(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 October 2017 (19.10.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/180061 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G06T 7/00* (2017.01)
G01N 33/533 (2006.01) *G06F 19/00* (2011.01)
G01N 33/68 (2006.01)

(21) International Application Number:
PCT/SG2017/050206

(22) International Filing Date:
11 April 2017 (11.04.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10201602855P 11 April 2016 (11.04.2016) SG

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

(72) Inventors: **ZINK, Daniele**; c/o Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, #09-01, The Nanos, Singapore 138669 (SG). **AKBAR HUSSAIN, Nur Faezah Begum**; c/o Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, #09-01, The Nanos, Singapore 138669 (SG). **LOO, Lit Hsin**; c/o Bioinformatics Institute, 30 Biopolis Street, #07-01 Matrix, Singapore 138671 (SG). **LAM, Ah Wah**; c/o Bioinformatics Institute, 30 Biopolis Street, #07-01 Matrix, Singapore 138671 (SG).

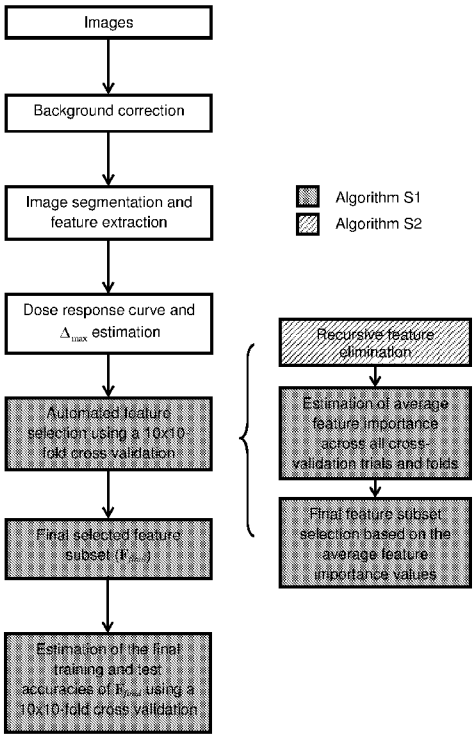
(74) Agent: **SPRUSON & FERGUSON (ASIA) PTE LTD**; P.O. Box 1531, Robinson Road Post Office, Singapore 903031 (SG).

(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, 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,

[Continued on next page]

(54) Title: HIGH THROUGHPUT METHOD FOR ACCURATE PREDICTION OF COMPOUND-INDUCED LIVER INJURY

Figure 2



(57) Abstract: A method and system for predicting liver injury in vivo due to hepatocyte damage by a test compound are provided. The method includes acquiring images of fluorescently stained cells obtained from a cell culture in which the cells have been treated with a dose-range of at least the test compound and its vehicle. The cells may be hepatic cells including primary or immortalized hepatocytes, hepatoma cells or induced pluripotent stem cell-derived hepatocyte-like cells. The acquired images are segmented. The method further includes extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of RELA (NF-KB p65), and (c) features of actin filaments at different subcellular regions and d) features of cellular organelles and their substructures in the segmented images. Finally, the method includes normalizing results from the treated samples to vehicle controls and predicting the probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted and selected phenotypic features using machine learning methods.

WO 2017/180061 A1

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).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

HIGH THROUGHPUT METHOD FOR ACCURATE PREDICTION OF COMPOUND-INDUCED LIVER INJURY

PRIORITY CLAIM

[0001] The present application claims priority from Singapore Patent Application No. 10201602855P filed 11 April 2016.

TECHNICAL FIELD

[0002] The present invention generally relates to compound-induced hepatotoxicity, and more particularly relates to a high-throughput method and system for accurate prediction of compound-induced liver injury based on hepatocyte damage *in vivo*.

BACKGROUND OF THE DISCLOSURE

[0003] The liver is the main human target organ for compound-induced toxicity. Compounds that damage the liver include drugs, herbal and fungal compounds, food and cosmetics ingredients, dietary supplements, industrial chemicals and environmental toxicants. Thus, compound-induced hepatotoxicity must be addressed by the chemical, food and nutrition, cosmetics and consumer care and pharmaceutical industries.

[0004] Drug-induced liver injury is a leading cause for acute liver failure requiring organ transplantation. Regulatory animal toxicity studies with combined rodent and non-rodent species identify only about fifty per cent of compounds that are hepatotoxic in humans. The low sensitivity of current pre-clinical hepatotoxicity models is one of the major reasons for attrition during late stages of development and is also the reason why drugs continue to be pulled from the market because of previously undetected hepatotoxicity.

[0005] Improved platforms for the prediction of hepatotoxicity are also required by the cosmetics industry, where there is a particular need for animal-free predictive platforms. Animal models cannot be used anymore for the testing of cosmetics developed or marketed in Europe, Norway, Israel and India due to animal testing bans. New Zealand, Canada, Taiwan, Brazil, Korea and also probably Argentina, USA and China are expected to follow.

[0006] Also, the food industry is now under substantial pressure from animal activist groups. Under this pressure, large food companies like Coca Cola have stopped using animal testing, the last food company having given into this pressure being Kikkoman. Accordingly, animal-free platforms for the accurate prediction of hepatotoxicity and other toxicities in humans are urgently required. Accepted *in vitro* or *in silico* models do not exist with respect to liver, kidney and other organs.

[0007] Thus, what is needed is methods and systems for high throughput *in vitro* modelling of compound-induced liver injury which provides accurate prediction of such compound-induced liver injury. Methods and systems based on high-throughput screening that are highly scalable and suitable for testing of large compound libraries at an industrial scale is also needed. Furthermore, other desirable features and characteristics will become apparent from the subsequent detailed description and the appended claims, taken in conjunction with the accompanying drawings and this background of the disclosure.

SUMMARY

[0008] In one aspect, there is provided a method for predicting liver injury induced by a test compound comprising: acquiring images of fluorescently stained cells obtained from a cell culture in which the cells have been treated with a dose-

range of at least the test compound and its vehicle; segmenting the acquired images; extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-□B p65), (c) features of actin filaments at different subcellular regions and d) features of cellular organelles and their substructures in the segmented images; normalizing results from the treated samples to vehicle controls; and predicting a probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted and selected phenotypic features using machine learning methods.

[0009] In an embodiment, extracting and selecting the one or more phenotypic features from the segmented images may comprise extracting and selecting the coefficient of variation of the intensity of a DNA marker at chromosomal regions in the segmented images.

[0010] In an embodiment, extracting and selecting the one or more phenotypic features from the segmented images may comprise extracting and selecting an average of a sum average of the intensity of a DNA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images.

[0011] In an embodiment, extracting and selecting the one or more phenotypic features from the segmented images may comprise extracting and selecting an average of the intensity of a RELA marker at chromosomal regions in the segmented images.

[0012] In an embodiment, extracting and selecting the one or more phenotypic features from the segmented images may comprise extracting and selecting a fraction

of the intensity of objects based on a RELA marker at nuclear regions in the segmented images.

[0013] In an embodiment, the cells may be selected from the group comprising hepatic cells, induced pluripotent stem cell (iPS)-derived hepatocyte-like cells, primary hepatocytes, an (immortalized) hepatocyte cell line, and hepatoma cells.

[0014] In an embodiment, the hepatic cells may be derived from hepatic progenitor-like cells.

[0015] In an embodiment, the cells are human cells. In another embodiment, the cells are animal cells.

[0016] In an embodiment, the hepatic cells may be HepaRG^{TM} cells.

[0017] In an embodiment, the cells of the cell culture may be cultured in a 2D culture.

[0018] In an embodiment, the may further comprising the step of fluorescently labelling the cells for detection of one or more of DNA, RELA (NF-κB p65), F-actin, gammaH2AX and the whole cells.

[0019] In an embodiment, treating the cells may comprise treating with the at least one test compound and the vehicle for a predetermined time duration of at least twelve hours or overnight.

[0020] In an embodiment, treating the cells may comprise treating with the at least one test compound and the vehicle for a predetermined time duration of sixteen hours or greater.

[0021] In an embodiment, predicting hepatotoxicity may comprise: treating the cells with a dose-range of at least one test compound and the vehicle; determining maximal response values (Δ_{max}) with respect to the phenotypic features; and using

machine learning methods to predict hepatotoxicity based on the compound-induced feature changes.

[0022] In an embodiment, using machine learning methods to predict hepatotoxicity may comprise using machine learning methods that use a median of the maximal response values (Δ_{max}) of the selected and normalized phenotypic feature changes of the cells for compound classification.

[0023] In an embodiment, the maximal response values may comprise response values at a concentration of 1 mM, 2 mM or 5 mM of the test compound.

[0024] In an embodiment, predicting liver toxicity may comprises: normalizing the phenotypic features from different test compounds to a same range when more than one test compound is tested; and predicting liver toxicity by using a machine learning classification algorithm that has been trained with a training set of compounds that is statistically independent from a test set of compounds.

[0025] In an embodiment, the machine learning classification algorithm may be a random-forest algorithm or a support vector machine.

[0026] In another aspect, there is provided a system for predicting liver injury induced by at least one test compound comprising: an apparatus for treating a cell culture with the at least one test compound and its vehicle and staining the cell culture for fluorescent detection; an imaging device arranged with respect to the apparatus to acquire images of the fluorescently stained cells for detecting the cells and their stained components; and a computing device coupled to the imaging device for receiving and segmenting the acquired images, the computing device comprising a feature extraction module for extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features

consisting of (a) features of DNA, (b) features of a RELA (NF-□B p65), and (c) features of actin filaments at different subcellular regions and d) features of cellular organelles and their substructures in the segmented images, the computing device further comprising a prediction module for normalizing results from the treated samples to vehicle controls and predicting a probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted and selected phenotypic features using machine learning methods.

[0027] In an embodiment, the feature extraction module may extract and select a phenotypic feature from the segmented images measuring a coefficient of variation of the intensity of a DNA marker at chromosomal regions in the segmented images.

[0028] In an embodiment, the feature extraction module may extract and select a phenotypic feature from the segmented images measuring an average of a sum average of the intensity of a DNA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images.

[0029] In an embodiment, the feature extraction module may extract and select a phenotypic feature from the segmented images measuring an average of the intensity of a RELA marker at chromosomal regions in the segmented images.

[0030] In an embodiment, the feature extraction module may extract and select a phenotypic feature from the segmented images measuring a fraction of the intensity of objects based on a RELA marker at nuclear regions in the segmented images.

[0031] In an embodiment, the prediction module of the computing device may predict hepatotoxicity by using machine learning methods based on maximum response values (Δ_{max}) with respect to the phenotypic feature changes of the cells treated with the at least one test compound and the vehicle.

[0032] In an embodiment, the prediction module of the computing device may predict hepatotoxicity by using machine learngn methods that uses a median of the maximum response values (Δ_{max}) of the selected and normalized phenotypic feature changes.

[0033] In an embodiment, the maximum response values may comprise response values at a compound concentration of 1mM, 2 mM or 5 Mm of the test compound.

[0034] In an embodiment, the prediction module of the computing device may predict hepatotoxicity based on changes of phenotypic features normalized to a same range when more than one test compound is used and by using a machine learning classification algorithm trained with a statistically independent training set of compounds.

[0035] In an embodiment, the machine learning classification algorithm may be a random forest algorithm or a support vector machine.

[0036] According to at least one embodiment of the present invention, a method for predicting liver injury *in vivo* due to hepatocyte damage by a test compound is provided. The method includes acquiring images of fluorescently stained cells obtained from a cell culture in which the cells have been treated with a dose-range of at least the test compound and its vehicle and segmenting the acquired images. The method further includes extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-κB p65), (c) features of actin filaments at different subcellular regions, and (d) features of cellular organelles and their substructures in the segmented images.

Finally, the method includes normalizing results from the treated samples to vehicle controls and predicting a probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted phenotypic features using a machine learning methods.

[0037] According to another embodiment of the present invention, a system for predicting liver injury *in vivo* due to hepatocyte damage by at least one test compound is provided. The system includes an apparatus for treating a cell culture with the at least one test compound and its vehicle and staining the cells for fluorescent detection, an imaging device and a computing device. The imaging device is arranged with respect to the apparatus to acquire images of the fluorescently stained cells for detecting the cells and their stained components. The computing device is coupled to the imaging device for receiving and segmenting the acquired images. The computing device includes a feature extraction module for extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-κB p65), (c) features of actin filaments at different subcellular regions, and (d) features of cellular organelles and their substructures in the segmented images. The computing device also includes a prediction module for normalizing results from the treated samples to vehicle controls and predicting a probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted phenotypic features using machine learning methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0038] The accompanying figures, where like reference numerals refer to identical or functionally similar elements throughout the separate views and which together with the detailed description below are incorporated in and form part of the specification, serve to illustrate various embodiments and to explain various principles and advantages in accordance with a present embodiment.

[0039] FIG. 1 depicts a block diagram of a system for predicting liver injury in accordance with a present embodiment, the liver injury induced by at least one test compound.

[0040] FIG. 2 depicts a flowchart of an imaging and data analysis method used by the computing device of the system of FIG. 1 for high throughput *in vitro* modelling for accurate prediction of compound-induced liver injury in accordance with the present embodiment.

[0041] FIG. 3, comprising FIGs. 3A to 3F, depicts exemplary immunofluorescence image data obtained with different compounds by nuclear and NF-$_K$B staining in accordance with the present embodiment.

[0042] FIG. 4 depicts a table of reference compound information used for validation of the high-throughput method with HepaRG cells in accordance with the present embodiment for accurate prediction of compound-induced injury.

[0043] FIG. 5 depicts a table of phenotypic feature sets and their prediction performances with HepaRG cells in accordance with the present embodiment.

[0044] FIG. 6 depicts a table of reference compound information used for pre-validation of the high-throughput method with induced pluripotent stem cell (iPS)-derived hepatocyte-like cells in accordance with the present embodiment for accurate prediction of compound-induced injury.

[0045] FIG. 7 shows image data illustrating phenotypic changes of iPS-derived hepatocyte-like cells induced by compounds that injure hepatocytes *in vivo.*

[0046] Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been depicted to scale.

DETAILED DESCRIPTION

[0047] The following detailed description is merely exemplary in nature and is not intended to limit the invention or the application and uses of the invention. Furthermore, there is no intention to be bound by any theory presented in the preceding background of the invention or the following detailed description. In the following description, the liver is the liver of an animal, such as non-human mammal and human, including, without limitation, animals classed as bovine, porcine, equine, canine, lupine, feline, murine, ovine, avian, piscine, caprine, corvine, acrine, or delphine. In one example, the liver is a human liver. It is the intent of the present embodiment to present a high-throughput system and method for the accurate prediction of hepatocyte toxicity *in vivo* that combines automated imaging of hepatocytes with automated computational analysis of the image data by phenotypic profiling results classification utilizing machine learning.

[0048] Unlike conventional systems and methods which determine the toxicity of compounds for cultivated hepatocytes, systems and methods in accordance with the present embodiment examine changes induced in *in vitro* cultivated cells and use those observed changes to predict hepatocyte toxicity *in vivo.*

[0049] Optimal results are obtained with HepaRG™ cells, which is one of the most promising hepatocyte models. The HepaRG™ cell line is a human bipotent progenitor cell line capable to differentiate toward two different cell phenotypes (i.e.,

biliary-like and hepatocyte-like cells). This cell line has been established from a liver tumor associated with chronic hepatitis C. HepaRG cells exhibit many characteristics of primary human hepatocytes, including morphology and expression of key metabolic enzymes, nuclear receptors, and drug transporters. Unlike HepG2 and Fa2N-4 cells, HepaRG cells have high P450 activity and complete expression of all nuclear receptors. iPS-derived human hepatocyte-like cells, primary human hepatocytes and other human hepatocyte cell lines may also be suitable.

[0050] Referring to FIG. 1, a block diagram 100 depicts a system for the prediction of liver injury in accordance with a present embodiment. The system 100 includes a cell culture and staining apparatus 110. Within the cell culture and staining apparatus 110, a cell culture is treated with the test compounds and a vehicle for compound exposure for a period of sixteen or more hours. After compound treatment, the treated cells are stained for fluorescent detection and various cellular components (DNA, RELA (NF- KB p65) and F-actin) and the whole cells are thereby fluorescently labeled. Thus, in accordance with the present embodiment, human hepatocytes are cultivated as two-dimensional (2D) monolayers in collagen-coated standard multi-well plates of the cell culture and staining apparatus 110. While currently HepaRG cells are used, primary human hepatocytes, iPS-derived hepatocyte-like cells and other types of hepatocytes or hepatocyte-like cells may be suitable in accordance with the present embodiment.

[0051] An imaging device 120 is optically arranged with respect to the cell culture and staining apparatus 110 to acquire images of the fluorescently stained cells for detecting the cells and their stained components. The imaging device 120 is a high content analyzer which automatically images the fluorescently stained cells to obtain immunofluorescent detection of RELA (NF-kB p65), F-actin, gammaH2AX and

4',6-diamidino-2-phenylindole (DAPI, DNA detection)) and whole cell stain (WCS) detection after the compound treatment.

[0052] The imaging device 120 provides the acquired images to a computing device 130 which is coupled to the imaging device 120 to receive and segment the acquired images. The computing device 130 analyzes phenotypic features by phenotypic profiling and uses machine learning methods for results classification and predictive performance analysis.

[0053] The computing device 130 includes a feature extraction module 132 which extracts one or more phenotypic features from the segmented images. In accordance with the present embodiment, a phenotypic feature is extracted by the feature extraction module 132 which is selected from the group of phenotypic features including (a) intensity or texture features of DNA marker at chromosomal or nuclear regions in the segmented images, and (b) intensity or ratiometric features of a RELA marker at chromosomal or nuclear regions in the segmented images. The computing device 130 also includes a prediction module 134 coupled to the feature extraction module 132 which uses the extracted phenotypic features to predict liver injury. More particularly, in accordance with the present embodiment, the prediction module 134 predicts hepatocyte toxicity *in vivo* using a supervised classifier developed using the phenotypic features. The system of the present embodiment advantageously provides high predictivity with a test sensitivity of 61.4%, a test specificity of 86.1% and a test balanced accuracy of 73.8%, a result significantly improved over conventional methods and systems.

[0054] Referring to FIG. 2, a flowchart 200 depicts an imaging and data analysis method used by the computing device 130 for high throughput prediction of compound-induced hepatocyte injury *in vivo* in accordance with the present

embodiment. The computing device 130 receives the images 202 and performs background correction 204 to improve the clarity of the image in preparation for image segmentation and feature extraction 206 by the feature extraction module 132. The prediction module performs a dose response curve fitting and Δ_{max} estimation process 208 before performing an automated feature selection 210 on the processed image.

[0055] The automated feature selection 210 in accordance with the present embodiment includes two novel methods: a recursive feature elimination algorithm 212 and a feature selection algorithm. The feature selection algorithm is performed after the recursive feature elimination algorithm 212 and includes a feature importance estimation step 214 and a final feature subset selection step 216. The feature importance estimation step 214 includes estimation of average feature importance across all 10x10-fold cross validation trials and folds. And the final feature subset selection step 216 includes final feature subset selection based on average feature importance values obtained at step 214.

[0056] The result is a final selected feature subset $\mathbf{F}_{final}$ 218. For validation, an estimation of the final training and test accuracies of $\mathbf{F}_{final}$ is performed using 10x10-fold cross validation 220.

[0057] Referring to FIG. 3, comprising FIGs. 3A to 3F, images 300, 310, 320, 330, 340, 350 depict exemplary immunofluorescence image data obtained with different compounds after fluorescent staining of RELA (red) and nuclear DNA (blue) in accordance with the present embodiment. Referring to FIG. 3A, the image 300 depicts fluorescently stained HepaRG cells obtained from a cell culture in which the cells have been treated with 100 µg/mL puromycin (positive control). Referring to FIGs. 3B and 3C (images 310, 320) the cells have been treated with1000 µg/mL

acarbose 1000 µg/mL tetracycline, respectively. Both compounds are toxic for hepatocytes *in vivo*. .

[0058] Referring to FIG. 3D, the image 330 depicts fluorescently stained cells of the vehicle control. And referring to FIGs. 3E and 3F, (images 340 and 350) the cells have been treated with100 µg/mL ethylene glycol and 1000 µg/mL lithium chloride, respectively.

[0059] The images 300, 310, 320, 330, 340, 350 depict HepaRG cells imaged by high-throughput screening with a high content analyzer (imaging device 120). . Hepatotoxicants in the images 300, 310, 320 induced phenotypic changes and an altered distribution of RELA (NF- κ B p65) 302 in comparison to the vehicle control image 330. Such phenotypic changes were not observed when cells were treated with non-hepatotoxic compounds as shown in the images 340, 350. The hepatotoxicant acarbose (the image 310) is not toxic for the human renal proximal tubular cells. Thus, in previous studies on renal models, changes of human primary renal proximal tubular cells (HPTC) or HPTC-like cells induced by the hepatotoxicant acarbose were not observed; however, the hepatotoxicant acarbose induced obvious changes in HepaRG cells as shown in the image 310. Ethylene glycol (nephrotoxic and neurotoxic) and lithium chloride (nephrotoxic) are both toxic in humans for other organs than the liver. Yet, these compounds did not induce obvious phenotypic changes in HepaRG cells as shown in the images 340, 350. These and other results suggest that the responses of the cells are highly organ-specific. In conventional hepatotoxicity testing, almost all non-hepatotoxic compounds with other toxicities in humans gave positive results with HepG2-based models. Therefore, in contrast to the HepaRG-based model in accordance with the present embodiment, the conventional HepG2-based models are not liver-specific, but detect cytotoxicity in general.

[0060] Referring to FIG. 4 a table of reference compound information used for validation of the high-throughput method in accordance with the present embodiment for accurate prediction of compound-induced injury is depicted. Those compounds that have been identified as toxic for hepatocytes in humans based on a literature survey using PubMed™ and Google® are indicated by "+".

[0061] Referring to FIG 5, the table shows the final selected feature subset F_{final} (218) consisting of 4 features. Δ_{max} (208) is indicated with respect to the estimated final training and test accuracies (220). The results were obtained with fluorescently labelled HepaRG cells and the set of 98 compounds (FIG. 4) used for pre-validation.

[0062] FIG. 6 depicts a table of reference compound information used for pre-validation of the high-throughput method with induced pluripotent stem cell (iPS)-derived hepatocyte-like cells in accordance with the present embodiment for accurate prediction of compound-induced injury.

[0063] FIG. 7 shows image data illustrating phenotypic changes of iPS-derived hepatocyte-like cells induced by compounds that injure hepatocytes *in vivo*. In particular, in Fig. 7, iPS-derived hepatocyte-like cells (iCell Hepatocytes) were treated with compounds at the indicated concentrations that are known to be toxic (compounds listed in top row – Azathioprine, Cadmium Chloride, Ranitidine HCL, Puromycin) or not toxic (compounds listed in bottom row – Aspartame, Caffeine, Nicotine, DMSO) for hepatocytes in vivo (classification based on literature on compound-induced clinical effects in humans) and respective positive (125 µM puromycin) and vehicle (0.5% dimethyl-sulfoxide, DMSO) controls . The cells were imaged by high content screening after immunofluorescent staining of RELA (red) and nuclear DNA (blue). These image data show examples phenotypic cellular

changes that were induced by compounds that damage hepatocytes in vivo (images at the top row).

[0064] iCell Hepatocytes (iCells) were purchased from Cellular Dynamics (Madison, Winconsin, USA). iCells were thawed with RPMI medium containing 1X B27, 20 ng/mL Oncostatin M, 0.1 μM Dexamethasone, 25 μg/mL Gentamicin and 1X iCell Hepatocyte Medium Supplement.

[0065] iCells were then seeded into collagen I coated 384-well plates at a seeding density of 30,000 cells per well. The cells were cultivated for 7 days before compound treatment. iCells were treated for 16 hours with test compounds at a dose-range of 7 concentrations which ranged from 2 μM to 1 mM.

[0066] The library used for pre-validation contained 24 compounds with well-documented clinical effects on hepatocytes in humans (subset of the 98 compounds listed in FIG. 4). The library contained diverse types of chemicals including drugs, industrial chemicals, agrochemicals, environmental toxicants and food additives.

[0067] After compound treatment, biomarkers were detected by immunostaining and the whole cells were stained with WCS. High-content screening was performed with a 20x objective using the ImageXpress Micro XLS system (Molecular Devices, Sunnyvale, CA, USA).

[0068] **EXAMPLE**

[0069] To demonstrate the present embodiment, HepG2 cells (used as controls) and HepaRG cells were purchased from the American Type Culture Collection (ATCC, Manassas, Virginia, USA) and ThermoFisher Scientific (Singapore), respectively. The HepG2 cells were cultured in T75 Corning Costar flasks with Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS)

and 1% penicillin-streptomycin. The HepG2 cells were seeded into 384-well black plates with transparent bottom (#781091; Greiner, Kremsmuenster, Austria,) at a density of 10,000 cells/cm^2. The HepG2 cells were cultured for three days to attain confluency before overnight compound treatment of sixteen hours. The HepaRG cells were directly seeded at a density of 250,000 cells/cm^2 into 384-well black plates with transparent bottom (#781091; Greiner) coated with 100 µg/mL collagen I. For the first twenty-four hours of cultivation a Williams E medium, supplemented with HepaRG Thaw, Plate & General Purpose medium was used. Subsequently, the HepaRG cells were cultivated for seven days in Williams E medium supplemented with HepaRG Tox medium and 1% GlutaMAX (all media and supplements from ThermoFisher Scientific) with medium replenishment on day 1 and day 4 of the culture. The HepaRG cells were cultured for 7 days before overnight compound treatment of 16 hours. The protocol of HepaRG culture for toxicity studies was obtained from ThermoFisher Scientific.

[0070] Compound treatment

[0071] The cells were treated for sixteen hours with 98 compounds as listed in the table of FIG. 4. Screening was performed by using a dose-range of the following seven concentrations for each compound: 2 µM, 15 µM, 63 µM, 125 µM, 250 µM, 500 µM and 1 mM. Puromycin (125 µM) and chlorpromazine (15 µM) were used as positive controls, while dexamethasone (250 µM and 1 mM) was used as a negative control. Untreated cells (no drug, no vehicle) and respective vehicle controls (dimethyl sulfoxide (DMSO) or water) were included in every plate. Four technical replicates were included for each treatment condition (four wells per data point). Z' values of each plate were calculated and values were all above 0.5, ranging from 0.62 to 0.89.

[0072] Immunostaining

[0073] Cells were fixed for six hours with 3.7% formaldehyde in phosphate-buffered saline (PBS) after sixteen hours of compound exposure. Cells were then blocked and permeabilized with PBS containing 5% bovine serum albumin (BSA) and 0.2% Triton X-100 for one hour at room temperature. Subsequently, samples were incubated for one hour at room temperature with either a rabbit polyclonal anti-RELA antibody (#ab16502; Abcam, Cambridge, UK) at 1 µg/mL or a mouse monoclonal anti-$_{gamma}$H2AX (phosphor S139) antibody (#ab26350; Abcam) at 2 µg/mL. Samples were then incubated with either a goat anti-rabbit secondary antibody conjugated with Alexa-488 (#A11008; Life Technologies, Singapore) or a goat anti-mouse secondary antibody conjugated with Alexa-647 (#A21235, Life Technologies) at 5 µg/mL. F-actin was detected on every plate with rhodamine phalloidin (#R415; 1:100; Invitrogen, Singapore). Cell nuclei and whole cells were stained on every plate with 4', 6-diamidino-2phenylindole dihydrochloride (DAPI; #268298; Merck, Darmstadt, Germany) at 4 ng/mL, and whole cell stain red (WCS; #8403402; 1:100; Cellomics, Rockford, United States of America).

[0074] Image acquisition

[0075] Automated imaging was performed with a 20x objective with the ImageXpress^{MICRO} system (Molecular Devices, Sunnyvale, CA, USA). Nine sites per well were imaged with four channels and the images were saved in a 16-bit TIFF format.

[0076] Image segmentation and feature extraction

[0077] To reduce non-uniform background illuminations, the images were corrected using the "rolling ball" algorithm implemented in ImageJ (NIH, v1.48). Cell

segmentations and feature measurements were performed using a cellXpress™ proprietary software platform.

[0078] Haralick's texture features

[0079] A grey-level co-occurrence matrix (GLCM) is a matrix that describes the distribution of co-occurring grey-level values at a given offset (Δx, Δy) in an $N_x \times N_y$ image, $\mathbf{I}(x, y)$, with N_g grey levels. In these notations, x and y are the row and column indices, respectively. The GLCM matrix is defined by Equation 1.

$$\boldsymbol{GLCM}_{\Delta x,\Delta y}(i,j) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \begin{cases} 1, & if\ I(x,y)\ and\ I(x+\Delta x, y+\Delta y = j \\ 0, & otherwise \end{cases} \quad (1)$$

where i and j are the grey-level or intensity values of the image. The normalized GLCM matrix is

$$p(i,j,\Delta x,\Delta y) = \frac{\boldsymbol{GLCM}_{\Delta x,\Delta y}(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \boldsymbol{GLCM}_{\Delta x,\Delta y}(i,j)} \quad (2)$$

The sum probability matrix is calculated to be

$$p_{x+y}(k,\Delta x,\Delta y) = \sum_{x=1}^{N_g} \sum_{y=1}^{N_g} p(i,j,\Delta x,\Delta y) \quad (3)$$

where $k = 2,3,\ldots,2N_g$. The sum average of GLCM are shown in Equations (4).

$$f_{SA}(\Delta x,\Delta y) = \sum_{k=2}^{2N_g} k\, p_{x+y}(k,\Delta x,\Delta y) \quad (4)$$

[0080] In the validation study, the images were the bounding boxes around the segmented cells with all the background pixels set to zero. The images were then quantized into $N_g = 256$ grey levels, and all the Haralick's features for 0 degree $(\Delta x = 0, \Delta y = 1)$, 45 degree $(\Delta x = 1, \Delta y = 1)$, 90 degree $(\Delta x = 1, \Delta y = 0)$, and 135 degree $(\Delta x = 1, \Delta y = -1)$ offsets were computed. For each feature, the mean of all the offset values was used and the cellXpress™ software platform was used to extract all the texture features.

[0081] Concentration response curve and Δ_{max} estimations

[0082] After feature extraction, the values of a feature were divided at all the tested compound concentrations by the values of the feature under the corresponding vehicle control conditions. Then, the ratios were log2-transformed ($_{\Delta}$). All further data analyses, including building concentration response curves and toxicity classifiers, were performed using customized scripts under the R statistical environment (the R foundation, v3.0.2) and the Windows 7 operating system (Microsoft, USA).

[0083] For each feature, the concentration response curve was estimated using a standard log-logistic model as shown in Equation (5).

$$\Delta\big(x,(b,c,d,e)\big) = \frac{d-c}{1+\exp\{b(\log(x)-\log(e))\}} \quad (5)$$

where x is the xenobiotics compound concentration, e is the response half-way between the lower limit c and upper limit d, and b is the relative slope around e. The "drc" library (v 2.3-96) under the R environment was used to fit the values of b, c, d, and e. After that, the maximum response values (Δ_{max}) were determined using the estimated response curves. In theory, Δ_{max} should be equal to the upper limit d. However, in practice, the responses of some compounds may not plateau even at the highest tested dosages. Therefore the estimated d value may not be accurate. Instead, Δ_{max} was fixed to be the response value at either 1, 2, or 5 mM, which was around the highest tested concentrations for most of the compounds. Finally, the median values of Δ_{max} across all the replicates were computed.

[0084] Feature normalization

[0085] Before data classification, each feature vector f_i was normalized to the same range [-1, 1] as shown in Equation (6).

$$\mathbf{f}_i \leftarrow 2\frac{(\mathbf{f}_i - f_{min})}{f_{max} - f_{min}} - 1 \quad (6)$$

where f_{min} and f_{max} are the minimum and maximum values of the feature. To ensure the training and test datasets were independent from each other, these two normalization coefficients were estimated only using the training data, but were applied to both the training and the test datasets.

[0086] Random forest classification

[0087] A random-forest algorithm was used to predict xenobiotic-induced nephrotoxicity, as this algorithm has previously been shown to outperform other commonly-used classifiers, including support vector machines, *k*-nearest neighbors and naïve Bayes. In addition, the "randomForest" library (v4.6-10) under the R environment was used.

[0088] Classification performance estimation

[0089] A stratified 10-fold cross validation procedure was used to estimate the toxicity prediction performance of the phenotypic features. The performance measurements used are shown in Equations (7), (8) and (9).

$$\text{Sensitivity} = \frac{TP}{TP+FN} \text{ x } 100\% \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \text{ x } 100\% \quad (8)$$

$$\text{Balanced Accuracy } (acc) = \frac{\text{Sensitivity+Specificity}}{2} \quad (9)$$

where *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives and *FN* is the number of false negatives. FIG. 5 depicts a table summarizing the overall prediction performance for single- and multi-feature classifiers for all three datasets in accordance with the present embodiment.

[0090] Referring to FIG. 5, a table lists the phenotypic feature sets and their prediction performance with HepaRG cells when tested against the ninety-eight compounds listed in FIG. 4 in accordance with present embodiment. The phenotypic feature sets include: a coefficient of variation of the intensity of a DNA marker at the chromosomal region, an average of the sum average of the DNA grey-level co-

occurrence matrix (GLCM) at the nuclear region, an average of the intensity of a RELA marker at the chromosomal region, and a fraction of the total intensity of the RELA objects at the nuclear region. Nuclear fraction of the RELA objects $= \frac{I(C \cap D)}{I(D)}$, where RELA objects are the subcellular regions that have high levels of RELA stains, $I(D)$ is the sum of the intensity values of all the pixels at the RELA-object regions, and $I(C \cap D)$ is the sum of the intensity values of all the pixels at the intersection of the cytoplasmic and RELA-object regions. Using a concentration of 1 mM to estimate the maximum response values (Δ_{max}) of the combination of these feature sets resulted in a test sensitivity of 47.9-61.4%, a test specificity of 86.1-92.1% and a test accuracy of 70.0-73.8%.

[0091] Thus it can be seen that the present embodiment provides an improved predictive model for drug-induced liver injury in humans utilizing an accurate and high-throughput screening (HTS)-based method and system for predicting compound-induced hepatocyte injury *in vivo*. Predicting hepatocyte toxicity *in vivo* in accordance with the present embodiment is based on HepaRG cells and not on HepG2 cells and high sensitivity and accuracy has been achieved (for the 98 compounds tested, test sensitivity is 61.4%, test specificity is 86.1%, and test balanced accuracy is 73.8%).

[0092] Unlike conventional models which determine the toxicity of compounds for cultivated hepatocytes, the system and methods in accordance with the present embodiment only look at changes induced in *in vitro* cultivated cells and use these changes to predict hepatocyte toxicity *in vivo*. The HTS model in accordance with the present embodiment makes binary predictions and gives a yes/no answer with respect to hepatocyte toxicity *in vivo*. No predictions however can be made with respect to the dose. This can be overcome by complementing the HTS model with

computational models or other *in vitro* models that predict the human dose response. Such models are still at the experimental stage, but in the future this may be an option. An advantage of the present embodiment is its high throughput. It can be efficiently applied at very early stages of compound screening and would allow the flagging of all compounds that are predicted to be toxic for hepatocytes *in vivo*. For most of these compounds, more detailed information on the dose response might not be required.

[0093] Further, the disclosure comprises embodiments according to the following clauses:

[0094] Clause 1. A method for predicting human liver injury induced by a test compound comprising: acquiring images of fluorescently stained cells obtained from a cell culture in which the cells have been treated with a dose-range of at least the test compound and its vehicle; segmenting the acquired images; extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-κB p65), and (c) features of actin filaments at different subcellular regions in the segmented images; normalizing results from the treated samples to vehicle controls; and predicting a probability of human liver injury by the test compound based on test compound-induced normalized phenotypic changes of the extracted phenotypic features using a supervised classifier.

[0095] Clause 2. The method in accordance with Clause 1, wherein extracting the one or more phenotypic features from the segmented images comprises extracting the phenotypic feature measuring a coefficient of variation of DNA intensity at chromosomal regions in the segmented images.

[0096] Clause 3. The method in accordance with Clause 1, wherein extracting the one or more phenotypic features from the segmented images comprises extracting two phenotypic features consisting of (a) a fraction of a total DNA object intensity at chromosomal regions in the segmented images and (b) an average of a sum of a variance of a RELA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images.

[0097] Clause 4. The method in accordance with Clause 1, wherein extracting the one or more phenotypic features from the segmented images comprises extracting four phenotypic features consisting of (a) a fraction of a total DNA object intensity at chromosomal regions in the segmented images, (b) an average of a sum of a variance of a RELA determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images, (c) a perimeter of one or more of the nuclear regions in the segmented images, and (d) ratios of total RELA intensity between the chromosomal regions and the nuclear regions in the segmented images.

[0098] Clause 5. The method in accordance with Clause 1, wherein the cells are selected from the group comprising differentiated hepatic cells, induced pluripotent stem cell (iPS)-derived human hepatocyte-like cells, primary human hepatocytes, and an (immortalized) human hepatocyte cell line.

[0099] Clause 6. The method in accordance with Clause 5, wherein the differentiated hepatic cells are derived from a human hepatic progenitor-like cell line.

[00100] Clause 7. The method in accordance with Clause 6, wherein the differentiated hepatic cells derived from a human hepatic progenitor-like cell line are HepaRG™ cells.

[00101] Clause 8. The method in accordance with any one of Clauses 1 to 7 wherein the cells of the cell culture are cultured in a 2D culture.

[00102] Clause 9. The method in accordance with any one of Clauses 1 to 8 further comprising the step of fluorescently labelling the cells cultured in the cell culture for detection of one or more of DNA, RELA (NF-κB p65), F-actin, gammaH2AX and a whole cell.

[00103] Clause 10. The method in accordance with any one of Clauses 1 to 9 wherein treating the cells comprises treating human hepatocytes with the at least one test compound and the vehicle for a predetermined time duration of at least twelve hours or overnight.

[00104] Clause 11. The method in accordance with Clause 10 wherein treating the cells comprises treating human hepatocytes with the at least one test compound and the vehicle for a predetermined time duration of sixteen hours or greater.

[00105] Clause 12. The method in accordance with any one of Clauses 1 to 11 wherein predicting human hepatotoxicity comprises: treating the cells with a dose-range of at least one test compound and the vehicle; determining maximal response values (Δ_{max}) with respect to the phenotypic features; and using the supervised classifier to predict human hepatotoxicity based on the compound-induced feature changes.

[00106] Clause 13. The method in accordance with Clause 12 wherein using the supervised classifier to predict human hepatotoxicity comprises using the supervised classifier that uses a median of the maximal response values (Δ_{max}) of the normalized phenotypic feature changes of the cells for compound classification.

[00107] Clause 14. The method in accordance with Clause 12, when the maximal response values comprise response values at a concentration of 1 mM, 2 mM or 5 mM of the test compound.

[00108] Clause 15. The method in accordance with Clause 1 wherein predicting liver toxicity in humans comprises: normalizing the phenotypic features from different test compounds to a same range when more than one test compound is tested; and predicting liver toxicity in humans by using a machine learning classification algorithm that has been trained with a training set of compounds that is statistically independent from a test set of compounds.

[00109] Clause 16. The Clause in accordance with Clause 15, wherein the machine learning classification algorithm is a random-forest algorithm.

[00110] Clause 17. The method in accordance with Clause 15, wherein the machine learning classification algorithm is a support vector machine.

[00111] Clause 18. A system for predicting human liver injury induced by at least one test compound comprising: an apparatus for treating a cell culture with the at least one test compound and its vehicle and staining the cell culture for fluorescent detection; an imaging device arranged with respect to the apparatus to acquire images of the fluorescently stained cells for detecting the cells and their stained components; and a computing device coupled to the imaging device for receiving and segmenting the acquired images, the computing device comprising a feature extraction module for extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-κB p65), and (c) features of actin filaments at different subcellular regions in the segmented images, the computing device further comprising a prediction module for normalizing results from the treated samples to vehicle controls and predicting a probability of human liver injury by the test compound

based on test compound-induced normalized phenotypic changes of the extracted phenotypic features using a supervised classifier.

[00112] Clause 19. The system in accordance with Clause 18 wherein the feature extraction module extracts a phenotypic feature from the segmented images measuring a coefficient of variation of DNA intensity at chromosomal regions in the segmented images.

[00113] Clause 20. The system in accordance with Clause 19 wherein the feature extraction module extracts two phenotypic features consisting of (a) a fraction of a total DNA object intensity at chromosomal regions in the segmented images and (b) an average of a sum of a variance of a RelA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images.

[00114] Clause 21. The system in accordance with Clause 19 wherein the feature extraction module extracts four phenotypic features consisting of (a) a fraction of a total DNA object intensity at chromosomal regions in the segmented images, (b) an average of a sum of a variance of a RelA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images, (c) a perimeter of one or more of the nuclear regions in the segmented images, and (d) ratios of total RelA marker intensity between the chromosomal regions and the nuclear regions in the segmented images.

[00115] Clause 22. The system in accordance with any of Clauses 19 to 21 wherein the prediction module of the computing device predicts human hepatotoxicity by using a supervised classifier based on maximum response values (Δ_{max}) with respect to the phenotypic feature changes of the cells treated with the at least one test compound and the vehicle.

[00116] Clause 23. The system in accordance with Clause 22 wherein the prediction module of the computing device predicts human hepatotoxicity by using a supervised classifier that uses a median of the maximum response values (Δ_{max}) of the normalized phenotypic feature changes.

[00117] Clause 24. The system in accordance with Clause 22, wherein the maximum response values comprises response values at a compound concentration of 1mM, 2 mM or 5 Mm of the test compound.

[00118] Clause 25. The system in accordance with any one of Clauses 20 to 24 wherein the prediction module of the computing device predicts human hepatotoxicity based on changes of phenotypic features normalized to a same range when more than one test compound is used and by using a machine learning classification algorithm trained with a statistically independent training set of compounds.

[00119] Clause 26. The system in accordance with Clause 25 wherein the machine learning classification algorithm is a random forest algorithm.

[00120] While exemplary embodiments have been presented in the foregoing detailed description of the invention, it should be appreciated that a vast number of variations exist. It should further be appreciated that the exemplary embodiments are only examples, and are not intended to limit the scope, applicability, operation, or configuration of the invention in any way. Rather, the foregoing detailed description will provide those skilled in the art with a convenient road map for implementing an exemplary embodiment of the invention, it being understood that various changes may be made in the function and arrangement of steps and method of operation described in the exemplary embodiment without departing from the scope of the invention as set forth in the appended claims.

CLAIMS

What is claimed is:

1. A method for predicting liver injury induced by a test compound comprising:

acquiring images of fluorescently stained cells obtained from a cell culture in which the cells have been treated with a dose-range of at least the test compound and its vehicle;

segmenting the acquired images;

extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-κB p65), (c) features of actin filaments at different subcellular regions and d) features of cellular organelles and their substructures in the segmented images;

normalizing results from the treated samples to vehicle controls; and

predicting a probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted and selected phenotypic features using machine learning methods.

2. The method in accordance with Claim 1, wherein extracting and selecting the one or more phenotypic features from the segmented images comprises extracting and selecting the coefficient of variation of the intensity of a DNA marker at chromosomal regions in the segmented images.

3. The method in accordance with Claim 1, wherein extracting and selecting the one or more phenotypic features from the segmented images comprises extracting and selecting an average of a sum average of the intensity of a DNA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images.

4. The method in accordance with Claim 1, wherein extracting and selecting the one or more phenotypic features from the segmented images comprises extracting and selecting an average of the intensity of a RELA marker at chromosomal regions in the segmented images.

5. The method in accordance with Claim 1, wherein extracting and selecting the one or more phenotypic features from the segmented images comprises extracting and selecting a fraction of the intensity of objects based on a RELA marker at nuclear regions in the segmented images.

6. The method in accordance with Claim 1, wherein the cells are selected from the group comprising hepatic cells, induced pluripotent stem cell (iPS)-derived hepatocyte-like cells, primary hepatocytes, an (immortalized) hepatocyte cell line, and hepatoma cells.

7. The method in accordance with Claim 6, wherein the hepatic cells are derived from hepatic progenitor-like cells.

8. The method in accordance with Claims 6 and 7, wherein the cells are human cells.

9. The method in accordance with Claims 6 and 7, wherein the cells are animal cells.

10. The method in accordance with Claim 7, wherein the hepatic cells are HepaRG™ cells.

11. The method in accordance with any one of Claims 1 to 10 wherein the cells of the cell culture are cultured in a 2D culture.

12. The method in accordance with any one of Claims 1 to 11 further comprising the step of fluorescently labelling the cells for detection of one or more of DNA, RELA (NF-κB p65), F-actin, gammaH2AX and the whole cells.

13. The method in accordance with any one of Claims 1 to 12 wherein treating the cells comprises treating with the at least one test compound and the vehicle for a predetermined time duration of at least twelve hours or overnight.

14. The method in accordance with Claim 13 wherein treating the cells comprises treating with the at least one test compound and the vehicle for a predetermined time duration of sixteen hours or greater.

15. The method in accordance with any one of Claims 1 to 14 wherein predicting hepatotoxicity comprises:

treating the cells with a dose-range of at least one test compound and the vehicle;

determining maximal response values (Δ_{max}) with respect to the phenotypic features; and

using machine learning methods to predict hepatotoxicity based on the compound-induced feature changes.

16. The method in accordance with Claim 15 wherein using machine learning methods to predict hepatotoxicity comprises using machine learning methods that use a median of the maximal response values (Δ_{max}) of the selected and normalized phenotypic feature changes of the cells for compound classification.

17. The method in accordance with Claim 15, when the maximal response values comprise response values at a concentration of 1 mM, 2 mM or 5 mM of the test compound.

18. The method in accordance with Claim 1 wherein predicting liver toxicity comprises:

normalizing the phenotypic features from different test compounds to a same range when more than one test compound is tested; and

predicting liver toxicity by using a machine learning classification algorithm that has been trained with a training set of compounds that is statistically independent from a test set of compounds.

19. The method in accordance with claim 18, wherein the machine learning classification algorithm is a random-forest algorithm.

20. The method in accordance with claim 18, wherein the machine learning classification algorithm is a support vector machine.

21. A system for predicting liver injury induced by at least one test compound comprising:

an apparatus for treating a cell culture with the at least one test compound and its vehicle and staining the cell culture for fluorescent detection;

an imaging device arranged with respect to the apparatus to acquire images of the fluorescently stained cells for detecting the cells and their stained components; and

a computing device coupled to the imaging device for receiving and segmenting the acquired images, the computing device comprising a feature extraction module for extracting and analyzing one or more phenotypic features from the segmented images, wherein the one or more phenotypic features are selected from the group of intensity, textural, morphological, or ratiometric features consisting of (a) features of DNA, (b) features of a RELA (NF-κB p65), and (c) features of actin filaments at different subcellular regions and d) features of cellular organelles and their substructures in the segmented images, the computing device further comprising a prediction module for normalizing results from the treated samples to vehicle controls and predicting a probability of liver injury by the test compound based on test compound-induced normalized changes of the extracted and selected phenotypic features using machine learning methods.

22. The system in accordance with Claim 21 wherein the feature extraction module extracts and selects a phenotypic feature from the segmented images measuring a coefficient of variation of the intensity of a DNA marker at chromosomal regions in the segmented images.

23. The system in accordance with Claim 21 wherein the feature extraction module extracts and selects a phenotypic feature from the segmented images measuring an average of a sum average of the intensity of a DNA marker determined from a grey-level co-occurrence matrix (GLCM) at nuclear regions in the segmented images.

24. The system in accordance with Claim 21 wherein the feature extraction module extracts and selects a phenotypic feature from the segmented images measuring an average of the intensity of a RELA marker at chromosomal regions in the segmented images.

25. The system in accordance with Claim 21 wherein the feature extraction module extracts and selects a phenotypic feature from the segmented images measuring a fraction of the intensity of objects based on a RELA marker at nuclear regions in the segmented images.

26. The system in accordance with any of Claims 22 to 25 wherein the prediction module of the computing device predicts hepatotoxicity by using machine learning methods based on maximum response values (Δ_{max}) with respect to the

phenotypic feature changes of the cells treated with the at least one test compound and the vehicle.

27. The system in accordance with Claim 26 wherein the prediction module of the computing device predicts hepatotoxicity by using machine learngn methods that uses a median of the maximum response values (Δ_{max}) of the selected and normalized phenotypic feature changes.

28. The system in accordance with Claim 26, wherein the maximum response values comprises response values at a compound concentration of 1mM, 2 mM or 5 Mm of the test compound.

29. The system in accordance with any one of Claims 23 to 28 wherein the prediction module of the computing device predicts hepatotoxicity based on changes of phenotypic features normalized to a same range when more than one test compound is used and by using a machine learning classification algorithm trained with a statistically independent training set of compounds.

30. The system in accordance with Claim 29 wherein the machine learning classification algorithm is a random forest algorithm.

31. The system in accordance with Claim 29 wherein the machine learning classification algorithm is a support vector machine.

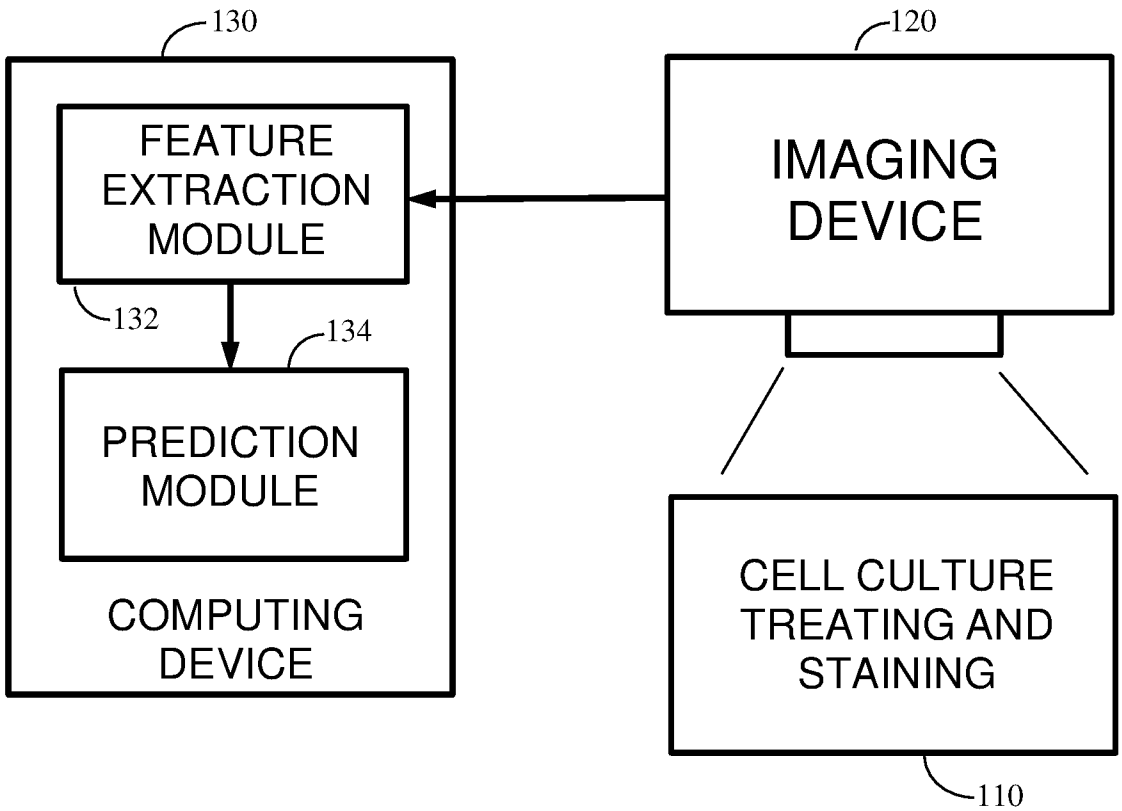
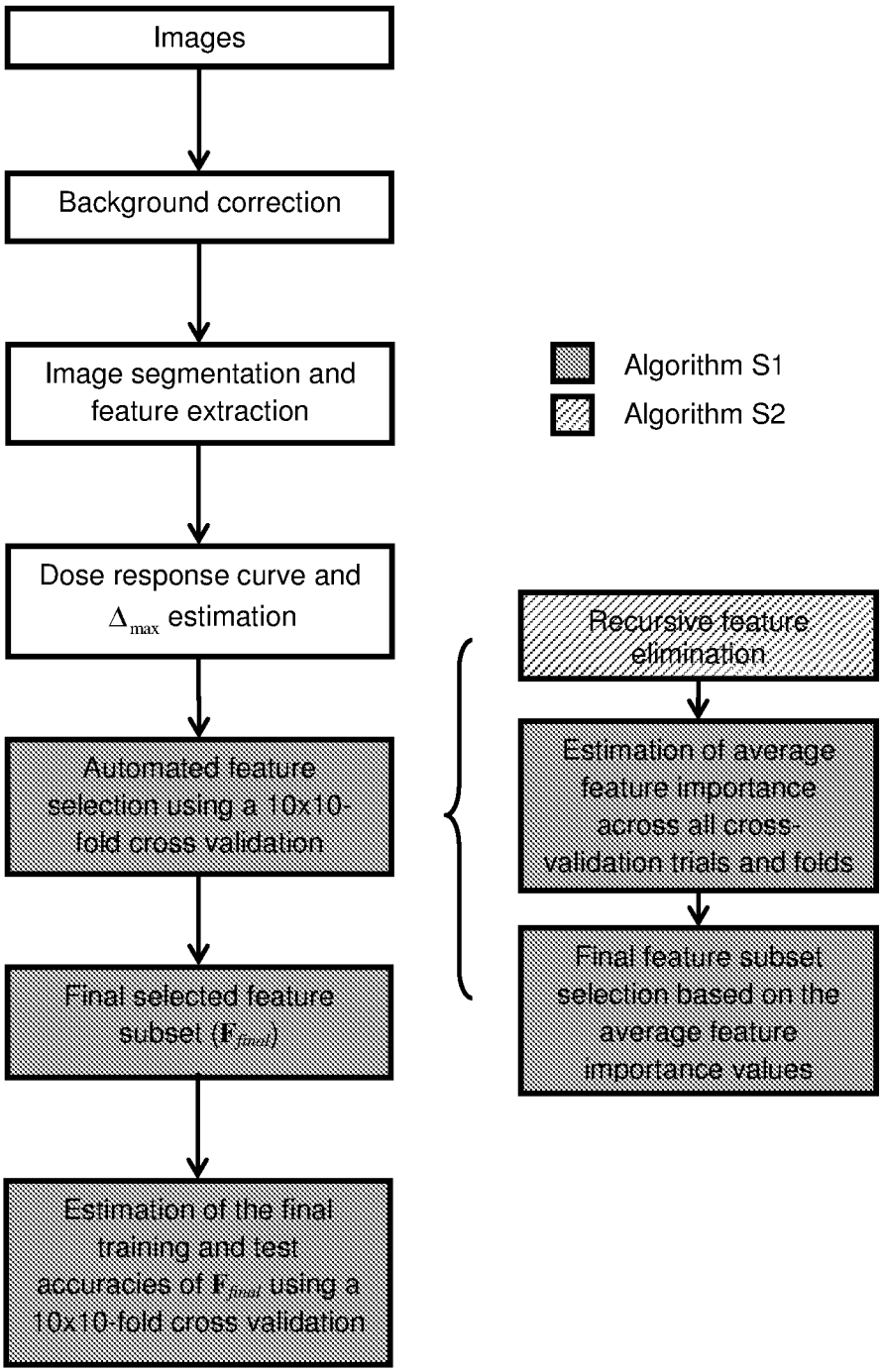
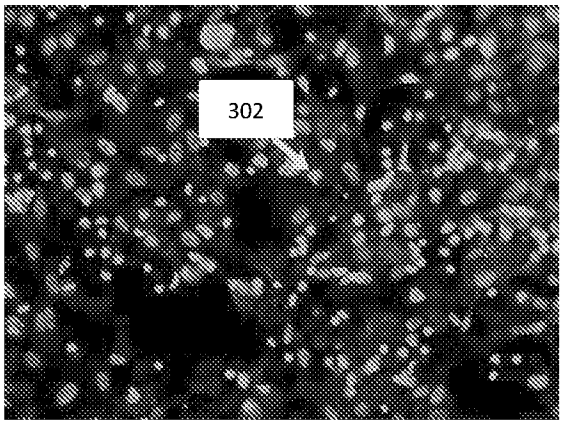

130
120
FEATURE EXTRACTION MODULE
IMAGING DEVICE
132
134
PREDICTION MODULE
COMPUTING DEVICE
CELL CULTURE TREATING AND STAINING
110
100

FIG. 1

Figure 2


Images
Background correction
Image segmentation and feature extraction
Dose response curve and Δ_{max} estimation
Automated feature selection using a 10x10-fold cross validation
Final selected feature subset ($\mathbf{F}_{final}$)
Estimation of the final training and test accuracies of $\mathbf{F}_{final}$ using a 10x10-fold cross validation
Algorithm S1
Algorithm S2
Recursive feature elimination
Estimation of average feature importance across all cross-validation trials and folds
Final feature subset selection based on the average feature importance values



300 FIG. 3A

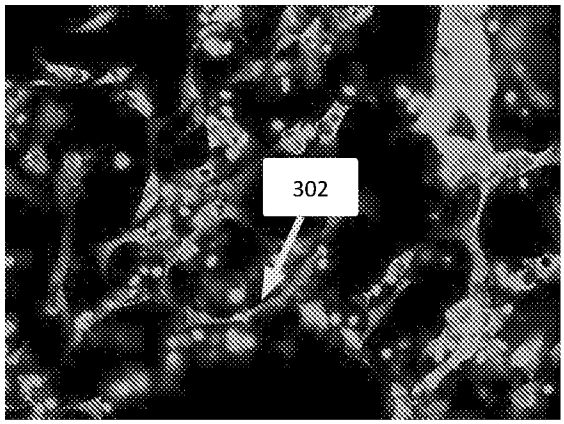
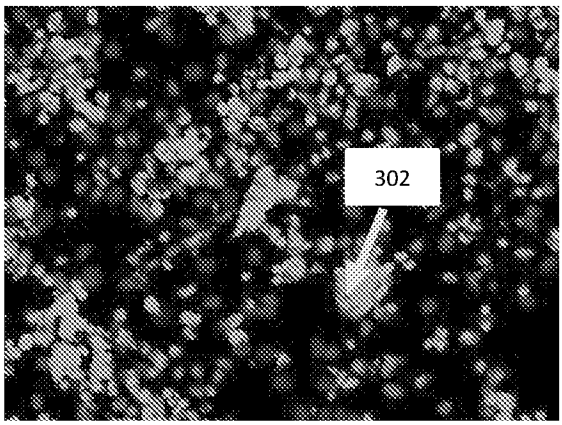


FIG. 3B 310



320 FIG. 3C

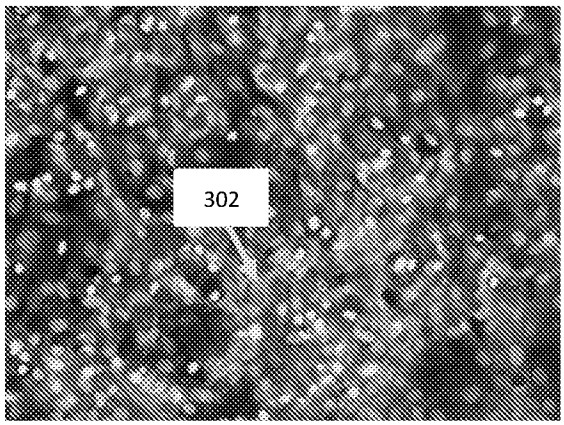
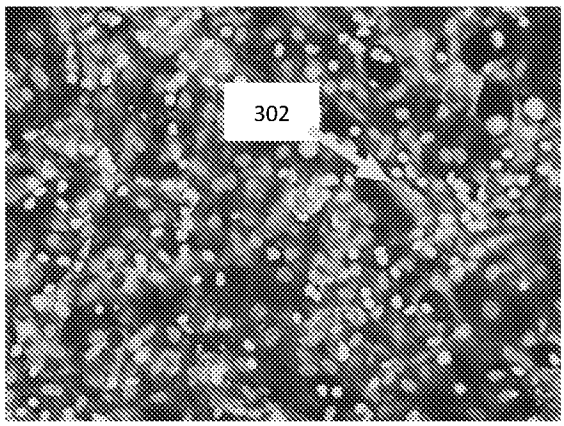


FIG. 3D 330



340 FIG. 3E

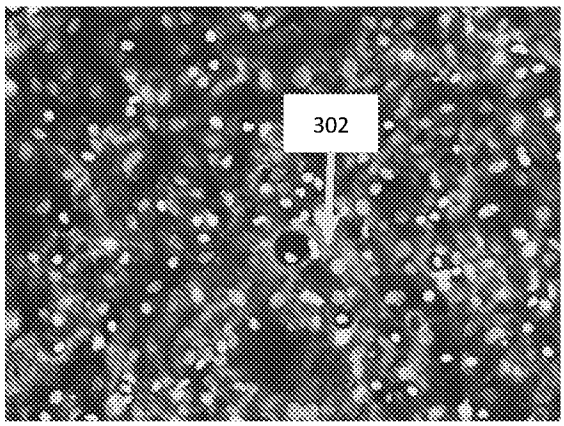


FIG. 3F 350

S/N	Chemicals	Hepatocyte Toxicity *In Vivo*
1	Acarbose	+
2	Acesulfame potassium	-
3	Acetamidophenol	-
4	Acetaminophen	-
5	Allantoin	-
6	Aristolochic acid	+
7	Arsenic(III) oxide	+
8	Artemisinin	-
9	Ascorbate	-
10	Aspartame	-
11	Azathioprine	+
12	Azithromycin	+
13	Betaine	-
14	Bicalutamide	+
15	Bismuth(III) oxide	+
16	Cadmium(II) chloride	+
17	Caffeine	-
18	Capecitabine	-
19	Cephaloridine	-
20	Chloral hydrate	-
21	Cinchophen	+
22	Ciprofloxacin	+
23	Cisplatin	-
24	Citrinin	-
25	Clofibrate	-
26	Colestipol	-
27	Copper(II) chloride	+
28	Cyclosporin A	+
29	Cyproterone acetate	+
30	D-alpha-Tocopherol	-
31	Dacarbazine	+
32	Diclofenac	+
33	Diltiazem HCl	+
34	DOPA	-
35	Dutasteride	-
36	Emtricitabine	-
37	Epigallocatechin gallate	+
38	Eplerenone	-
39	Ethylene glycol	-
40	Felbamate	+
41	Finasteride	-
42	5-Flurouracil	+

FIG. 4

43	Flutamide	+
44	Furosemide	+
45	Garcinia acid	-
46	Gentamycin sulfate	-
47	Glutathione	-
48	Glycine	-
49	Gold(I) chloride	+
50	Halothane	+
51	Ibuprofen	+
52	Ifosfamide	-
53	Isoniazid	+
54	Ketoconazole	+
55	Ketorolac	-
56	Leflunomide	+
57	Lincomycin	-
58	Lithium chloride	-
59	Melatonin	-
60	Menthol	-
61	Metformin HCl	-
62	Methotrexate	+
63	Methyl vanillate	-
64	Modafinil	-
65	N-acetyl-cysteine	-
66	Naloxone	-
67	Naltrexone HCl	-
68	Naproxen	+
69	Nevirapine	+
70	Nicotine	-
71	Penicillin	-
72	Phenacetin	-
73	Potassium dichromate	+
74	Probenecid	-
75	Promethazine	-
76	Ranitidine HCl	+
77	Retinoic acid	-
78	Ribavirin	-
79	Rifampicin	+
80	Simvastatin	-
81	Sodium chloride	-
82	Sodium valproate	+
83	Stavudine (d4T)	+
84	Sucrose	-
85	Sunitinib malate	+
86	Tacrolimus	-
87	Tetracycline	+
88	Theophylline	-

FIG. 4 (continued)

89	Tigecycline	-
90	Tolcapone	+
91	Tolvaptan	+
92	Torsemide	-
93	Triptolide	+
94	Trisodium citrate dihydrate	-
95	Troglitazone	+
96	Valacyclovir	-
97	Vancomycin	+
98	Vardenafil HCl	-

FIG. 4 (continued)

No.	Feature set	Δ_{max} concentration	Test Sensitivity	Test specificity	Test accuracy
1	Coefficient of variation of DNA intensity at the chromosomal region	1mM	61.4%	86.1%	73.8%
2	Average of the sum average of DNA GLCM at the nuclear region	1mM	56.0%	89.5%	72.7%
3	Average intensity of RELA at the chromosomal region	1mM	53.1%	89.6%	71.4%
4	Fraction of the total intensity of RELA objects at the nuclear region	1mM	47.9%	92.1%	70.0%

FIG. 5

S/N	Chemicals	Hepatocyte Toxicity *In Vivo*
1	Acesulfame Potassium	-
2	Arsenic (III) Oxide	+
3	Aspartame	-
4	Azathioprine	+
5	Cadmium (II) Chloride	+
6	Caffeine	-
7	Chloral Hydrate	-
8	Citrinin	-
9	Clofibrate	-
10	Cyclosporin A	+
11	Diclofenac	+
12	Felbamate	+
13	5-Flurouracil	+
14	Garcinia Acid	-
15	Glutathione	-
16	Glycine	-
17	Gold (I) Chloride	+
18	Isoniazid	+
19	Modafinil	-
20	Nicotine	-
21	Ranitidine HCl	+
22	Tetracycline	+
23	Torsemide	-
24	Vancomycin	+

FIG. 6

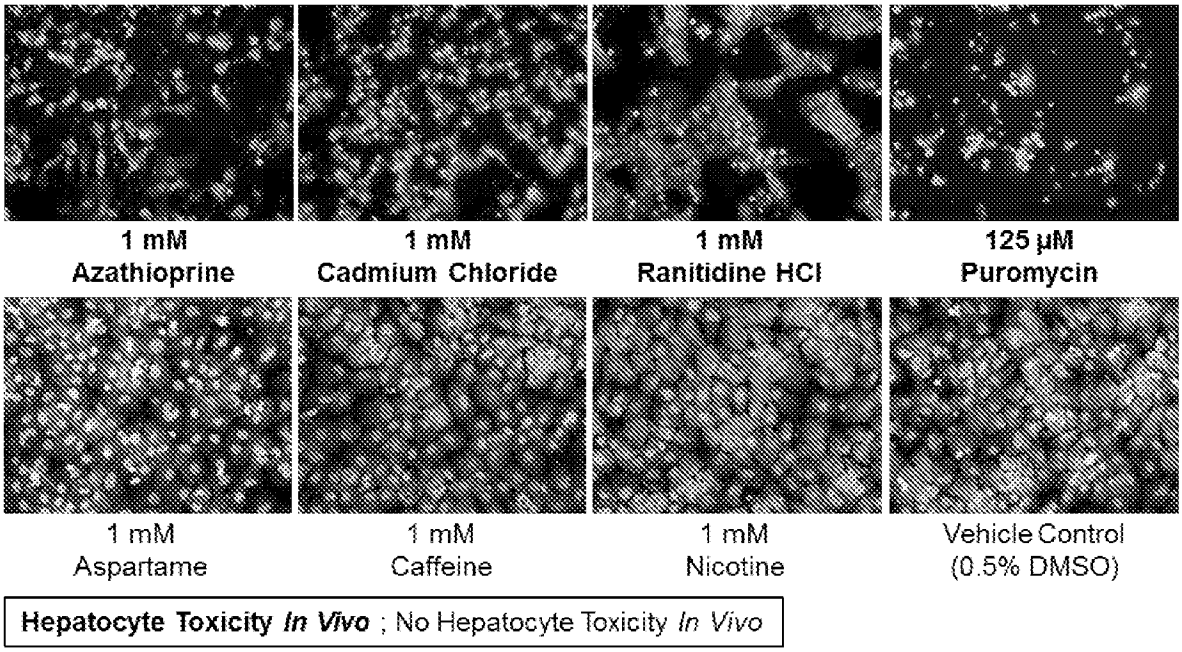

1 mM
Azathioprine
1 mM
Cadmium Chloride
1 mM
Ranitidine HCl
125 µM
Puromycin
1 mM
Aspartame
1 mM
Caffeine
1 mM
Nicotine
Vehicle Control
(0.5% DMSO)
Hepatocyte Toxicity *In Vivo* ; No Hepatocyte Toxicity *In Vivo*

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2017/050206

A. CLASSIFICATION OF SUBJECT MATTER

See Supplemental Box

According to International Patent Classification (IPC)

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q; G01N

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)

FAMPAT/CAPLUS/BIOSIS/EMBASE/MEDLINE: Predict, drug, liver, injury, DNA, RELA, actin, imaging and similar terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 2005/0014217 A1 (MATTHEAKIS L.C. ET AL.) 20 January 2005 Para. [0003], [0025]-[0179], [0197]; Fig. 1B & 3	1 and 6-21 2-5 and 22-31
X Y	WO 2009/002565 A1 (CELLUMEN, INC.) 31 December 2008 Page 3-Page 44; Fig. 1 & 3	1 and 6-21 2-5 and 22-31
X Y	PerkinElmer "Application Note: Cytotoxicity Studies on Live Primary Human Hepatocytes using Operetta". May 2011 [Retrieved on 2017-07-28 from https://www.perkinelmer.com/lab-solutions/resources/docs/APP_CytotoxicitystudiesonliveprimaryhepatocytesOperetta.pdf] Whole document	1 and 6-21 2-5 and 22-31

☒ 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	Date of mailing of the international search report
28/07/2017 *(day/month/year)*	10/08/2017 *(day/month/year)*

Name and mailing address of the ISA/SG

IPOS INTELLECTUAL PROPERTY OFFICE OF SINGAPORE

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

Email: pct@ipos.gov.sg

Authorized officer

Sung Ying Ying (Dr)

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2017/050206

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	SIRENKO O. ET AL., High-Content Assays for Hepatotoxicity Using Induced Pluripotent Stem Cell-Derived Cells. *Assay Drug Dev Technol*, 14 November 2013, Vol. 12, No. 1, pages 43-54 [Retrieved on 2017-07-28] <DOI: 10.1089/ADT.2013.520> Whole document	1 and 6-21 2-5 and 22-31
X Y	MENNECOZZI M. ET AL., Hepatotoxicity Screening Taking a Mode-Of-Action Approach Using HepaRG Cells and HCA. ALTEX-Proceedings, 2012, Vol. 1, No. 1, pages 193-204 [Retrieved on 2017-07-28] <DOI: NOT AVAILABLE> Whole document	1 and 6-21 2-5 and 22-31
X Y	TOLOSA L. ET AL., Development of a Multiparametric Cell-based Protocol to Screen and Classify the Hepatotoxicity Potential of Drugs. Toxicol Sci, 13 February 2012, Vol. 127, No. 1, pages 187-198 [Retrieved on 2017-07-28] <DOI: 10.1093/TOXSCI/KFS083> Whole document	1 and 6-21 2-5 and 22-31
X Y	TENCZA S.B. AND SIPE M.A., Detection and Classification of Threat Agents via High-content Assays of Mammalian Cells. J Appl Toxicol, September 2004, Vol. 24, No. 5, pages 371-377 [Retrieved on 2017-07-28] <DOI: 10.1002/JAT.1025 > Whole document	1 and 6-21 2-5 and 22-31
X Y	XU J.J. ET AL., Cellular Imaging Predictions of Clinical Drug-Induced Liver Injury. Toxicol Sci, 3 June 2008, Vol. 105, No. 1, pages 97-105 [Retrieved on 2017-07-28] <DOI: 10.1093/TOXSCI/KFN109> Whole document	1 and 6-21 2-5 and 22-31
X Y	US 2012/0219204 A1 (HONG S.J. AND GHOSH R.N.) 30 August 2012 Whole document	1 and 6-21 2-5 and 22-31
Y	LOO L.-H. ET AL., Image-based multivariate profiling of drug responses from single cells. Nature Methods, 1 April 2007, Vol. 4, No. 5, pages 445-453 [Retrieved on 2017-07-28] <DOI: 10.1038/NMETH1032> Whole document and Supplementary Methods	2-5 and 22-31
A	SU R. ET AL., High-throughput imaging-based nephrotoxicity prediction for xenobiotics with diverse chemical structures. Arch Toxicol, 27 November 2015, Vol. 90, No. 11, pages 2793-2808 [Retrieved on 2017-07-28] <DOI: 10.1007/S00204-015-1638-Y> Whole document	-

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2017/050206

Supplemental Box
(Classification of Subject Matter)

Int. Cl.
C12Q 1/68 (2006.01)
G01N 33/533 (2006.01)
G01N 33/68 (2006.01)
G06T 7/00 (2017.01)
G06F 19/00 (2011.01)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2017/050206

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 2005/0014217 A1	20/01/2005	EP 1646867 A1	19/04/2006
		GB 2418731 A	05/04/2006
		US 2005/0014216 A1	20/01/2005
		US 2006/0234332 A1	19/10/2006
		US 2006/0228695 A1	12/10/2006
		WO 2005/008242 A1	27/01/2005
WO 2009/002565 A1	31/12/2008	CN 101784895 A	21/07/2010
		EP 2171443 A1	07/04/2010
US 2012/0219204 A1	30/08/2012	NONE	