

## CLAIMS:

1. A method of analyzing tissue features based on multiplexed biometric image data comprising:

storing a data set comprising cell profile data including multiplexed biometric images capturing the expression of a plurality of biomarkers with respect to a plurality of fields of view in which individual cells are delineated and segmenting into compartments, wherein the cell profile data is generated from a plurality of tissue samples drawn from a cohort of patients having a commonality, the data set further comprising an association of the cell profile data with at least one piece of meta-information including a field of view level assessment or a patient-level assessment related to the commonality;

generating a plurality of sets of clusters of similar cells from the data set, wherein each of the plurality of sets of clusters comprises a unique number of clusters, wherein each cell is assigned to a single cluster in each of the plurality of sets of clusters, wherein each of the plurality of clusters in each of the plurality of sets of clusters comprises cells having a plurality of selected attributes more similar to the plurality of selected attributes of other cells in that cluster than to the plurality of selected attributes of cells in other clusters in the set;

within each of the plurality of sets of clusters, observing a proportion of the cells assigned to each cluster;

examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment or the patient-level assessment related to the commonality; and

selecting one of the plurality of sets of clusters comprising a predictive set of clusters based on a comparison of the performance of at least one model of the plurality of sets of clusters.

2. The method of claim 1 wherein data set is associated with a plurality of batches, the method further comprising:

normalizing the cell profile data with respect to the plurality of batches by subtracting a median intensity of the whole cell for all cells within one of the plurality of batches from each of a median intensity of the whole cell, a median intensity of the nucleus, a median intensity of the membrane, and a median intensity of the cytoplasm for each cell in the batch;

wherein generating a plurality of sets of clusters comprises generating a plurality of sets of clusters of similar cells from the normalized data set.

3. The method of claim 1 wherein cell similarity is based on a comparison of at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.
4. The method of claim 1 wherein the at least one attribute of a cell is selected from four features of a cell consisting of a median intensity of the whole cell, a nucleus intensity ratio, a membrane intensity ratio, and a cytoplasm intensity ratio,

wherein the nucleus intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the cytoplasm from the median intensity of the nucleus;

wherein the membrane intensity ratio is calculated by subtracting half of the sum of the median intensity of the nucleus and the median intensity of the cytoplasm from the median intensity of the membrane; and

wherein the cytoplasm intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the nucleus from the median intensity of the cytoplasm.

5. The method of claim 1 wherein cell similarity is based on a comparison of at least two attributes of a cell, wherein each of the at least two attributes is based on the expression of the at least one of the plurality of biomarkers.
6. The method of claim 1 wherein cell similarity is based on a comparison of at least three attributes of a cell, wherein each of the at least three attributes is based on the expression of the at least one of the plurality of biomarkers.
7. The method of claim 1 wherein cell similarity is based on a comparison of at least four attributes of a cell, wherein each of the at least four attributes is based on the expression of the at least one of the plurality of biomarkers.
8. The method of claim 1 wherein cell profiles of normal cells are excluded from the data set used to generate the plurality of sets of clusters of similar cells.

9. The method of claim 1 further comprising determining the similarity of cells by applying a K-medians clustering algorithm to at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

10. The method of claim 1 further comprising determining the similarity of cells by applying a K-means clustering algorithm to at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

11. The method of claim 1 wherein the observed proportion of cells is the observed proportion of the cells of each field of view assigned to each cluster.

12. The method of claim 1

wherein examining the observed proportions comprises examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment related to the commonality; and

wherein selecting a predictive set of clusters comprises selecting a predictive set of clusters based on a comparison of the performance of the field of view level assessment models based on the plurality of sets of clusters.

13. The method of claim 1 wherein the observed proportion of cells is the observed proportion of the cells of each patient assigned to each cluster.

14. The method of claim 1

wherein examining the observed proportions comprises examining the observed proportions for an association with a prognosis [survival time] of a condition or a disease; and

wherein selecting one of the plurality of sets of clusters comprises selecting one of the plurality of sets of clusters based on a comparison of a performance of a patient level assessment model based on the plurality of sets of clusters.

15. The method of claim 1 wherein the cell data comprises training data and test data, wherein the plurality of sets of clusters of similar cells are generated from training data, and wherein the performance of the at least one model for comparison is determined from the testing data.

16. The method of claim 1 further comprising

comparing the performance of at least one model with respect to the number of clusters in each of the plurality of sets of clusters.

17. The method of claim 1 wherein selecting a predictive set of clusters further comprises selecting one of the plurality of sets of clusters having a number of clusters above which a greater number of clusters in the set of cluster does not offer a statistically significant increase in performance.

18. The method of claim 1 wherein selecting a predictive set of clusters further comprises selecting one of the plurality of sets of clusters having a number of clusters below which a greater number of clusters in the set of cluster provides a decrease in performance.

19. The method of claim 1 further comprising examining the observed proportions in the selected set of clusters for a univariate association with the at least one piece of meta-information.

20. The method of claim 1 further comprising examining the observed proportions in the selected set of clusters for a multivariate association with the at least one piece of meta-information.

21. The method of claim 1 further comprising selecting a predictive set of clusters based on a performance of the at least one model of the set of clusters corresponding to a concordance of greater than a threshold.

22. The method of claim 1 further comprising identifying at least one predictive cluster from the predictive set of clusters.

23. A method of analyzing cell cluster features based on multiplexed biometric images comprising:

storing a data set comprising cell profile data including multiplexed biometric images capturing the expression of a plurality of biomarkers with respect to a plurality of fields of view in which individual cells are delineated and segmenting into compartments;

identifying a first cluster in a plurality of clusters of similar cells from the data set, wherein each cell is assigned to one of the plurality of clusters, wherein each cluster in the plurality of clusters includes cells having a plurality of selected attributes more similar to the plurality of selected attributes of other cells in that cluster than to the plurality of selected attributes of cells in other clusters in the set;

creating a montage of a first cell in the first cluster, wherein the montage comprises a portion of at least some multiplexed images describing the first cell's expression of each of a plurality of biomarkers, wherein each portion of the at least some images includes the first cell and a small region of interest around the first cell; and

displaying the montage of the first cell in the first cluster to enable a user to understand a feature of the first cluster.

24. The method of claim 23 wherein the montage of the first cell comprises a series of juxtaposed portions of the at least some images of a field of view describing the first cell's expression of each of a plurality of biomarkers.

25. The method of claim 23 wherein the montage of the first cell comprises a series of superimposed portions of the at least some images of a field of view describing the first cell's expression of each of a plurality of biomarkers.

26. The method of claim 23 further comprising:

creating a montage of a second cell in the first cluster, wherein the montage comprises a portion of at least some images of a field of view describing the second cell's expression of each of a plurality of biomarkers, wherein each portion of the at least some images includes the second cell and a small region of interest around the second cell; and

displaying the montage of the second cell in the first cluster to enable a user to understand the feature of the first cluster.

27. The method of claim 23 further comprising:

displaying the montage of the first cell in the first cluster and the montage of the second cell in the first cluster simultaneously to enable a user to understand the feature of the first cluster.

28. A system for analyzing tissue features based on multiplexed biometric image data comprising:

a storage device for storing a data set comprising cell profile data including multiplexed biometric images capturing the expression of a plurality of biomarkers with respect to a plurality of fields of view in which individual cells are delineated and segmenting into compartments, wherein the cell profile data is generated from a plurality of tissue samples drawn from a cohort

of patients having a commonality, the data set further comprising an association of the cell profile data with at least one piece of meta-information including a field of view level assessment or a patient-level assessment related to the commonality;

at least one processor for executing code that causes the at least one processor to perform the steps of:

generating a plurality of sets of clusters of similar cells from the data set, wherein each of the plurality of sets of clusters comprises a unique number of clusters, wherein each cell is assigned to a single cluster in each of the plurality of sets of clusters, wherein each of the plurality of clusters in each of the plurality of sets of clusters comprises cells having a plurality of selected attributes more similar to the plurality of selected attributes of other cells in that cluster than to the plurality of selected attributes of cells in other clusters in the set;

within each of the plurality of sets of clusters, observing a proportion of the cells assigned to each cluster; and

examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment or the patient-level assessment related to the commonality; and

a visual display device that enables one of the plurality of sets of clusters, comprising a predictive set of clusters, to be selected based on a comparison of the performance of at least one model of the plurality of sets of clusters.

29. The system of claim 28 wherein data set is associated with a plurality of batches, and wherein the at least one processor further executes code that causes the at least one processor to perform the steps of:

normalizing the cell profile data with respect to the plurality of batches by subtracting a median intensity of the whole cell for all cells within one of the plurality of batches from each of a median intensity of the whole cell, a median intensity of the nucleus, a median intensity of the membrane, and a median intensity of the cytoplasm for each cell in the batch;

wherein generating a plurality of sets of clusters comprises generating a plurality of sets of clusters of similar cells from the normalized data set.

30. The system of claim 28 wherein cell similarity is based on a comparison of at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

31. The system of claim 28 wherein the at least one attribute of a cell is selected from four features of a cell consisting of a median intensity of the whole cell, a nucleus intensity ratio, a membrane intensity ratio, and a cytoplasm intensity ratio,

wherein the nucleus intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the cytoplasm from the median intensity of the nucleus;

wherein the membrane intensity ratio is calculated by subtracting half of the sum of the median intensity of the nucleus and the median intensity of the cytoplasm from the median intensity of the membrane; and

wherein the cytoplasm intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the nucleus from the median intensity of the cytoplasm.

32. The system of claim 28 wherein the at least one processor determines cell similarity based on a comparison of at least two attributes of a cell, wherein each of the at least two attributes is based on the expression of the at least one of the plurality of biomarkers.

33. The system of claim 28 wherein the at least one processor determines cell similarity based on a comparison of at least three attributes of a cell, wherein each of the at least three attributes is based on the expression of the at least one of the plurality of biomarkers.

34. The system of claim 28 wherein the at least one processor determines cell similarity based on a comparison of at least four attributes of a cell, wherein each of the at least four attributes is based on the expression of the at least one of the plurality of biomarkers.

35. The system of claim 28 wherein the at least one processor further executes code that causes the at least one processor to perform the step of excluding cell profiles of normal cells from the data set used to generate the plurality of sets of clusters of similar cells.

36. The system of claim 28 wherein the at least one processor determines the similarity of cells by applying a K-medians clustering algorithm to at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

37. The system of claim 28 wherein the at least one processor determines the similarity of cells by applying a K-means clustering algorithm to at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

38. The system of claim 28 wherein the observed proportion of cells comprises the observed proportion of the cells of each field of view assigned to each cluster.

39. The system of claim 28

wherein examining the observed proportions comprises examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment related to the commonality; and

wherein selecting a predictive set of clusters comprises selecting a predictive set of clusters based on a comparison of the performance of the field of view level assessment models based on the plurality of sets of clusters.

40. The system of claim 28 wherein the observed proportion of cells is the observed proportion of the cells of each patient assigned to each cluster.

41. The system of claim 28

wherein examining the observed proportions comprises examining the observed proportions for an association with a prognosis [survival time] of a condition or a disease; and

wherein selecting one of the plurality of sets of clusters comprises selecting one of the plurality of sets of clusters based on a comparison of a performance of a patient level assessment model based on the plurality of sets of clusters.

42. The system of claim 28 wherein the at least one processor further divides the cell data into training data and test data, generates the plurality of sets of clusters of similar cells from training data, and determines the performance of the at least one model for comparison from the testing data.

43. The system of claim 28 wherein the at least one processor further executes code that causes the at least one processor to perform the step of:

comparing the performance of at least one model with respect to the number of clusters in each of the plurality of sets of clusters.

44. The system of claim 28 wherein the visual display device further enables selection of one of the plurality of sets of clusters having a number of clusters above which a greater number of clusters in the set of cluster does not offer a statistically significant increase in performance.

45. The system of claim 28 wherein the visual display device further enables selection of one of the plurality of sets of clusters having a number of clusters below which a greater number of clusters in the set of cluster provides a decrease in performance.

46. The system of claim 28 further comprising examining the observed proportions in the selected set of clusters for a univariate association with the at least one piece of meta-information.

47. The system of claim 28 further comprising examining the observed proportions in the selected set of clusters for a multivariate association with the at least one piece of meta-information.

48. The system of claim 28 the visual display device further enables selection of one of the plurality of sets of clusters based on a performance of the at least one model of the set of clusters corresponding to a concordance of greater than a threshold.

49. The system of claim 28 wherein the at least one processor further executes code that causes the at least one processor to perform the step of identifying at least one predictive cluster from the predictive set of clusters.

50. A system for analyzing tissue features based on multiplexed biometric image data comprising:

a storage device for storing a data set comprising cell profile data including multiplexed biometric images capturing the expression of a plurality of biomarkers with respect to a plurality of fields of view in which individual cells are delineated and segmenting into compartments; and

a visual display device that enables a first cluster in a plurality of clusters of similar cells from the data set to be identified, wherein each cell is assigned to one of the plurality of clusters, wherein each cluster in the plurality of clusters includes cells having a plurality of selected attributes more similar to the plurality of selected attributes of other cells in that cluster than to the plurality of selected attributes of cells in other clusters in the set; and

at least one processor for executing code that causes the at least one processor to create a montage of a first cell in the first cluster, wherein the montage comprises a portion of at least some multiplexed images describing the first cell's expression of each of a plurality of biomarkers, wherein each portion of the at least some images includes the first cell and a small region of interest around the first cell;

wherein the visual display device further displays the montage of the first cell in the first cluster to enable a user to understand a feature of the first cluster.

51. The system of claim 50 wherein the montage of the first cell comprises a series of juxtaposed portions of the at least some images of a field of view describing the first cell's expression of each of a plurality of biomarkers.

52. The system of claim 50 wherein the montage of the first cell comprises a series of superimposed portions of the at least some images of a field of view describing the first cell's expression of each of a plurality of biomarkers.

53. The system of claim 50 further comprising:

wherein the at least one processor further creates a montage of a second cell in the first cluster, wherein the montage comprises a portion of at least some images of a field of view describing the second cell's expression of each of a plurality of biomarkers, wherein each portion of the at least some images includes the second cell and a small region of interest around the second cell; and

wherein the visual display device further displays the montage of the second cell in the first cluster to enable a user to understand the feature of the first cluster.

54. The method of claim 23 wherein the visual display device further displays the montage of the first cell in the first cluster and the montage of the second cell in the first cluster simultaneously to enable a user to understand the feature of the first cluster.

#### CLAIMS:

1. A method of analyzing tissue features based on multiplexed biometric image data comprising:

storing a data set comprising cell profile data including multiplexed biometric images capturing the expression of a plurality of biomarkers with respect to a plurality of fields of view in which individual cells are delineated and segmented into compartments, wherein the cell profile data is generated from a plurality of tissue samples drawn from a cohort of patients having a commonality, the data set further comprising an association of the cell profile data with at least one piece of meta-information including a field of view level assessment or a patient-level assessment related to the commonality;

generating a plurality of sets of clusters of similar cells from the data set, wherein each of the plurality of sets of clusters comprises a unique number of clusters, wherein each cell is assigned to a single cluster in each of the plurality of sets of clusters, wherein each of the plurality of clusters in each of the plurality of sets of clusters comprises cells having a plurality of selected attributes more similar to the plurality of selected attributes of other cells in that cluster than to the plurality of selected attributes of cells in other clusters in the set;

within each of the plurality of sets of clusters, observing a proportion of the cells assigned to each cluster;

examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment or the patient-level assessment related to the commonality; and

selecting one of the plurality of sets of clusters comprising a predictive set of clusters based on a comparison of the performance of at least one model of the plurality of sets of clusters.

2. The method of claim 1 wherein the data set is associated with a plurality of batches, the method further comprising:

normalizing the cell profile data with respect to the plurality of batches by subtracting a median intensity of the whole cell for all cells within one of the plurality of batches from each of a median intensity of the whole cell, a median intensity of the nucleus, a median intensity of the membrane, and a median intensity of the cytoplasm for each cell in the batch;

wherein generating a plurality of sets of clusters comprises generating a plurality of sets of clusters of similar cells from the normalized data set.

3. The method of claim 1 wherein cell similarity is based on a comparison of at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

4. The method of claim 1 wherein the at least one attribute of a cell is selected from four features of a cell consisting of a median intensity of the whole cell, a nucleus intensity ratio, a membrane intensity ratio, and a cytoplasm intensity ratio,

wherein the nucleus intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the cytoplasm from the median intensity of the nucleus;

wherein the membrane intensity ratio is calculated by subtracting half of the sum of the median intensity of the nucleus and the median intensity of the cytoplasm from the median intensity of the membrane; and

wherein the cytoplasm intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the nucleus from the median intensity of the cytoplasm.

5. The method of claim 1 wherein cell similarity is:

(a) based on a comparison of at least two attributes of a cell, wherein each of the at least two attributes is based on the expression of the at least one of the plurality of biomarkers;

(b) based on a comparison of at least three attributes of a cell, wherein each of the at least three attributes is based on the expression of the at least one of the plurality of biomarkers; or

(c) based on a comparison of at least four attributes of a cell, wherein each of the at least four attributes is based on the expression of the at least one of the plurality of biomarkers.

6. The method of claim 1 wherein cell profiles of normal cells are excluded from the data set used to generate the plurality of sets of clusters of similar cells.

7. The method of claim 1 further comprising determining the similarity of cells by applying a K-medians clustering algorithm to at least one attribute of a cell based on the expression of at least one of the plurality of biomarkers.

8. The method of claim 1 wherein:

(a) the observed proportion of cells is the observed proportion of the cells of each field of view assigned to each cluster; or

(b) examining the observed proportions comprises examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment related to the commonality; and wherein selecting a predictive set of clusters comprises selecting a predictive set of clusters based on a comparison of the performance of the field of view level assessment models based on the plurality of sets of clusters.

9. The method of claim 1 wherein:

(a) the observed proportion of cells is the observed proportion of the cells of each patient assigned to each cluster; or

(b) examining the observed proportions comprises examining the observed proportions for an association with a prognosis [survival time] of a condition or a disease; and wherein selecting one of the plurality of sets of clusters comprises selecting one of the plurality of sets of clusters based on a comparison of a performance of a patient level assessment model based on the plurality of sets of clusters.

10. The method of claim 1 wherein the cell data comprises training data and test data, wherein the plurality of sets of clusters of similar cells are generated from training data, and wherein the performance of the at least one model for comparison is determined from the testing data.

11. The method of claim 1 further comprising comparing the performance of at least one model with respect to the number of clusters in each of the plurality of sets of clusters.

12. The method of claim 1 wherein:

(a) selecting a predictive set of clusters further comprises selecting one of the plurality of sets of clusters having a number of clusters above which a greater number of clusters in the set of cluster does not offer a statistically significant increase in performance; or

(b) selecting a predictive set of clusters further comprises selecting one of the plurality of sets of clusters having a number of clusters below which a greater number of clusters in the set of cluster provides a decrease in performance.

13. The method of claim 1, further comprising identifying at least one predictive cluster from the predictive set of clusters.

14. A system for analyzing tissue features based on multiplexed biometric image data comprising:

a storage device for storing a data set comprising cell profile data including multiplexed biometric images capturing the expression of a plurality of biomarkers with respect to a plurality of fields of view in which individual cells are delineated and segmented into compartments, wherein the cell profile data is generated from a plurality of tissue samples drawn from a cohort of patients having a commonality, the data set further comprising an association of the cell profile data with at least one piece of meta-information including a field of view level assessment or a patient-level assessment related to the commonality;

at least one processor for executing code that causes the at least one processor to perform the steps of:

generating a plurality of sets of clusters of similar cells from the data set, wherein each of the plurality of sets of clusters comprises a unique number of clusters, wherein each cell is assigned to a single cluster in each of the plurality of sets of clusters, wherein each of the plurality of clusters in each of the plurality of sets of clusters comprises cells having a plurality of selected attributes more similar to the plurality of selected attributes of other cells in that cluster than to the plurality of selected attributes of cells in other clusters in the set;

within each of the plurality of sets of clusters, observing a proportion of the cells assigned to each cluster; and

examining the observed proportions for an association with the at least one piece of meta-information including the field of view level assessment or the patient-level assessment related to the commonality; and

a visual display device that enables one of the plurality of sets of clusters, comprising a predictive set of clusters, to be selected based on a comparison of the performance of at least one model of the plurality of sets of clusters.

15. The system of claim 14 wherein the at least one attribute of a cell is selected from four features of a cell consisting of a median intensity of the whole cell, a nucleus intensity ratio, a membrane intensity ratio, and a cytoplasm intensity ratio,

wherein the nucleus intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the cytoplasm from the median intensity of the nucleus;

wherein the membrane intensity ratio is calculated by subtracting half of the sum of the median intensity of the nucleus and the median intensity of the cytoplasm from the median intensity of the membrane; and

wherein the cytoplasm intensity ratio is calculated by subtracting half of the sum of the median intensity of the membrane and the median intensity of the nucleus from the median intensity of the cytoplasm.