-DDT Mirex Lindane Chlorothalonil Pentachloronitrobenzene	<u>Phenols</u> 2,2'-Methylenebis(ethyl-6-tert-butylphenol) Pentachlorophenol 4-(1,1,3,3-Tetramethylbutyl)phenol Phenol Bisphenol A Resorcinol 2,6-Di-tert-butylphenol 3,3',5,5'-Tetrabromo bisphenol A	<u>Pyrethroids</u> Cyfluthrin Fenpropathrin S-Bioallethrin Permethrin
<u>Conazoles</u> Difenoconazole Fenbuconazole Cyproconazole Triadimenol Propiconazole	<u>Organophosphates</u> Phorate Diazinon Methyl parathion Chlorpyrifos Chlorpyrifos-methyl Malathion Bensulide Trichlorfon	<u>Phenoxies</u> Diclofop-methyl 2,4-Dichlorophenoxyacetic acid 2,4-DB Triclosan MCPA Fluroxypyr	<u>Xanthines</u> Theophylline Caffeine
<u>Genols</u> Methyleugenol Eugenol			

Most of the 64 chemicals are relevant to human exposures and used in food and/or consumer products. For example, benzophenone is used in sunscreens, eugenols is used in perfumes, and caffeine (a xanthine) is used in beverages. Several of the 13 selected classes are commonly used in agricultural products (such as herbicides or insecticides), and residues of them may be found in food products.

FIG. 5 is a comparison of POD values estimated based on an exemplary embodiment of the system or method against the POD values obtained from ToxCast assays. White circles represent chemicals that were found to be “active” by the exemplary embodiment, while black circles represent chemicals found to be “inactive” by the exemplary embodiment. R is the Pearson’s correlation coefficient between PODs derived based on the exemplary embodiment and ToxCast.

For the comparison, *in vitro* bioactivity values of the 64 chemicals described for FIG. 4 above were obtained from the public ToxCast database (version: October 2015), available at <https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data>. The ToxCast database consists of 1192 endpoints from 359 unique toxicity assays. For each chemical, the AC₅₀ values of endpoints that were only tested at one concentration and found to be “inactive” were assigned to an arbitrary large number, e.g. 1000 µM. Otherwise, the given AC₅₀ values of the endpoints were used. If the number of tested endpoints is less than or equal to three, the POD value of the chemical was set to the minimum of the AC₅₀ values of the tested endpoints. Otherwise, the POD value of the chemical was set to the 6th percentile of the AC₅₀ values of all the tested endpoints.

As shown, the POD values estimated based on an exemplary embodiment of the system or method (“HIPPTox-PODs”) are highly and positively correlated to the POD values obtained from ToxCast assays (“ToxCast-PODs”) (Pearson’s correlation coefficient = 0.62, $P = 1.08 \times 10^{-4}$). See A of FIG. 5. Furthermore, most of the POD values are near the diagonal line, indicating that the values are also numerically close to each other. When the ToxCast assays were separated into assays based on human cells/tissues (see B) and assays based on non-human cells/tissues (see C), HIPPTox-PODs were found to be positively correlated to ToxCast-PODs obtained based on human assays (Pearson’s correlation coefficient = 0.57, $P = 6.56 \times 10^{-4}$), but not as significantly correlated to ToxCast-PODs that were based on non-human assays (Pearson’s correlation coefficient = 0.23, $P = 0.19$). However, it is recognised that the highest correlation was found between HIPPTox-PODs and the ToxCast-PODs based on all ToxCast assays (Pearson’s correlation coefficient = 0.62). The results show that embodiments of the system/method for determining a POD of a chemical has a biological coverage that is approximately similar to ToxCast, and can be used to approximate the PODs derived from the ToxCast assays.

FIG. 6 is a flowchart of a method 600 for determining a POD of a chemical in an exemplary embodiment.

5 In the exemplary embodiment, at step 602, one or more features obtained from images of one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type are measured. At step 604, one or more features obtained from images of one or more control samples having been treated with a solvent control at one or a plurality of control
10 concentrations and being obtained from a control population of the first cell type are measured. At step 606, each sample of the one or more test samples at each of the plurality of test concentrations and the one or more control samples at the one or each of the plurality of control concentrations is classified based on measured feature values. At
15 step 608, a first concentration-accuracy curve is formulated based on the step of classifying each sample. At step 610, a first initial POD value of the chemical is derived based on the first concentration-accuracy curve.

 In various embodiments, classifying each sample comprises applying a supervised classifier to classify cells of the one or more test samples treated at each of the plurality
20 of test concentrations and the one or more control samples treated at the one or each of the plurality of concentrations into an assigned "treated" class or an assigned "control" class based on the measured features.

 In various embodiments, to obtain a classification accuracy value, the following is
25 performed: a number of test samples (P) and a number of control sample (N) is counted, a number of correctly classified test samples (TP) and a number of correctly classified control samples (TN) is counted, and a balanced accuracy value based on $(TP/P + TN/N)/2$ is set.

30 In various embodiments, formulating a first concentration-accuracy curve comprises fitting a plurality of statistical models to a first classification result of the classifying step and determining a preferred fit.

In various embodiments, the statistical models comprise at least one of a log-logistic model, a logistic model, a Weibull model, a gamma model, and a constant model may be used.

5 In various embodiments, the method further comprises labeling the chemical as “inactive” if the preferred fit is based on a constant model and labeling the chemical as “active” if the preferred fit is not a constant model.

10 In various embodiments, the method further comprises repeating steps 602, 604 and 606 for a second test population and for a second control population and formulating at least a second concentration-accuracy curve based on the second test population and the second control population to derive a second initial POD value, and obtaining an average initial POD value based on the second initial POD value.

15 In various embodiments, the method further comprises repeating one or more of steps 602, 604, 606 608 and 610 for a different cell type to obtain a different initial POD value that is associated with said different cell type.

20 In various embodiments, the method further comprises determining the POD of the chemical based on the different initial POD value.

25 Different exemplary embodiments can be implemented in the context of data structure, program modules, program and computer instructions executed in a computer implemented environment. A general purpose computing environment is briefly disclosed herein. FIG. 7 is a schematic drawing of a computer system suitable for implementing an exemplary embodiment of the system or method.

30 One or more exemplary embodiments may be implemented as software, such as a computer program being executed within a computer system 700, and instructing the computer system 700 to conduct a method of an exemplary embodiment.

The computer system 700 comprises a computer unit 702, input modules such as a keyboard 704 and a pointing device 706 and a plurality of output devices such as a display

708, and printer 710. A user can interact with the computer unit 702 using the above devices. The pointing device can be implemented with a mouse, track ball, pen device or any similar device. One or more other input devices (not shown) such as a joystick, game pad, satellite dish, scanner, touch sensitive screen or the like can also be connected to the computer unit 702. The display 708 may include a cathode ray tube (CRT), liquid crystal display (LCD), field emission display (FED), plasma display or any other device that produces an image that is viewable by the user.

The computer unit 702 can be connected to a computer network 712 via a suitable transceiver device 714, to enable access to e.g. the Internet or other network systems such as Local Area Network (LAN) or Wide Area Network (WAN) or a personal network. The network 712 can comprise a server, a router, a network personal computer, a peer device or other common network node, a wireless telephone or wireless personal digital assistant. Networking environments may be found in offices, enterprise-wide computer networks and home computer systems etc. The transceiver device 714 can be a modem/router unit located within or external to the computer unit 702, and may be any type of modem/router such as a cable modem or a satellite modem.

It will be appreciated that network connections shown are exemplary and other ways of establishing a communications link between computers can be used. The existence of any of various protocols, such as TCP/IP, Frame Relay, Ethernet, FTP, HTTP and the like, is presumed, and the computer unit 702 can be operated in a client-server configuration to permit a user to retrieve web pages from a web-based server. Furthermore, any of various web browsers can be used to display and manipulate data on web pages.

The computer unit 702 in the example comprises a processor 718, a Random Access Memory (RAM) 720 and a Read Only Memory (ROM) 722. The ROM 722 can be a system memory storing basic input/output system (BIOS) information. The RAM 720 can store one or more program modules such as operating systems, application programs and program data.

The computer unit 702 further comprises a number of Input/Output (I/O) interface units, for example I/O interface unit 724 to the display 708, and I/O interface unit 726 to the keyboard

704. The components of the computer unit 702 typically communicate and interface/couple connectedly via an interconnected system bus 728 and in a manner known to the person skilled in the relevant art. The bus 728 can be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures.

It will be appreciated that other devices can also be connected to the system bus 728. For example, a universal serial bus (USB) interface can be used for coupling a video or digital camera to the system bus 728. An IEEE 1394 interface may be used to couple additional devices to the computer unit 702. Other manufacturer interfaces are also possible such as FireWire developed by Apple Computer and i.Link developed by Sony. Coupling of devices to the system bus 728 can also be via a parallel port, a game port, a PCI board or any other interface used to couple an input device to a computer. It will also be appreciated that, while the components are not shown in the figure, sound/audio can be recorded and reproduced with a microphone and a speaker. A sound card may be used to couple a microphone and a speaker to the system bus 728. It will be appreciated that several peripheral devices can be coupled to the system bus 728 via alternative interfaces simultaneously.

An application program can be supplied to the user of the computer system 700 being encoded/stored on a data storage medium such as a CD-ROM or flash memory carrier. The application program can be read using a corresponding data storage medium drive of a data storage device 730. The data storage medium is not limited to being portable and can include instances of being embedded in the computer unit 702. The data storage device 730 can comprise a hard disk interface unit and/or a removable memory interface unit (both not shown in detail) respectively coupling a hard disk drive and/or a removable memory drive to the system bus 728. This can enable reading/writing of data. Examples of removable memory drives include magnetic disk drives and optical disk drives. The drives and their associated computer-readable media, such as a floppy disk provide nonvolatile storage of computer readable instructions, data structures, program modules and other data for the computer unit 702. It will be appreciated that the computer unit 702 may include several of such drives. Furthermore, the computer unit 702 may include drives for interfacing with other types of computer readable media.

The application program is read and controlled in its execution by the processor 718. Intermediate storage of program data may be accomplished using RAM 720. The method(s) of the exemplary embodiments can be implemented as computer readable instructions, computer executable components, or software modules. One or more software modules may alternatively be used. These can include an executable program, a data link library, a configuration file, a database, a graphical image, a binary data file, a text data file, an object file, a source code file, or the like. When one or more computer processors execute one or more of the software modules, the software modules interact to cause one or more computer systems to perform according to the teachings herein.

The operation of the computer unit 702 can be controlled by a variety of different program modules. Examples of program modules are routines, programs, objects, components, data structures, libraries, etc. that perform particular tasks or implement particular abstract data types. The exemplary embodiments may also be practiced with other computer system configurations, including handheld devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, personal digital assistants, mobile telephones and the like. Furthermore, the exemplary embodiments may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a wireless or wired communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

The description herein may be, in certain portions, explicitly or implicitly described as algorithms and/or functional operations that operate on data within a computer memory or an electronic circuit. These algorithmic descriptions and/or functional operations are usually used by those skilled in the information/data processing arts for efficient description. An algorithm is generally relating to a self-consistent sequence of steps leading to a desired result. The algorithmic steps can include physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transmitted, transferred, combined, compared, and otherwise manipulated.

Further, unless specifically stated otherwise, and would ordinarily be apparent from the following, a person skilled in the art will appreciate that throughout the present

specification, discussions utilizing terms such as “measuring”, “computing”, “calculating”, “determining”, “deriving”, “classifying”, “replacing”, “generating”, “formulating”, “initializing”, “outputting”, and the like, refer to action and processes of an instructing processor/computer system, or similar electronic circuit/device/component, that manipulates/processes and transforms data represented as physical quantities within the described system into other data similarly represented as physical quantities within the system or other information storage, transmission or display devices etc. Furthermore, unless specifically stated otherwise and as would be appreciated, when the term “imaging” is used, it refers to action and process of an image capturing module for capturing image data, such as a camera (e.g. a microscope camera), that may be operatively connected to or controlled by the instructing processor/computer system, or similar electronic circuit/device/component or operatively separate from the instructing processor/computer system, or similar electronic circuit/device/component.

The description also discloses relevant device/apparatus for performing the steps of the described methods or relevant device/apparatus for implementing the described systems. Such apparatus may be specifically constructed for the purposes of the methods/systems, or may comprise a general purpose computer/processor and/or other device selectively activated or reconfigured by a computer program stored in a storage member. The algorithms and displays described herein are not inherently related to any particular computer or other apparatus. It is understood that general purpose devices/machines may be used in accordance with the teachings herein. Alternatively, the construction of a specialized device/apparatus to perform the method steps may be desired.

In addition, it is submitted that the description also implicitly covers a computer program, in that it would be clear that the steps of the methods or the systems described herein may be put into effect by computer code. It will be appreciated that a large variety of programming languages and coding can be used to implement the teachings of the description herein. Moreover, the computer program if applicable is not limited to any particular control flow and can use different control flows without departing from the scope of the invention.

Furthermore, one or more of the steps of the computer program if applicable may be performed in parallel and/or sequentially. Such a computer program if applicable may be

stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a suitable reader/general purpose computer. In such instances, the computer readable storage medium is non-transitory. Such storage medium also covers
5 all computer-readable media e.g. medium that stores data only for short periods of time and/or only in the presence of power, such as register memory, processor cache and Random Access Memory (RAM) and the like. The computer readable medium may even include a wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in bluetooth technology. The computer program when loaded and executed on a suitable
10 reader effectively results in an apparatus that can implement the steps of the described methods.

The exemplary embodiments may also be implemented as hardware modules. A module is a functional hardware unit designed for use with other components or modules. For
15 example, a module may be implemented using digital or discrete electronic components, or it can form a portion of an entire electronic circuit such as an Application Specific Integrated Circuit (ASIC). A person skilled in the art will understand that the exemplary embodiments can also be implemented as a combination of hardware and software modules.

20 Additionally, when describing some embodiments, the disclosure may have disclosed a method and/or process (e.g. a process implemented by a system) as a particular sequence of steps. However, unless otherwise required, it will be appreciated the method or process should not be limited to the particular sequence of steps disclosed. Other sequences of steps may be possible. The particular order of the steps disclosed herein should not be construed
25 as undue limitations. Unless otherwise required, a method and/or process disclosed herein should not be limited to the steps being carried out in the order written. The sequence of steps may be varied and still remain within the scope of the disclosure.

The term "point of departure" or "POD" as used herein refers to a quantitative indicator
30 of toxicity of a chemical that is derived from a concentration accuracy curve. It is akin to the traditional point of departure which is the point on a dose-response curve established from experimental data or observational data generally corresponding to an estimated low effect, or no effect level. Thus, the "point of departure" determined herein can be used as a starting

point for later extrapolations and analyses such as extrapolation to the Reference Dose (RfD) used for risk assessment. In exemplary embodiments, the "point of departure" may correspond to the EC₁₀ value derived from a concentration accuracy curve or an average EC₁₀ value based on multiple concentration accuracy curves for an "active" chemical. In exemplary
5 embodiments, the "point of departure" corresponds to a null value for an "inactive" chemical.

The terms "coupled" or "connected" as used in this description are intended to cover both directly connected or connected through one or more intermediate means, unless
10 otherwise stated.

Further, in the description herein, the word "substantially" whenever used is understood to include, but not restricted to, "entirely" or "completely" and the like. In addition, terms such as "comprising", "comprise", and the like whenever used, are intended to be non-restricting descriptive language in that they broadly include elements/components recited after
15 such terms, in addition to other components not explicitly recited. For an example, when "comprising" is used, reference to a "one" feature is also intended to be a reference to "at least one" of that feature. Terms such as "consisting", "consist", and the like, may, in the appropriate context, be considered as a subset of terms such as "comprising", "comprise", and the like. Therefore, in embodiments disclosed herein using the terms such as "comprising", "comprise",
20 and the like, it will be appreciated that these embodiments provide teaching for corresponding embodiments using terms such as "consisting", "consist", and the like. Further, terms such as "about", "approximately" and the like whenever used, typically means a reasonable variation, for example a variation of +/- 5% of the disclosed value, or a variance of 4% of the disclosed value, or a variance of 3% of the disclosed value, a variance of 2% of the disclosed value or
25 a variance of 1% of the disclosed value.

Furthermore, in the description herein, certain values may be disclosed in a range. The values showing the end points of a range are intended to illustrate a preferred range. Whenever a range has been described, it is intended that the range covers and teaches all
30 possible sub-ranges as well as individual numerical values within that range. That is, the end points of a range should not be interpreted as inflexible limitations. For example, a description of a range of 1% to 5% is intended to have specifically disclosed sub-ranges 1% to 2%, 1% to 3%, 1% to 4%, 2% to 3% etc., as well as individually, values within that range such as 1%,

2%, 3%, 4% and 5%. The intention of the above specific disclosure is applicable to any depth/breadth of a range.

It will be appreciated by a person skilled in the art that other variations and/or modifications may be made to the specific embodiments without departing from the scope of the invention as broadly described. For example, in the description herein, features of different exemplary embodiments may be mixed, combined, interchanged, incorporated, adopted, modified, included etc. or the like across different exemplary embodiments. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

APPLICATIONS

Embodiments of the system for determining a POD of a chemical and embodiments of the method for determining a POD of a chemical possess several unique features/advantages as compared to existing systems/methods for determining a POD of a chemical.

Firstly, embodiments of the method/system may integrate information from a large number of molecular and/or phenotypic readouts at the readout level, and generates a single POD from the integrated information. In contrast, most existing POD assessment methods determine a single POD for each measured readout, even if multiple readouts may be measured from the same biological samples. These existing methods process each readout independently to derive a POD, and do not integrate information at the readout level. These existing methods usually only integrate or average the resulting PODs after the PODs from each readout have been computed.

An example of an existing system/method is the United State Environmental Protection Agency (US EPA)'s Toxicity Forecaster (ToxCast) Program that used ~300-1000 in vitro assays with more than 1000 bioactivity readouts to test more than 1000 reference chemicals. For each readout from an assay, ToxCast fitted the measured response data for a chemical with a concentration response curve, and determined three different POD estimates, namely activity concentration at 10% of maximal activity (AC_{10}), activity concentration at baseline (ACB), and activity concentration at cutoff (ACC), from the fitted curve

(<https://www.epa.gov/chemical-research/toxcast-data-generation-toxcast-pipeline-tcpl>).

About 300-1000 POD values were generated for each chemical. These values may be summarized by computing their mean, 5th-percentile, or minimum value.

5 Another example is a previous method that determined BMDs based on high-dimensional gene expression data (>20,000 genes) collected from rats exposed to a chemical. For every significantly changed readout (in this case, the RNA expression level of a gene), the proposed method fitted the measured response data with a dose response curve, and determined a BMD value from the fitted curve. Hundreds to thousands of BMD values were
10 determined for all the changed genes in the same rat tissue samples. Then, the genes were divided according to different Gene Ontology (GO) groups, and an average POD value was computed for every group.

 Another further example is a previous method that assessed liver toxicity based on five
15 phenotypic readouts measured from microscopy images of human liver cells exposed to a chemical. Again, for every phenotypic readout, the proposed method fitted the measured response data with a concentration response curve, and determined a half maximal inhibitory concentration (IC₅₀) value from the fitted curve. Five IC₅₀ values were determined for the five phenotypic readouts measured from the same liver cell samples. Then, the lowest IC₅₀ value
20 was determined from these values.

 Such existing methods have several shortcomings. Typically, a single molecular and/or phenotypic readout covers only limited and specific biological pathways or mechanisms. Furthermore, certain phenotypic responses to a chemical may be different at low
25 concentrations of the chemical and at high concentrations of the chemicals. Thus, readouts that can detect responses at high concentrations may not be able to detect responses at low concentrations. Therefore, such methods that process each readout independently may overestimate the PODs values of a chemical.

30 In contrast, embodiments of the method/system may overcome the above problems associated with existing methods by using a multi-dimensional supervised classifier that is capable of processing all the readouts simultaneously and determining differences in the multi-dimensional readout space that can distinguish between chemically-treated and control

samples at each tested concentration. Therefore, embodiments of the method/system integrate information at the readout level, and may capture broader biological information from the measured response data than existing methods.

5 Secondly, embodiments of the method/system do not require any reference or training chemicals with known toxicity annotations or labels. Many existing *in vitro* toxicity prediction methods rely on such reference or training chemicals to train their classifiers. However, for many adverse effects, very few chemicals are known or have been validated to be able to directly induce such effects in humans. Thus, these previous methods cannot be applied to
10 build models for these adverse effects. In contrast, embodiments of the method/system make use of only measurements from control (solvent-treated) and chemical-treated cells. This is because in embodiments of the method/system, a classifier is trained to distinguish between these two conditions. Therefore, embodiments of the method/system can usefully be applied to study toxicity or adverse effects that have no or very few known reference chemicals.

15 Thirdly, embodiments of the method/system are also distinguished from existing methods in that a supervised classifier is used to derive a “concentration-accuracy curve” from molecular/phenotypic readouts. In embodiments of the method/system, the y-axis of the curve is classification accuracy that describes/indicates the magnitude of the chemical effects in the
20 measured multi-dimensional bioactivity readouts. It is recognised that almost all existing *in vitro* assays (including those used by the ToxCast database) are based instead on concentration-response curves with the y-axis being the magnitude of a specific bioactivity readout.

25 Fourthly, embodiments of the method/system are capable of detecting/analysing a large number of readouts to determine a POD based on relatively few imaging assays. In one embodiment, only three imaging assays are used to obtain 165 readouts/features. By comparison, ToxCast uses more than 300 unique toxicity assays (more than 100 times that of
30 embodiments of the method/system) to obtain more 1000 endpoints (less than 10 times more that of embodiments of the method/system). Embodiments of the method/system are therefore more efficient and cost-effective as compared to existing methods such as ToxCast. Indeed, although the per-chemical cost (both in terms of money and time) of constructing the ToxCast database for ~1000 reference chemicals is lower than the cost needed to run animal tests for

the same set of chemicals, the per-chemical cost to test a small number (~1-10) of new chemicals using all the ~300-1000 ToxCast assays is prohibitively high. Thus, in many practical scenarios, such as when evaluating the safety of a few novel chemical ingredients to be used in a product, the ToxCast database may provide limited or no direct information on the PODs of these chemicals. In such scenarios, it is also not practical to run all the ToxCast assays on these few chemicals. Under these circumstances, embodiments of the method/system may be used to estimate the PODs at a much lower cost and faster turn-around time.

Overall, embodiments of the system for determining a POD of a chemical and embodiments of the method for determining a POD of a chemical provide an *in vitro* and/or *in silico* system/method that does not necessitate the use of animals or any reference or training chemicals with known toxicity information or annotations. In addition, embodiments of the system/method are highly efficient because they integrate information from a large number of molecular and/or phenotypic readouts at the readout level, and thus require a smaller number of assays or experiments than existing methods that process and analyze each readout independently. Notwithstanding, embodiments of the system/method provide a biological coverage that is similar to existing methods such as ToxCast. Embodiments of the system/method are thus cost efficient and feasible even for small-scale and/or rapid tests of chemicals. Embodiments of the system/method are recognised to be suitable for use in the fields of toxicology, environmental health, chemistry, and pharmacology, such as to facilitate the assessment of the risk of chemicals to human health.

CLAIMS

1. A system for determining a point of departure (POD) of a chemical, the system comprising:

5 a feature measurement module arranged to measure one or more features obtained from images of (a) one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type, and (b) one or more control samples having been treated with a solvent control at one or a plurality of control concentrations and
10 being obtained from a control population of the first cell type;

a classification module arranged to classify each sample of the one or more test samples at each of the plurality of test concentrations and the one or more control samples at the one or each of the plurality of control concentrations based on feature values generated from the feature measurement module;

15 a curve formulation module arranged to formulate a first concentration-accuracy curve based on a classification result of the classification module; and

a POD determination module arranged to derive a first initial POD value of the chemical based on the first concentration-accuracy curve.

20 2. The system of claim 1, further comprising the classification module being configured to apply a supervised classifier to classify cells of the one or more test samples treated at each of the plurality of test concentrations and the one or more control samples treated at the one or each of the plurality of control concentrations into an assigned "treated" class or an assigned "control" class based on the one
25 or more features measured from the feature measurement module.

3. The system of claim 2, wherein the classification module is further arranged to provide a classification accuracy value based on counting a number of test samples (P) and a number of control sample (N); counting a number of correctly
30 classified test samples (TP) and a number of correctly classified control samples (TN); and setting a balanced accuracy value based on $(TP/P + TN/N)/2$.

4. The system of claim 2 or claim 3, wherein the supervised classifier comprises a Support Vector Machine.
5. The system of any of the preceding claims, further comprising the curve
5 formulation module being configured to fit a plurality of statistical models to a first classification result of the classification module and to determine a preferred fit to obtain the first concentration-accuracy curve.
6. The system of claim 5, wherein the plurality of statistical models comprise at least
10 one of a log-logistic model, a logistic model, a Weibull model, a gamma model, and a constant model.
7. The system of claim 5 or claim 6, further comprising the POD determination
15 module being arranged to label the chemical as "inactive" if the preferred fit is based on a constant model and to label the chemical as "active" if the preferred fit is not a constant model.
8. The system of any of the preceding claims, further comprising:
20 the feature measurement module being arranged to measure one or more features obtained from images of (a) one or more test samples having been treated with the chemical at the plurality of test concentrations and being obtained from a second test population of the first cell type, and (b) one or more control samples having been treated with the solvent control at the one or plurality of control concentrations and being obtained from a second control population of the first cell
25 type;
the classification module being arranged to classify each sample of the one or more test samples from the second test population at each of the plurality of test concentrations and the one or more control samples from the second control population at the one or each of the plurality of control concentrations;
30 the curve formulation module being arranged to formulate a second concentration-accuracy curve based on a second classification result of the classification module; and

the POD determination module being arranged to derive a second initial POD value of the chemical based on the second concentration-accuracy curve and to obtain an average initial POD value based on the second initial POD value.

- 5 9. The system of any of the preceding claims, further comprising:
 the feature measurement module being arranged to measure one or more features obtained from images of (a) one or more test samples having been treated with the chemical at the plurality of test concentrations and being obtained from a test population of a different cell type, and (b) one or more control samples having
10 been treated with the solvent control at the one or plurality of control concentrations and being obtained from a control population of the different cell type;
 the classification module being arranged to classify each sample of the one or more test samples from the test population of a different cell type at each of the plurality of test concentrations and the one or more control samples from the
15 control population of the different cell type at the one or each of the plurality of control concentrations;
 the curve formulation module being arranged to formulate a different concentration-accuracy curve that is associated with said different cell type based on a third classification result of the classification module; and
20 the POD determination module being arranged to derive a different initial POD value that is associated with said different cell type based on the different concentration-accuracy curve.
- 25 10. The system of claim 9, further comprising the POD determination module being configured to determine the POD of the chemical based on the different initial POD value.
- 30 11. The system of any of the preceding claims, wherein the first cell type comprises at least one of a lung cell, a lung-derived cell, a lung-derived cell line cell, a liver cell, a liver-derived cell, a liver-derived cell line cell, a kidney cell, a kidney-derived cell or a kidney-derived cell line cell.
12. A method of determining a POD of a chemical, the method comprising:

(i) measuring one or more features from images of one or more test samples having been treated with the chemical at a plurality of test concentrations and being obtained from a test population of a first cell type;

(ii) measuring one or more features from images of one or more control samples having been treated with a solvent control at one or a plurality of control concentrations and being obtained from a control population of the first cell type;

(iii) classifying each sample of the one or more test samples at each of the plurality of test concentrations and the one or more control samples at the one or each of the plurality of control concentrations based on measured feature values;

(iv) formulating a first concentration-accuracy curve based on the step of classifying each sample; and

(v) deriving a first initial POD value of the chemical based on the first concentration-accuracy curve.

13. The method of claim 12, wherein classifying each sample comprises applying a supervised classifier to classify cells of the one or more test samples treated at each of the plurality of test concentrations and the one or more control samples treated at the one or each of the plurality of control concentrations into an assigned “treated” class or an assigned “control” class based on the measured features.

14. The method of claim 13, further comprising obtaining a classification accuracy value based on:

counting a number of test samples (P) and a number of control samples (N);

counting a number of correctly classified test samples (TP) and a number of correctly classified control samples (TN); and

setting a balanced accuracy value based on $(TP/P + TN/N)/2$.

15. The method of any of claims 12 to 14, wherein formulating a first concentration-accuracy curve comprises fitting a plurality of statistical models to a first classification result of the classifying step and determining a preferred fit.

16. The method of claim 15, wherein the plurality of statistical models comprise at least one of a log-logistic model, a logistic model, a Weibull model, a gamma model, and a constant model.
- 5 17. The method of claim 15 or claim 16, further comprising labeling the chemical as “inactive” if the preferred fit is based on a constant model and labeling the chemical as “active” if the preferred fit is not a constant model.
- 10 18. The method of any of claims 12 to 17, further comprising repeating steps (i) to (iii) for a second test population and for a second control population and formulating at least a second concentration-accuracy curve based on the second test population and the second control population to derive a second initial POD value, and obtaining an average initial POD value based on the second initial POD value.
- 15 19. The method of any of claims 12 to 18, further comprising repeating steps (i) to (v) for a different cell type to obtain a different initial POD value that is associated with said different cell type.
- 20 20. The method of claim 19, further comprising determining the POD of the chemical based on the different initial POD value.
- 25 21. The method of any of claims 12 to 20, wherein the first cell type comprises at least one of a lung cell, a lung-derived cell, a lung-derived cell line cell, a liver cell, a liver-derived cell, a liver-derived cell line cell, a kidney cell, a kidney-derived cell or a kidney-derived cell line cell.

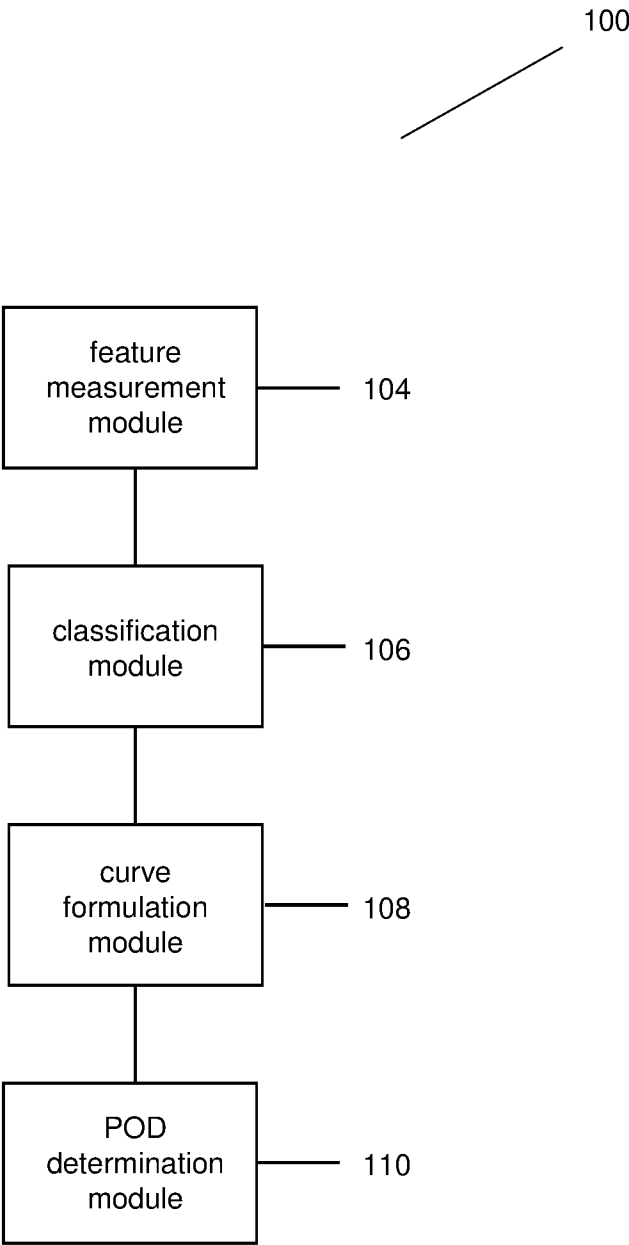


FIG. 1

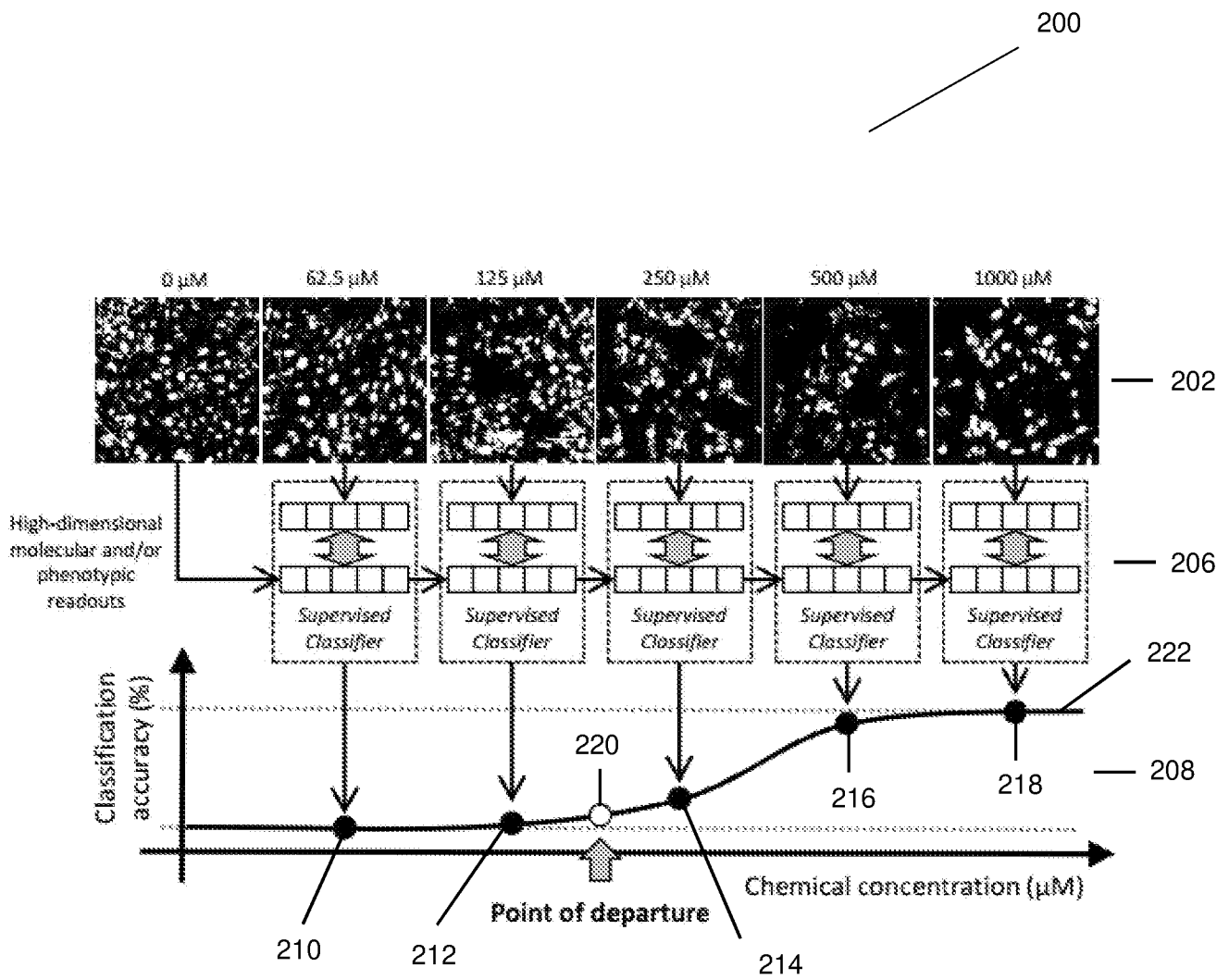


FIG. 2

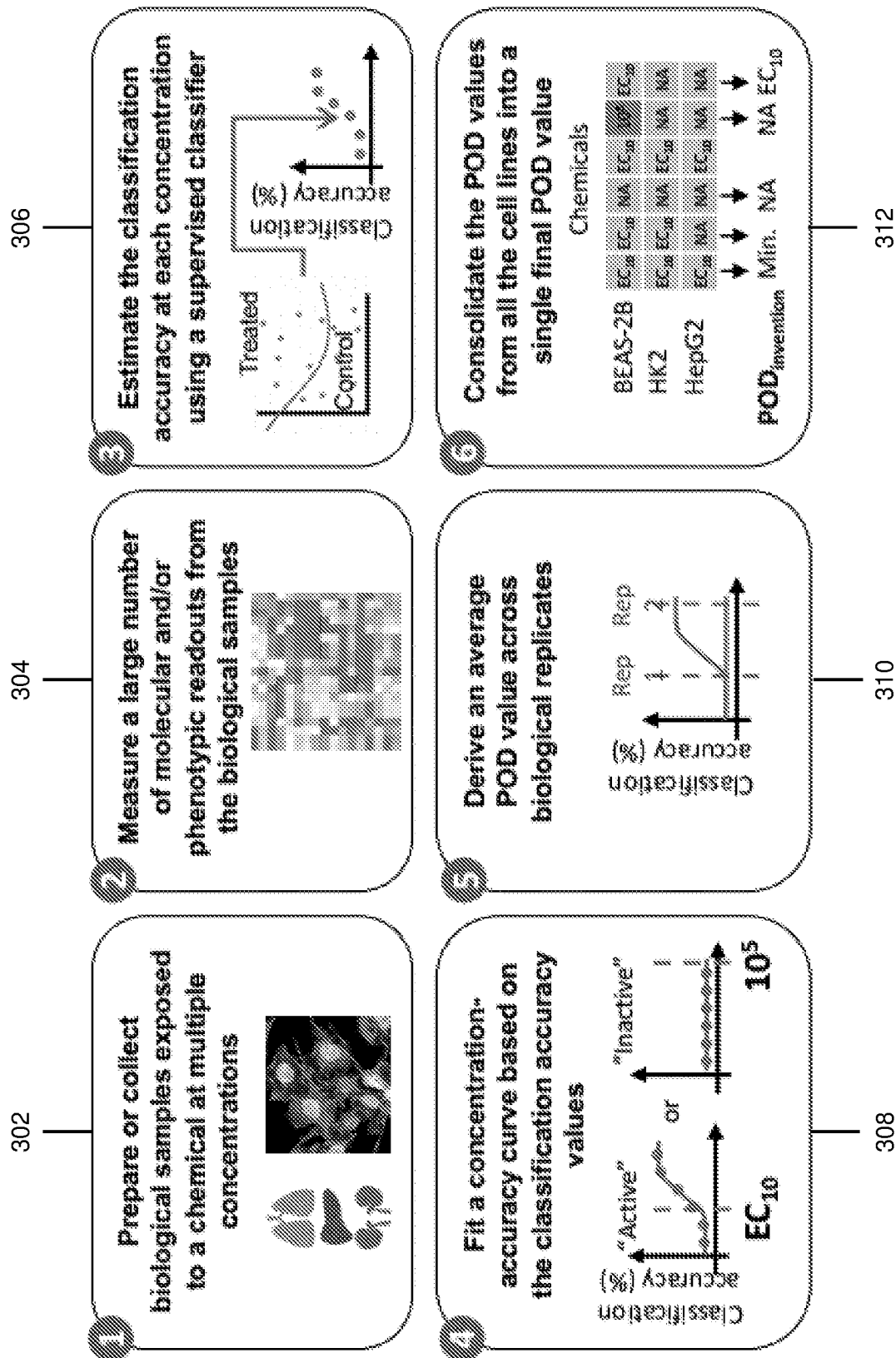


FIG. 3

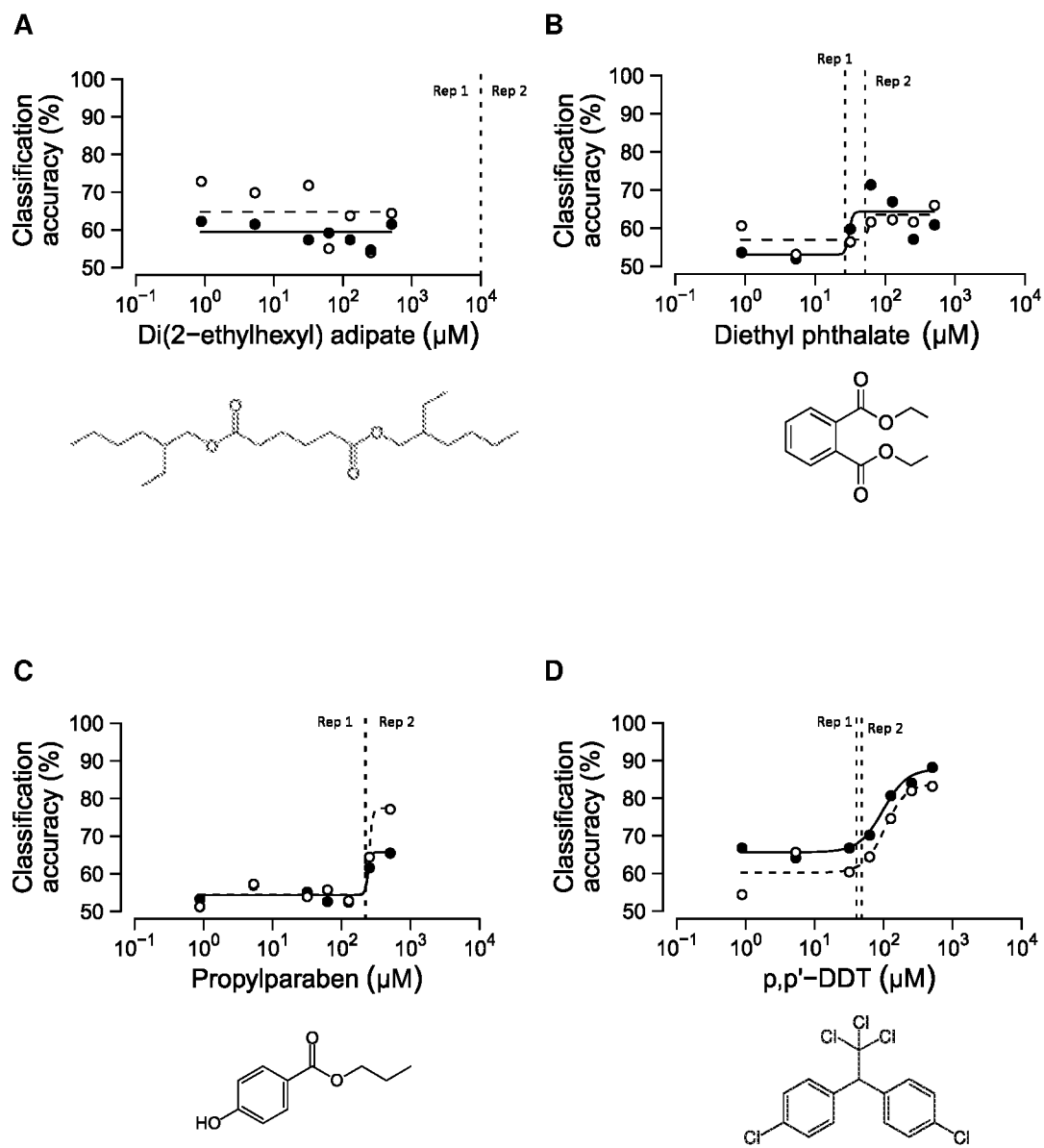


FIG. 4

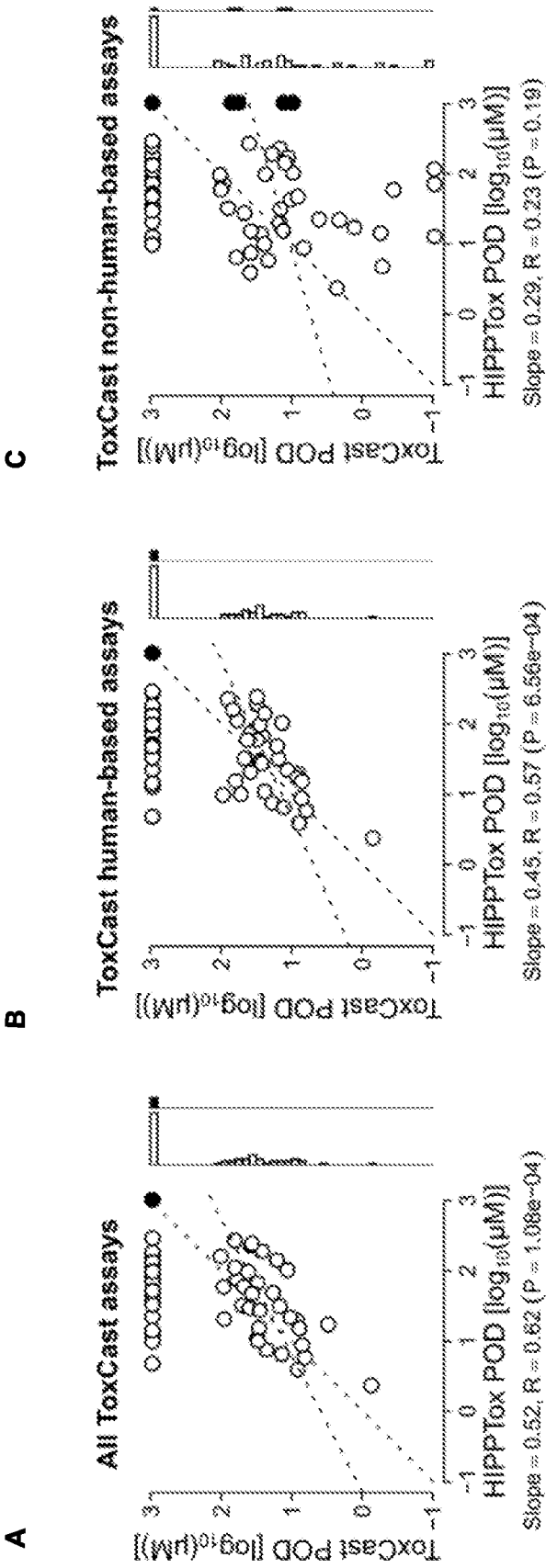


FIG. 5

6/7

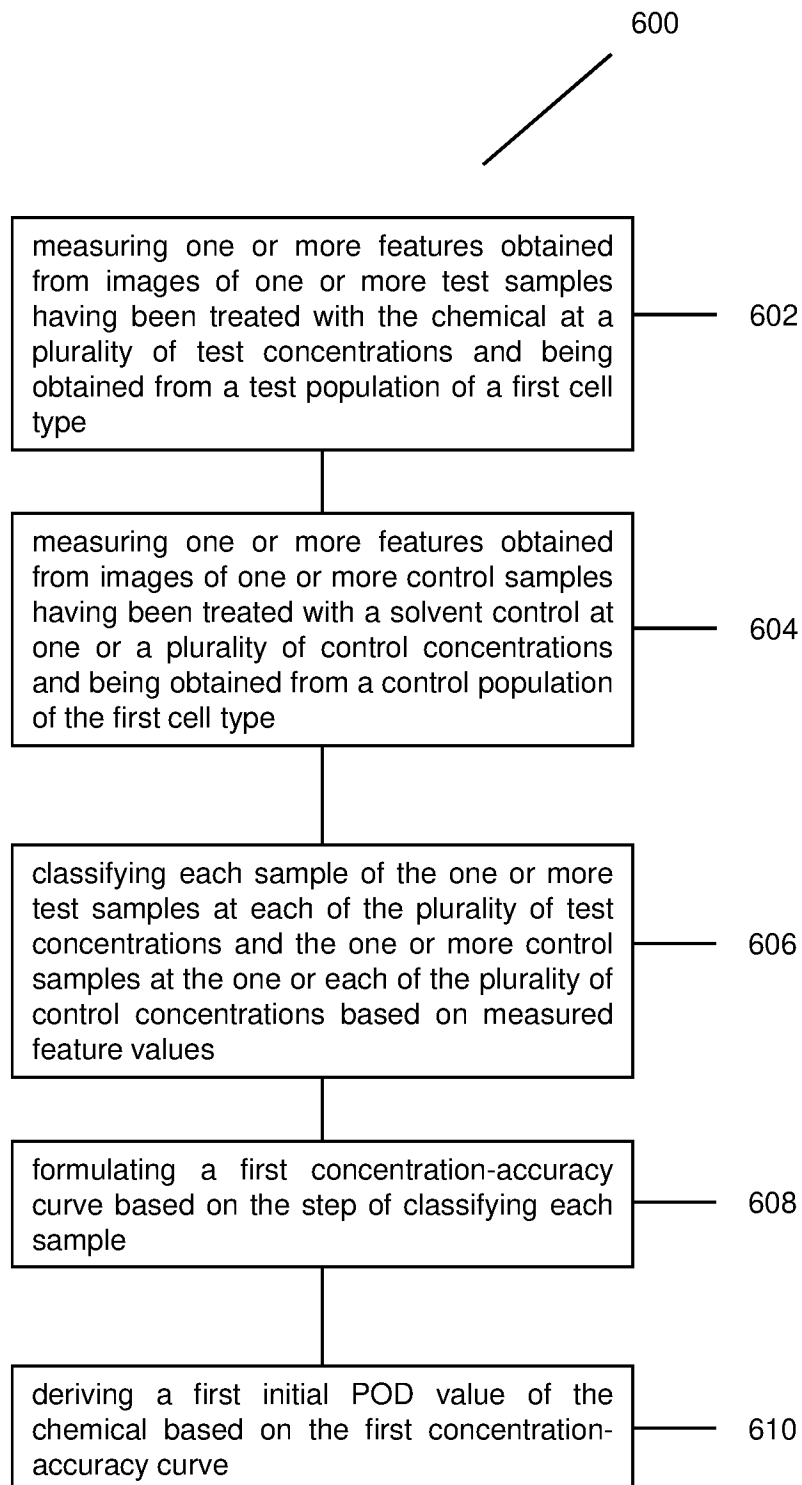


FIG. 6

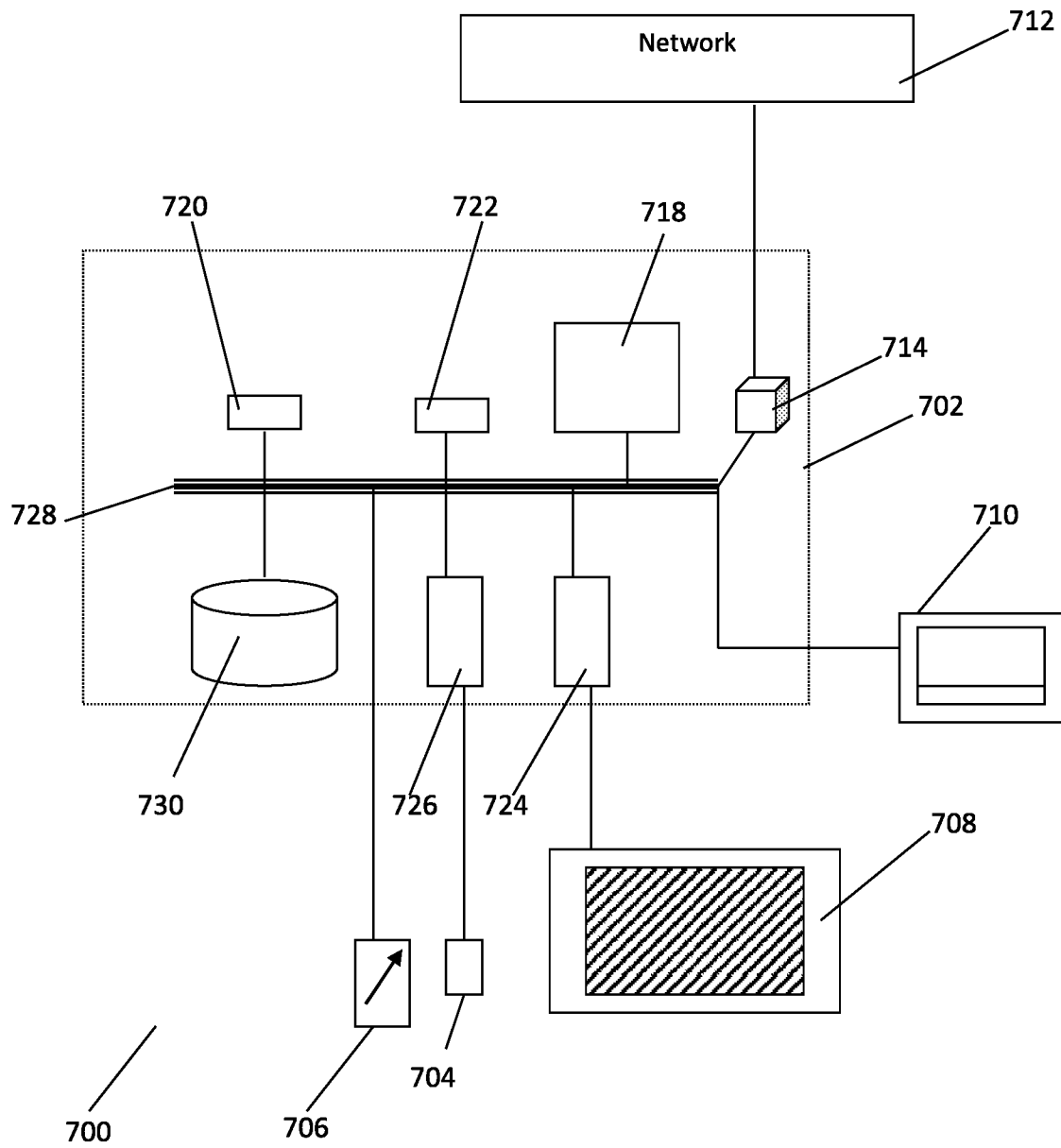


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2019/050314

A. CLASSIFICATION OF SUBJECT MATTER

G16B 40/20 (2019.01) G16C 20/70 (2019.01) G06N 20/10 (2019.01)

According to International Patent Classification (IPC)

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16B, G16C

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CAPLUS, EMBASE, MEDLINE, BIOSIS, SCISEARCH, FAMPAT: point of departure, classification, concentration, image, high dimension, *in vitro*, confusion, dissimilarity, and the related terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LOO L.H., ET AL., Image-based multivariate profiling of drug responses from single cells. <i>Nature Methods</i> , 1 April 2007, Vol. 4, No. 5, pages 445-453 [Retrieved on 2019-07-25] <DOI: 10.1038/NMETH1032> Fig. 1; page 455 right column – page 466 right column	1-21
A	US 2017/0270239 A1 (GRAFSTROM R., ET AL.) 21 September 2017 Whole document, in particular Fig. 11	--
A	LEE J. J., ET AL., Building predictive in vitro pulmonary toxicity assays using high-throughput imaging and artificial intelligence. <i>Archives of Toxicology</i> , 28 April 2018, Vol. 92, No. 6, pages 2055-2075 [Retrieved on 2019-07-25] <DOI: 10.1007/S00204-018-2213-0> Whole document	--

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

*Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

25/07/2019 (day/month/year)

Date of mailing of the international search report

29/08/2019 (day/month/year)

Name and mailing address of the ISA/SG

IPOS Intellectual Property Office of Singapore
51 Bras Basah Road
#01-01 Manulife Centre
Singapore 189554

Email: pct@ipos.gov.sg

Authorized officer

Cao Ye (Dr)

IPOS Customer Service Tel. No.: (+65) 6339 8616

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2019/050314

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JUDSON, R., ET AL., A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. <i>BMC Bioinformatics</i> , 19 May 2008, Vol. 9, No. 1, pages 241 [Retrieved on 2019-07-25] <DOI: 10.1186/1471-2105-9-241> Whole document	--

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2019/050314

Note: This Annex lists known patent family members relating to the patent documents cited in this International Search Report. This Authority is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2017/0270239 A1	21/09/2017	EP 3149204 A1 WO 2015/183173 A1	05/04/2017 03/12/2015