CLAIMS

What is claimed:

1. A method for visualizing a pattern in a discrete attribute value dataset, the method comprising:

at a computer system comprising one or more processing cores and a memory:

storing the discrete attribute value dataset in the memory, wherein the discrete attribute value dataset comprises:

a respective discrete attribute value for each corresponding gene in a plurality of genes, for each respective cell in a plurality of cells, and

a respective ATAC fragment count for each corresponding ATAC peak in a plurality of ATAC peaks, for each respective cell in the plurality of cells, wherein

the plurality of cells comprises 100 cells,

the plurality of genes comprises 100 genes,

the plurality of ATAC peaks comprises 50 ATAC peaks,

each respective cell in the plurality of cells is assigned to a respective cluster group in a first plurality of cluster groups based on a first clustering of discrete attribute values for the plurality of genes across the plurality of cells, and

each respective cell in the plurality of cells is also assigned to a respective cluster group in a second plurality of cluster groups based on a second clustering of ATAC fragment count values for the plurality of ATAC peaks across the plurality of cells;

displaying, in a first panel, a two-dimensional projection of the plurality of cells based on assignment of the plurality of cells to one of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups; and

indicating within the two-dimensional projection, for each respective cell in the plurality of cells, membership in the other of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups, thereby visualizing the pattern in the discrete attribute value dataset.

2. The method of claim 1, wherein membership of each respective cell in the plurality of cells in the other of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups is

indicated by coloring the respective cell a color that is uniquely associated with a cluster group to which the respective cell has been assigned in the other of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups.

3. The method of claim 1, the method further comprising:

computing, for each respective gene in the plurality of genes for each respective cluster in the first plurality of clusters or the second plurality of clusters, a difference in the discrete attribute value for the respective gene across the respective subset of cells in the respective cluster relative to the discrete attribute value for the respective gene across the first plurality of clusters or the second plurality of clusters other than the respective cluster, thereby deriving a differential value for each respective gene in the plurality of gene for each respective cluster in the first plurality of clusters or the second plurality of clusters, and

displaying in a second panel, concurrently with the first panel, a heat map that comprises a representation of the differential value for each respective gene in the plurality of genes for each respective cluster in the first plurality of clusters or the second plurality of clusters thereby visualizing the pattern in the discrete attribute value dataset.

4. The method of claim 3, wherein the differential value for each respective gene in the plurality of genes for each respective cluster in the first plurality of clusters or the second plurality of clusters is a fold change in (i) a first measure of central tendency of the discrete attribute value for the respective gene measured in each of the cells in the plurality of cells in the respective cluster and (ii) a second measure of central tendency of the discrete attribute value for the respective gene measured in each of the cells of all clusters in the first plurality of clusters or all clusters in the second plurality of clusters, other than the first respective cluster.

5. The method of claim 4, wherein the fold change is a $\log_2$ fold change or a $\log_{10}$ fold change.

6. The method of any one of claims 3-5, wherein the method further comprises normalizing the discrete attribute value prior for each respective gene in the plurality of genes prior to computing the differential value for each respective gene in the plurality of genes for each respective cluster in the first plurality of clusters or the second plurality of clusters.

7. The method of claim 1, wherein:

the first clustering of discrete attribute values for the plurality of genes across the plurality of cells is a clustering of a first plurality of dimension reduction values for each respective cell in the plurality of cells, across the plurality of cells, wherein each respective dimensional reduction value in the first plurality of dimension reduction values for each respective cell in the plurality of cells is derived, using a first dimension reduction algorithm, from the discrete attribute values of each gene in the respective cell, and

the second clustering of ATAC fragment counts for the plurality of ATAC peaks across the plurality of cells is a clustering of a second plurality of dimension reduction values for each respective cell in the plurality of cells, across the plurality of cells, wherein each respective dimensional reduction value in the second plurality of dimension reduction values for each respective cell in the plurality of cells is derived, using the first dimension reduction algorithm, from the ATAC fragment counts of each ATAC peak in the respective cell.

8. The method of claim 7, wherein the first dimension reduction algorithm is principal component analysis.

9. The method of claim 7 or 8, wherein

the two-dimensional projection of the plurality of cells is based on the assignment of the plurality of cells to the first plurality of cluster groups, and

the two-dimensional projection of the plurality of cells is obtained from t-distributed stochastic neighbor or UMAP embedding of the first plurality of dimension reduction values for each respective cell in the plurality of cells, across the plurality of cells.

10. The method of claim 7 or 8, wherein

the two-dimensional projection of the plurality of cells is based on the assignment of the plurality of cells to the second plurality of cluster groups, and

the two-dimensional projection of the plurality of cells is obtained from t-distributed stochastic neighbor or UMAP embedding of the second plurality of dimension reduction values for each respective cell in the plurality of cells, across the plurality of cells.

11. The method of any one of claims 1-10, wherein

the first clustering of discrete attribute values for the plurality of genes across the plurality of cells comprises application of a Louvain modularity algorithm, k-means clustering, a fuzzy k-means clustering algorithm, or Jarvis-Patrick clustering, and

the second clustering of ATAC fragment count values for the plurality of ATAC peaks across the plurality of cells comprises application of a Louvain modularity algorithm, k-means clustering, a fuzzy k-means clustering algorithm, or Jarvis-Patrick clustering.

12. The method of any one of claims 1-11, wherein the respective discrete attribute value for each corresponding gene in a plurality of genes, for each respective cell in a plurality of cells represents a whole transcriptome shotgun sequencing experiment that quantifies gene expression from each respective single cell in the plurality of cells in counts of transcript reads mapped to the genes.

13. The method of claim 1, wherein each gene in a particular cell in the plurality of cells is uniquely represented within the discrete attribute value dataset with a first barcode that is unique to the particular cell.

14. The method of claim 1, wherein the discrete attribute value of each gene in a particular cell in the plurality of cells is determined after the particular cell has been separated from all the other cells in the plurality of cells into its own microfluidic partition.

15. The method of any one of claims 1-14, wherein the plurality of cells comprises 1000 cells, 2000 cells, 5000 cells, 10,000 cells, 25,000 cells, 50,000 cells or 100,000 cells.

16. The method of any one of claims 1-15, wherein the plurality of genes comprises 150 genes, 200 genes, 300 genes, 400 genes, 1000 genes, 2000 genes, 3000 genes, 4000 genes, or 5000 genes.

17. The method of any one of claims 1-16, wherein

the discrete attribute value dataset further comprises a feature-linkage matrix that stores, for each respective gene in the plurality of genes and for each respective ATAC peak in the plurality of ATAC peaks:

a collection of ATAC peaks and genes that are within a threshold distance of the respective gene or respective ATAC peak in a reference genome, and for each respective ATAC peak or respective gene in the collection:

a correlation in an ATAC fragment count of the respective ATAC peak or a correlation in the discrete attribute value of the respective gene to the first ATAC peak or first gene across the plurality of cells, and

the method further comprises:

receiving a selection of a first gene in the plurality of genes or a first ATAC peak in the plurality of ATAC peaks, and

using the feature-linkage matrix to obtain and provide a first plot comprising an indicator for each gene in the plurality of genes or each peak in the plurality of peaks linked to the first gene or the first ATAC peak, by order of distance apart from the first gene or the first ATAC peak in the reference genome.

18. A computer system comprising one or more processing cores and a memory, the memory storing instructions that use the one or more processing cores to perform a method comprising:

storing a discrete attribute value dataset in the memory, wherein the discrete attribute value dataset comprises:

a respective discrete attribute value for each corresponding gene in a plurality of genes, for each respective cell in a plurality of cells, and

a respective ATAC fragment count for each corresponding ATAC peak in a plurality of ATAC peaks, for each respective cell in the plurality of cells, wherein

the plurality of cells comprises 100 cells,

the plurality of genes comprises 100 genes,

the plurality of ATAC peaks comprises 50 ATAC peaks,

each respective cell in the plurality of cells is assigned to a respective cluster group in a first plurality of cluster groups based on a first clustering of discrete attribute values for the plurality of genes across the plurality of cells, and

each respective cell in the plurality of cells is also assigned to a respective cluster group in a second plurality of cluster groups based on a second clustering of ATAC fragment count values for the plurality of ATAC peaks across the plurality of cells;

displaying, in a first panel, a two-dimensional projection of the plurality of cells based on assignment of the plurality of cells to one of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups; and

indicating within the two-dimensional projection, for each respective cell in the plurality of cells, membership in the other of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups, thereby visualizing the pattern in the discrete attribute value

19. A non-transitory computer readable storage medium, wherein the non-transitory computer readable storage medium stores instructions, which when executed by a computer system, cause the computer system to perform a method for visualizing a pattern in a discrete attribute value dataset, the method comprising:

obtaining the discrete attribute value dataset, wherein the discrete attribute value dataset comprises:

a respective discrete attribute value for each corresponding gene in a plurality of genes, for each respective cell in a plurality of cells, and

a respective ATAC fragment count for each corresponding ATAC peak in a plurality of ATAC peaks, for each respective cell in the plurality of cells, wherein

the plurality of cells comprises 100 cells,

the plurality of genes comprises 100 genes,

the plurality of ATAC peaks comprises 50 ATAC peaks,

each respective cell in the plurality of cells is assigned to a respective cluster group in a first plurality of cluster groups based on a first clustering of discrete attribute values for the plurality of genes across the plurality of cells, and

each respective cell in the plurality of cells is also assigned to a respective cluster group in a second plurality of cluster groups based on a second clustering of ATAC fragment count values for the plurality of ATAC peaks across the plurality of cells;

displaying, in a first panel, a two-dimensional projection of the plurality of cells based on assignment of the plurality of cells to one of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups; and

indicating within the two-dimensional projection, for each respective cell in the plurality of cells, membership in the other of (i) the first plurality of cluster groups or (ii) the second plurality of cluster groups, thereby visualizing the pattern in the discrete attribute value.