

CLAIMS:

1. An in vitro method of predicting whether a test compound will be toxic for a specific cell type in vivo, the method comprising:

- (a) contacting at least one test population of cells with the test compound at a single concentration or over a range of concentrations,
- (b) labelling and imaging the cells with one or more biomolecular markers,
- (c) segmenting the cells and identifying whole-cell regions and one or more subcellular regions from the cells,
- (d) determining if cell loss or death has occurred at the highest test concentrations; if so, stop and predict the compound is toxic,
- (e) obtaining one or more quantified spatial-dependent and -independent phenotypic features in the test populations,
- (f) obtaining multiple dose response curves (DRCs) from the features,
- (g) obtaining quantified parameters of the DRCs wherein the DRC parameters are quantitated using the maximum response value $\Delta_{\max}$ for each phenotypic feature from a DRC of the test compound, and
- (h) comparing the quantitated DRC parameters to a reference set of quantitated DRC parameter data; said reference quantitated DRC parameter data being derived from two groups;
 - (i) compounds with known in vivo toxicity to the cell type, and (ii) compounds not known to be toxic to the cell type in vivo.

2. The method of claim 1, wherein said specific cell type is selected from the group comprising renal proximal tubular cells (PTCs), bronchial epithelial cells (BECs), and/or alveolar cells (AVCs).

3. The method of claim 1 or 2, wherein said one or more quantitated phenotypic features are associated with characteristics selected from the group comprising DNA damage response, actin filament integrity, whole cell morphology and cell count.

4. The method of any one of claims 1 to 3, wherein said one or more phenotypic features are quantitated based on (i) one or more of the spatial-dependent features selected from the group comprising textural features, spatial correlation features, and ratios of markers at different subcellular regions; and (ii) one or more of the spatial-independent features selected from the group comprising intensity features, cell count, and morphology.

5. The method of claim 4, wherein said textural features include one or more of the statistics of the Haralick's grey-level co-occurrence matrix (GLCM) at specific sub- or whole-cellular regions, namely mean correlation of DNA GLCM at the nuclear region; mean entropy of DNA GLCM at the nuclear region; mean angular second moment of DNA GLCM at the nuclear region; standard deviation of the sum variance of DNA GLCM at the nuclear region; mean sum entropy of actin GLCM at the whole cell region; mean entropy of actin GLCM at the whole cell region; standard deviation of the information measure of correlation 2 of γ H2AX GLCM at the whole-cell region; and mean sum average of γ H2AX GLCM at the whole cell region.

6. The method of claim 4 or 5, wherein said one or more of the spatial-dependent features comprises:

a) a staining intensity feature selected from one or more of the group comprising normalized spatial correlation coefficient between DNA and actin intensities at the whole cell region; total actin intensity level at the inner cytoplasmic region; normalized spatial correlation coefficient between DNA and γ H2AX intensities at the whole cell region; normalized spatial correlation coefficient between DNA and γ H2AX intensities at the nuclear region and coefficient of variation of the DNA intensity at the nuclear region; or

b) a staining intensity ratio feature selected from one or more of the group comprising ratio of the total γ H2AX to DNA intensities at the whole cell region; the ratio of the total γ H2AX to actin intensities at the nuclear region; and ratio of the total γ H2AX intensity levels at the nuclear region to the whole cell region.

7. The method of any one of claims 1 to 6, wherein the said one or more phenotypic features are selected from a group i)-iv) comprising:

i) mean sum entropy of the actin GLCM at the whole-cell region; coefficient of variation (CV) of the DNA intensity at the nuclear region; mean entropy of the actin GLCM at the whole-cell region; and mean angular second moment (ASM) of DNA GLCM at the nuclear region;

ii) total actin intensity level at the inner cytoplasmic region; mean angular second moment (ASM) of DNA GLCM at the nuclear region; standard deviation of the information measure of correlation 2 of γ H2AX GLCM at the whole-cell region; and cell count;

iii) normalized spatial correlation coefficient between DNA and γ H2AX intensities at the whole-cell region; normalized spatial correlation coefficient between DNA and actin intensities at the whole-cell region; mean sum average of γ H2AX GLCM at the whole-cell

region; ratio of the total γ H2AX to DNA intensities at the whole-cell region; and standard deviation of the sum variance of DNA GLCM at the nuclear region; and

iv) mean entropy of the DNA GLCM at the nuclear region; ratio of the total γ H2AX intensity levels at the nuclear region to the whole-cell region; mean correlation of actin GLCM; and mean correlation of DNA GLCM at the nuclear region.

8. The method of any one of claims 1 to 7, wherein cell toxicity is predicted using random-forest algorithm.

9. The method of any one of claims 1 to 8, wherein the at least one test population of cells are derived from somatic cells.

10. The method of any one of claims 1 to 9, wherein said contacting:

i) is performed over a period of time of at least 1-48 hours; and/or

ii) comprises adding the test compound to the at least one test population of cells at a concentration of about 1 μ g/ml to about 1000 μ g/ml.

11. The method of any one of claims 1 to 10, wherein said imaging techniques comprise high-throughput microscopy image capture.

12. A computer-implemented method of predicting in vivo cell toxicity of a test compound using at least one test population of the cells subjected to the test compound in vitro, the method comprising:

(a) receiving, by a computer processor, an image of the test population of the cells;

(b) extracting, by the computer processor, one or more spatial-dependent phenotypic features associated with the test population of cells from the image, the one or more spatial-dependent phenotypic feature characterizing a spatial distribution of biomolecules associated with the cells;

(c) obtaining one or more quantitated dose response curve (DRC) parameters describing the DRC of the respective one or more spatial-dependent phenotypic features, wherein the quantitated DRC parameter is obtained using the maximum response value $\Delta_{\max}$; and

(d) inputting said one or more quantitated DRC parameters to a predictive model to generate a prediction of in vivo cell toxicity of the test compound, wherein the cells are preferably renal proximal tubular cells (PTCs), bronchial epithelial cells (BECs), or alveolar cells (AVCs).

13. A method according to claim 12, wherein said image comprises a plurality of images each representing the test population of cells imaged using a respective imaging channel emphasizing a type of biomolecules associated with the cells.

14. A method according to claim 13, wherein each of the plurality of images represents a distribution of a type of biomarkers targeting the corresponding type of biomolecules.

15. A method according to any one of claims 12 to 14, wherein step (b) comprises segmenting the cells using the image, and extracting the one or more spatial-dependent phenotypic features using intensity values of the image corresponding to the segmented cells.

16. A method according to any one of claims 12 to 15, wherein the one or more spatial-dependent phenotypic features are selected from the group comprising features characterizing DNA structure alterations, chromatin structure alterations and Actin filament structure alterations of the cells.

17. A method according to any one of claims 12 to 16, wherein the predicative model is obtained using a supervised learning algorithm trained with a set of training data.

18. A method according to claim 17, wherein said set of training data comprises a plurality of candidate quantitated spatial-dependent dose response curve (DRC) parameters characterizing a corresponding plurality of spatial-dependent phenotypic features associated with control populations of cells; said control populations of cells having been respectively subjected to: (i) compounds known to be toxic to the cells in vivo; (ii) compounds not known to be toxic to the cells in vivo.

19. A method according to any one of claims 12 to 18 further comprises extracting one or more spatial-independent phenotypic features associated with the at least one test population of cells, and obtaining the one or more quantitated dose response curve (DRC) parameters further using the one or more spatial-independent phenotypic features.

20. A method according to any one of claims 12 to 19, wherein step (c) comprises obtaining the quantitated DRC parameter at a pre-defined concentration of the test compound.