

[0006] With the methods of pooled screening of cells from different genetic backgrounds provided herein, individual cell lines from various healthy and/or patient lines are labelled with a unique genetic barcode, which is designed to be processed via *in situ* sequencing after any assay. Phenotypes can then be classified and linked back to the original cell lines' genotypes, enabling statistical genetics analysis to be conducted in highly-controlled *in vitro* assays. This technology also allows pooled screening studies to be applied to cell lines with similar background genetics (*i.e.*, trios and isogenically engineered lines), as well as multiplexed with perturbation-based genetic screens. Also herein are machine learning-enabled improvements of genetic barcoding methods.

[0007] The methods provided herein solve the problem of being unable to conduct large-population studies *in vitro* without facing difficult, costly, and artifact-sensitive scaling challenges. The pooled approach greatly reduces the effects of evaporation, temperature, and other plate-wise or well-wise artifacts that can obscure larger screens. These methods also enable population-genetics based assays to be conducted on specific cell types in a highly controlled manner, rather than on clinical outcomes which are extremely variable and affected by countless convoluting factors. Other methods, as described above, involve either sequencing-only readouts or the barcoding of the perturbagen (pooled optical barcoding), which miss the importance of non-coding variants in a given assay.

[0008] All references cited herein, including patent applications, patent publications, and scientific literature, are herein incorporated by reference in their entirety, as if each individual reference were specifically and individually indicated to be incorporated by reference.

SUMMARY

[0009] Provided herein is a method of pooled screening of cells from different genetic backgrounds, comprising: a) labeling two or more populations of cells of different genetic backgrounds with two or more unique nucleic acid barcode sequences, each unique nucleic acid barcode sequence corresponding to a different population of cells; b) combining the two or more populations of cells to obtain a single mixed population of cells; c) performing *in situ* single-cell sequencing on the cells; and d) analyzing known or identifying new phenotypes of the cells in the mixed population. In some embodiment examples, the two or more populations of cells are from

different cell lines. In some [embodimentexamples](#), the different cell lines are healthy cell lines. In some [embodimentexamples](#), the different cell lines are patient cell lines. In some [embodimentexamples](#), the different cell lines are isogenically engineered cell lines. In some [embodimentexamples](#), the different cell lines include any combination of healthy, diseased, and isogenically engineered cell lines. In some [embodimentexamples](#), the cells are induced pluripotent stem cells (iPSCs). In some [embodimentexamples](#), the iPSCs are differentiated prior to analyzing the phenotype. In some [embodimentexamples](#), the method further comprises culturing the cells prior to analyzing the phenotype. In some [embodimentexamples](#), the single mixed population of cells is on a substrate or in three-dimensional culture. In some [embodimentexamples](#), the substrate is a cell culture dish. In some [embodimentexamples](#), the method further comprises performing single cell RNAseq. In some [embodimentexamples](#), the method further comprises growing the two or more populations of cells for two or more generations prior to step b). In some [embodimentexamples](#), the method comprises stably integrating the unique nucleic acid barcode sequences into the genomes of the two or more populations of cells. In some [embodimentexamples](#), the unique nucleic acid barcode sequences are delivered into the cell using a virus. In some [embodimentexamples](#), the virus is a lentivirus. In some [embodimentexamples](#), the virus encodes a selectable marker. In some [embodimentexamples](#), the selectable marker is an antibiotic resistance gene. In some [embodimentexamples](#), the virus encodes a fluorescent protein. In some [embodimentexamples](#), each unique nucleic acid barcode sequence is at least 1 base pair in length. In some [embodimentexamples](#), each unique nucleic acid barcode sequence is 1 to about 18 base pairs in length. In some [embodimentexamples](#), each unique nucleic acid barcode sequence is 8 base pairs in length. In some [embodimentexamples](#), the two or more populations of cells are sequenced prior to labeling with the unique nucleic acid barcode sequences. In some [embodimentexamples](#), the sequencing is whole genome sequencing. In some [embodimentexamples](#), the two or more populations of cells were obtained from related individuals. In some [embodimentexamples](#), the two or more populations of cells were obtained from humans. In some [embodimentexamples](#), the method comprises labeling ten or more populations of cells of different genetic backgrounds with ten or more unique nucleic acid barcode sequences, each unique nucleic acid barcode sequence corresponding to a different population of cells. In some [embodimentexamples](#), analyzing the phenotype of the cells comprises an assay selected from the group consisting of high content imaging, calcium imaging, immunohistochemistry, cell morphology imaging, protein aggregation imaging, cell-cell interaction imaging, live cell imaging, and any other imaging-based assay modality. In some [embodimentexamples](#), step d) comprises

analyzing the phenotype of the cells by capturing a microscopic image or a time series of microscopic images of the cells and evaluating phenotypic features presented in the image or images. In some [embodimentexamples](#), the method further comprises a computer-implemented technique for aligning between a first plurality of images and a second plurality of images, comprising: a) generating a first reference coordinate space of the first plurality of images, wherein the first plurality of images is of a well on a culture plate; b) extracting a first patch of the first plurality of images; c) generating a second reference coordinate space of the second plurality of images, wherein the second plurality of images is of the well on the culture plate; d) extracting a second patch of the second plurality of images; e) computing an affine transformation function between the first patch and the second patch to obtain a plurality of transformation parameters; and f) generating a coordinate transformation function between the first reference coordinate space and the second reference coordinate space based on the plurality of transformation parameters.

[0010] Also provided herein is a computer-implemented method for aligning between a first plurality of images and a second plurality of images, comprising: a) generating a first reference coordinate space of the first plurality of images, wherein the first plurality of images is of a well on a culture plate; b) extracting a first patch of the first plurality of images; c) generating a second reference coordinate space of the second plurality of images, wherein the second plurality of images is of the well on the culture plate; d) extracting a second patch of the second plurality of images; e) computing an affine transformation function between the first patch and the second patch to obtain a plurality of transformation parameters; and f) generating a coordinate transformation function between the first reference coordinate space and the second reference coordinate space based on the plurality of transformation parameters. In some [embodimentexamples](#), the first plurality of images is a plurality of barcoding images. In some [embodimentexamples](#), the second plurality of images is a plurality of marker-based/marker-free readout image. In some [embodimentexamples](#), the first plurality of images and the second plurality of images provide different coverage of the well. In some [embodimentexamples](#), the first plurality of images and the second plurality of images are taken at different times. In some [embodimentexamples](#), the first plurality of images and the second plurality of images have different resolutions. In some [embodimentexamples](#), the first plurality of images is taken by a first microscope and the second plurality of images is taken by a second microscope. In some [embodimentexamples](#), the first microscope is a fluorescent microscope. In some [embodimentexamples](#), the second microscope is a non-fluorescent microscope. In some [embodimentexamples](#),

the first plurality of images is captured by a first imager. In some [embodimentexamples](#), the method further comprises detecting one or more physical characteristics of the well in the first plurality of images; and generating the first reference coordinate space based on the detected one or more physical characteristics. In some [embodimentexamples](#), the one or more physical characteristics of the well comprises: a shape of the well, an edge of the well, a location of the well. In some [embodimentexample](#), two or more images in the first plurality of images are offset from each other by an overlap ratio, and wherein the first reference coordinate space is generated based on the overlap ratio. In some [embodimentexamples](#), the first reference coordinate space is generated based on metadata of the first imager. In some [embodimentexamples](#), the method further comprises selecting one or more marker images from the first plurality of images based on one or more landmarks captured in the one or more images, wherein the first patch is obtained from the one or more marker images from the first plurality of images. In some [embodimentexamples](#), the one or more landmarks comprise: one or more cells, one or more well boundaries, one or more beads, one or more nuclei, or any combination thereof. In some [embodimentexamples](#), the second plurality of images is captured by a second imager. In some [embodimentexamples](#), the method further comprises detecting one or more physical characteristics of the well in the second plurality of images; and generating the second reference coordinate space based on the detected one or more physical characteristics. In some [embodimentexamples](#), the one or more physical characteristics of the well comprises: a shape of the well, an edge of the well, a location of the well. In some [embodimentexamples](#), two or more images in the second plurality of images are offset from each other by an overlap ratio, and wherein the second reference coordinate space is generated based on the overlap ratio. In some [embodimentexample](#), the second reference coordinate space is generated based on metadata of the second imager. In some [embodimentexamples](#), the method further comprises selecting one or more marker images from the second plurality of images based on one or more landmarks captured in the one or more images, wherein the second patch is obtained from the one or more marker images of the second plurality of images. In some [embodimentexamples](#), the one or more landmarks comprise: one or more cells, one or more well boundaries, one or more beads, one or more nuclei, or any combination thereof. In some [embodimentexamples](#), the first patch covers a center of the first image. In some [embodimentexamples](#), the second patch covers a center of the second image. In some [embodimentexamples](#), at least a portion of the first patch and at least a patch of the second patch correspond to the same subject. In some [embodimentexamples](#), the transformation parameters comprise one or more of: a translation parameter, a

scaling parameter, and a rotation parameter. In some embodimentexamples, the first and second images are obtained from an assay of pooled screening of cells from different genetic backgrounds, comprising: a) labeling two or more populations of cells of different genetic backgrounds with two or more unique nucleic acid barcode sequences, each unique nucleic acid barcode sequence corresponding to a different population of cells; b) combining the two or more populations of cells to obtain a single mixed population of cells; c) performing *in situ* single-cell sequencing on the cells; and d) analyzing known or identifying new phenotypes of the cells in the mixed population.

[0011] In some embodimentexamples, the method further comprises utilizing a classifier configured to receive an image and output a classification result. In some embodimentexamples, the classifier comprises a plurality of layers. In some embodimentexamples, the classifier is a convolutional neural network. In some embodimentexamples, the classifier is a DenseNet classifier. In some embodimentexamples, the classification result is a classification based on the genetic background of each single cell captured in the image. In some embodimentexamples, the method further comprises generating an embedding of the image. In some embodimentexamples, the embedding is generated from an activation output layer prior to the last layer of the classifier. In some embodimentexamples, the embedding is reduced in dimension using a linear dimensionality reduction method such as UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) algorithm, principal component analysis (PCA), t-distributed stochastic neighborhood embedding (t-SNE), etc. In some embodimentexamples, the embedding is reduced in dimension using a UMAP algorithm to obtain one or more UMAP plots for visualization of the embedding. In some embodimentexamples, the method further comprises evaluating a treatment based on at least a portion of the embedding. In some embodimentexamples, the method further comprises evaluating a treatment based on at least a portion of the one or more UMAP plots.

[0011a] In one aspect, the present disclosure refers to a method of pooled screening of cells from different genetic backgrounds, comprising: labeling two or more populations of cells of different genetic backgrounds with two or more unique nucleic acid barcode sequences, each unique nucleic acid barcode sequence corresponding to a different population of cells; combining the two or more populations of cells to obtain a single mixed population of cells in a well on a culture plate; performing in situ single-cell sequencing on the cells; obtaining a first plurality of images, wherein the first plurality of images comprises one or more barcoding images of the well on the culture plate; obtaining a second plurality of images, wherein the second plurality of images comprises one or more marker-based or marker-free readout images of the well on the culture plate; generating a first reference coordinate space of the first plurality of images; extracting a first patch of the first plurality of images; generating a second reference coordinate space of the second plurality of images; extracting a second patch of the second plurality of images; computing an affine transformation function between the first patch and the second patch to obtain a plurality of transformation parameters; and generating a coordinate transformation function between the first reference coordinate space and the second reference coordinate space based on the plurality of transformation parameters; identifying cell-specific image data for one or more cells in the single mixed population of cells by correlating at least one image of the first plurality of images with at least one image of the second plurality of images based on the coordinate transformation function; and identifying one or more phenotypes of the one or more cells based on the cell-specific image data.

[0012] It is to be understood that one, some, or all of the properties of the various embodimentexamples described herein may be combined to form other embodimentexamples of the present inventiondisclosure. These and other aspectexamples of the inventiondisclosure will become apparent to one of skill in the art. These and other embodimentexamples of the inventiondisclosure are further described by the detailed description that follows.