WHAT IS CLAIMED IS:

1. A method for identifying targets to distinguish cell types, comprising:

(a) receiving a target counts data structure, wherein the target counts data structure comprises expression profiles of a plurality of cells, and wherein the expression profiles of the plurality of cells comprise a number of each target of a plurality of targets for each cell of the plurality of cells;

(b) hierarchically clustering expression profiles of the plurality of cells based on the target counts data structure and distances between the expression profiles of the plurality of cells to generate a dendrogram representing the expression profiles of the plurality of cells, wherein the dendrogram comprises a plurality of nodes, wherein the plurality of nodes comprises a root node, a plurality of leaf nodes, and a plurality of non-root, non-leaf nodes, wherein each leaf node of the plurality of leaf nodes represents an expression profile of a different cell of the plurality of cells, and wherein the root node represents the expression profiles of the plurality of cells, wherein each of the plurality of leaf nodes and the plurality of non-root, non-leaf nodes is associated with a parent node, and wherein each of the root node and the plurality of non-root, non-leaf nodes is associated with a left child node and a right child node and represents expression profiles represented by the left child node and the right child node of the node;

(c) while traversing through each node of the plurality of nodes of the dendrogram from the root node of the dendrogram to the plurality of leaf nodes of the dendrogram:

(1) determining whether a splitting of the node into child nodes of the node is valid or invalid by:

determining the splitting to be valid if the distance between the left child node and the right child node is above a splitting threshold; and

determining the splitting to be invalid if the distance between the left child node and the right child node is not above the splitting threshold, wherein the distance between the left child node and the right child node is determined based on a statistical test performed on each target of the plurality of targets between expression profiles represented by the left child node and the right child node; and

(2) if the splitting of the node into the child nodes of the node is invalid, adding the node to a merging cluster set;

(d) iteratively, for each first node in the merging cluster set, if a distance between the first node in the merging cluster set and a second node in the merging cluster set that is closest to the first node is within a merging distance threshold, merging the first node with the second node to generate a merged node comprising expression profiles represented by the first node and the second node; and

(e) for each node in the merging cluster set, identifying targets for distinguishing cell types based on expression profiles of the plurality of targets of cells represented by the node.

2. The method of claim 1, wherein the target counts data structure comprises a target counts matrix, and wherein each row or each column of the target counts matrix comprises a number of each target of a plurality of targets for a different individual cell of the plurality of cells.

3. The method of any one of claims 1-2, wherein hierarchically clustering the expression profiles of the plurality of cells based on the target counts data structure and the distances between the expression profiles of the plurality of cells in (b) comprises:

assigning each expression profile of the plurality of cells to a different leaf node; and

iteratively combining a first node and a second node of the plurality of nodes to generate a parent node of the first node and the second node if the second node is the closest node of the plurality of nodes to the first node.

4. The method of claim 3, wherein an indication of an intra-node correlation of the first node and an intra-node correlation of the second node is higher than an inter-node correlation of the first node and the second node.

5. The method of claim 4, wherein the indication of the intra-node correlation of the first node and the intra-node correlation of the second node is based on at least one of:

an intra-node maximum correlation of the first node and the second node,

an intra-node average correlation of the first node and the second node,

an intra-node median correlation of the first node and the second node,

an intra-node minimum correlation of the first node and the second node, and

any combinations thereof,

and wherein the intra-node correlation of the first node is based on at least one of:

an intra-node maximum correlation of the first node,

an intra-node average correlation of the first node,

an intra-node median correlation of the first node,

an intra-node minimum correlation of the first node, and

any combinations thereof,

and wherein the intra-node correlation of the second node is based on at least one of:

an intra-node maximum correlation of the second node,

an intra-node average correlation of the second node,

an intra-node median correlation of the second node,

an intra-node minimum correlation of the second node, and

any combinations thereof, and

wherein the inter-node correlation of the first node and the second node is based on at least one of:

an inter-node maximum correlation of the first node and the second node,

an inter-node average correlation of the first node and the second node,

an inter-node median correlation of the first node and the second node,

an inter-node minimum correlation of the first node and the second node, and

any combinations thereof.

6. A method for identifying targets to distinguish cell types, comprising:

(a) receiving expression profiles of a plurality of cells, wherein the expression profiles comprise a number of each target of a plurality of targets for each cell of the plurality of cells;

(b) clustering the expression profiles of the plurality of cells to generate a plurality of clusters of expression profiles based on the distances between the expression profiles of the plurality of cells, wherein each cluster has one or more associations with one or both of (1) a parent cluster and (2) two or more child clusters, wherein the parent cluster represents expression profiles of one or more cells of the plurality of cells represented by the cluster, and wherein the cluster represents expression profiles represented by the two or more child clusters;

(c) for each cluster with two or more child clusters, if associations between the cluster with the two or more child clusters are invalid, adding the cluster to a merging cluster set,

wherein the associations between the cluster with the two or more child clusters is determined to be invalid when the distance between a left child cluster and a right child cluster is not above an association threshold,

wherein the associations between the cluster with the two or more child clusters is determined to be valid when the distance between the left child cluster and the right child cluster is above the association threshold,

wherein the distance between the left child cluster and the right child cluster is determined based on a statistical test performed on each target of the plurality of targets between expression profiles represented by the left child cluster and the right child cluster;

(d) iteratively, for each first cluster in the merging cluster set, if a distance between the first cluster in the merging cluster set and a second cluster in the merging cluster set that is closest to the first cluster is within a merging distance threshold, merging the first cluster and the second cluster to generate a merged cluster, wherein the merged cluster comprises expression profiles of the first cluster and the second cluster; and

(e) for each cluster in the merging cluster set, identifying targets for distinguishing cell types based on expression profiles of the plurality of targets of cells represented by the cluster.

7. The method of claim 6, wherein clustering the expression profiles of the plurality of cells into the plurality of clusters of expression profiles based on the distances between the expression profiles of the plurality of cells comprises:

hierarchically clustering the expression profiles of the plurality of cells to generate a dendrogram representing the expression profiles of the plurality of cells based on the distances between the expression profiles of the plurality of cells, wherein the dendrogram

comprises the plurality of clusters, wherein the plurality of clusters comprise a root cluster, a plurality of leaf clusters, and a plurality of non-root, non-leaf clusters.

8. The method of claim 7, wherein for each cluster with two or more child clusters, if associations between the cluster with the two or more child clusters are invalid, adding the cluster to a merging cluster set comprises:

while traversing through each cluster of the dendrogram from the root cluster of the dendrogram to the plurality of leaf clusters of the dendrogram:

(1) determining whether associations of the cluster with child clusters of the cluster are valid or invalid; and

(2) if the associations are invalid, adding the cluster to a merging cluster set.

9. The method of claim 8, comprising, at each cluster when traversing the plurality of clusters of the dendrogram: (3) adding the cluster to the merging cluster set if the cluster represents an expression profile of a single cell.

10. The method of claim 8 or 9, comprising, at each cluster when traversing the plurality of clusters of the dendrogram: assigning a cluster label to the cluster.

11. The method of claim 10,

wherein, if the cluster represents an expression profile of a single cell, the cluster label of the cluster comprises a single cell designation,

otherwise, if the cluster is the left child cluster of the parent cluster, the cluster label of the cluster comprises the cluster label of the parent cluster and a left designation, and

otherwise, the cluster label of the cluster comprises the cluster label of the parent cluster and a right designation.

12. The method of any one of claims 6-11, comprising, prior to clustering the expression profiles of the plurality of cells to generate the plurality of clusters of expression profiles based on the distances between the expression profiles of the plurality of cells in (b):

(i) determining a distance data structure of the expression profiles of the plurality of cells.

13. The method of claim 12, wherein the distance data structure comprises a distance matrix of the expression profiles of the plurality of cells.

14. The method of claim 13, wherein clustering the expression profiles of the plurality of cells to generate the plurality of clusters of expression profiles based on the distances between the expression profiles of the plurality of cells in (b) comprises: clustering the expression profiles of the plurality of cells to generate the plurality of clusters of expression profiles based on the distance matrix.

15. The method of any one of claims 12-14, comprising, prior to determining the distance data structure in (i), log-transforming the target counts data structure into a log-transformed target counts data structure,

wherein determining the distance data structure of elements of the target counts data structure comprises determining the distance data structure of the log-transformed target counts data structure, and

wherein clustering the expression profiles of the plurality of cells to generate the plurality of clusters of expression profiles based on the distances between the expression profiles of the plurality of cells in (b) comprises: clustering the expression profiles of the

plurality of cells based on the log-transformed target counts data structure and the distance data structure to generate the plurality of clusters.

16. The method of any one of claims 6-15, wherein clustering the expression profiles of the plurality of cells based on distances between the expression profiles of the plurality of cells in (b) comprises:

assigning each expression profile of the plurality of cells to a different leaf cluster in the plurality of clusters; and

iteratively combining a first cluster and a second cluster of the plurality of clusters to generate a parent cluster of the first cluster and the second cluster if the second cluster is the closest cluster of the plurality of cluster to the first cluster.

17. The method of any one of claims 6-16, wherein the distance between the left child cluster and the right child cluster is determined based on the maximum *p*-value of the statistical test performed on the each target of the plurality of targets between expression profile represented by the left child cluster and each expression profile represented by the right child cluster.

18. The method of any one of claims 6-17, wherein for each cluster in the merging cluster set, identifying the targets for distinguishing the cell types based on the expression profiles of the plurality of targets of the cells represented by the cluster comprises:

determining that the number of molecular labels with distinct sequences associated with the targets for distinguishing the cell types in the expression profile represented by the cluster compared to the number of molecular labels with distinct sequences associated with

the targets for distinguishing the cell types in the expression profile represented by another cluster in the merging cluster set is greater by at least a significance threshold.