(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 June 2018 (21.06.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/111982 A1

(51) International Patent Classification:
G01N 15/14 (2006.01) *C12N 15/10* (2006.01)
G01N 33/53 (2006.01) *C12N 5/071* (2010.01)
G01N 33/48 (2006.01) *C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2017/065987

(22) International Filing Date:
13 December 2017 (13.12.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/433,930 14 December 2016 (14.12.2016) US

(71) Applicant: **FLOWJO, LLC** [US/US]; 385 Williamson Way, Ashland, Oregon 97520 (US).

(72) Inventors: **ALMARODE, James**; 31011 SW Larson Road, Lebanon, Oregon 97355 (US). **SPIDLEN, Josef**; #20-2561 Runnel Drive, Coquitlam, British Columbia V3E2S3 (CA). **STADNISKY, Michael David**; 755 South Mountain Avenue, Ashland, Oregon 97520 (US).

(74) Agent: **DILEO, Christopher M.** et al.; Knobbe Martens Olson & Bear, LLP, 2040 Main Street, Fourteenth Floor, Irvine, CA 92614 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: APPLIED COMPUTER TECHNOLOGY FOR MANAGEMENT, SYNTHESIS, VISUALIZATION, AND EXPLORATION OF PARAMETERS IN LARGE MULTI-PARAMETER DATA SETS

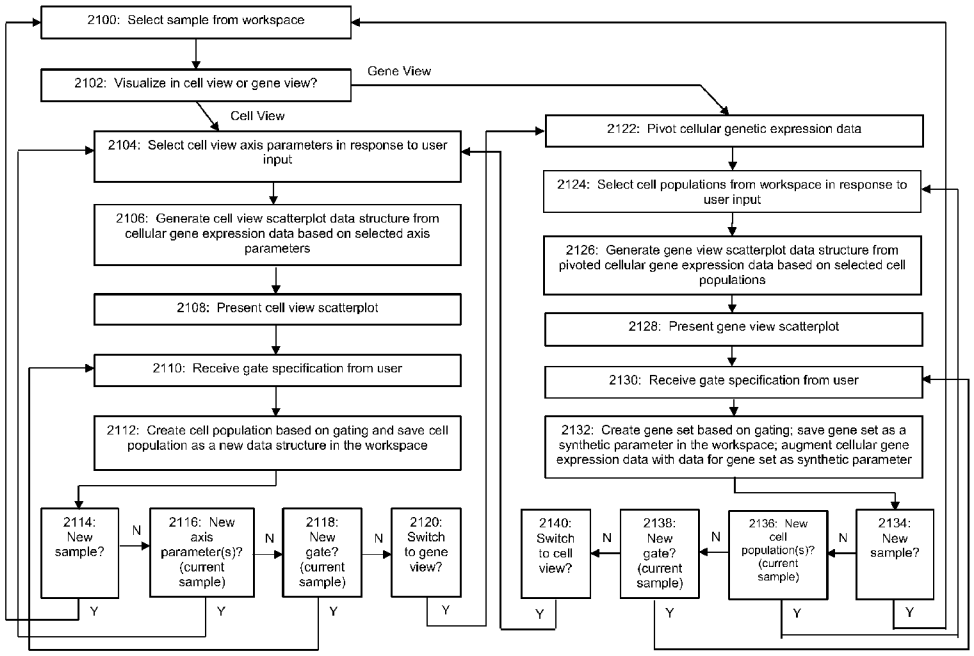


Figure 21

(57) Abstract: Computer technology is disclosed that applies innovative data processing and visualization techniques to large multi-parameter data sets such as cellular gene expression data to find new relationships such as relationships between cells and genes and create new associative data structures within the data sets that represent these relationships. For example, scatterplots of gene expression data can be iteratively pivoted between a cell view and a gene view to find cell populations and gene sets of interest to a user.

[Continued on next page]

WO 2018/111982 A1

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

Applied Computer Technology for Management, Synthesis, Visualization, and Exploration of Parameters in Large Multi-Parameter Data Sets

Cross-Reference and Priority Claim to Related Patent Application:

This patent application claims priority to U.S. provisional patent application serial no. 62/433,930, filed December 14, 2016, and entitled "Applied Computer Technology for Management, Synthesis, Visualization, and Exploration of Parameters in Large Multi-Parameter Data Sets", the entire disclosure of which is incorporated herein by reference.

Introduction:

The volume of genetic and gene expression information that is available for both bulk populations and individual cells has grown to the point where it has become unwieldy for investigators. For example, cellular gene expression data can include gene expression data for thousands of genes (e.g., 10,000-30,000 or more genes), can now be measured for individual cells, and thousand cells can be measured per sample. This has presented a tremendous technical problem in the art of visualizing, analyzing, exploring, and making sense of cellular gene expression data.

For example, with conventional approaches to using computers to facilitate visualization of cellular gene expression data, the visualization is a final end point and the visualization is reached as a result of a user manually writing scripts using the R programming language, which requires the user to have knowledge of different libraries in order to perform data input, reformatting, manipulations, calculations and graphing. These scripts usually have to be customized for particular data sets, and their creation requires expert knowledge of the programming language, existing libraries, and required inputs to produce results. Moreover, such conventional approaches prevent the deep exploration of heterogeneous cell populations.

As a solution to this technical problem, the inventors disclose the application of computer technology using innovative scatterplot displays across various dimensions of cellular expression data including cell (or cell population) view scatterplots in which cells are visualized as individual data points (e.g., gene vs gene scatterplots of cells) and gene view scatterplots in which genes are visualized as individual data points (e.g., cell population vs

cell population scatterplots of genes). Gating can be performed within these scatterplots to create cell populations and gene sets respectively that can serve as biologically-relevant dimensions to be added as new data objects in a workspace for use in augmenting the cellular gene expression data and opening new avenues for meaningful investigation. By way of contrast, performing such analysis in an isolated, siloed manner on the basis of individual genes quickly becomes unwieldy, whereas the ability to pivot between cell view scatterplots and gene view scatterplots allows users to find biologically-relevant grouping of genes that can then be further investigated as synthetic parameters of a cell view scatterplot.

As noted above, with conventional visualization systems in the art, the visualization serves as an endpoint in the process and cannot serve as a starting point for further creating further visualization refinements for further investigations. As an example, samples of immune cells from metastatic melanoma patients may contain T cells, and conventional visualization systems in the art would be able only to identify this subset within the immune cells. However, the innovative computer systems described herein allow for deep exploration and analysis of the T cell subset to identify multiple subsets within T cells, for example, T cells which are "exhausted", track this state to individual genes which can then be targeted to reverse this exhaustion, activate T cells, and thus possibly stimulate an immune response to eradicate the metastases, as explained in greater detail below with reference to example embodiments.

Accordingly, through the innovative visualization techniques described herein, computer technology can be applied to cellular gene expression data to find new relationships between cells and genes and create new associative data structures within the cellular gene expression data that represents these relationships.

Through these and other features, example embodiments of the invention provide significant technical advances in the applied bioinformatics arts.

Brief Description of the Drawings:

Figure 1 discloses an example computer system that can be used to support the innovative data processing and visualization techniques described herein.

Figure 2A depicts examples of cellular gene expression data sets.

Figure 2B depicts a table view of example cellular gene expression data.

Figure 3 depicts an example cell view graph window scatterplot.

Figure 4 depicts an example process flow for execution to create a cell view scatterplot.

Figure 5 shows an example cell view scatterplot user interface that includes parameter selection menus.

Figure 6 shows an example of how gating can be performed within a cell view scatterplot user interface to create a cell population.

Figure 7 shows an example user interface that permits a user to spread the zeros of a cell view scatterplot.

Figure 8 shows an example cell view scatterplot where parameters other than genes are selected as axis parameters.

Figure 9 shows an example of how gating can be performed within the cell view scatterplot of Figure 8 to create a cell population.

Figure 10 shows an example user interface for adjusting display settings in a scatterplot presentation.

Figure 11 depicts an example gene view graph window scatterplot including cell population selection menus.

Figure 12 depicts an example of how cellular gene expression data can be pivoted to create cell population data for a gene view scatterplot.

Figures 13A and B depict examples of how complementary cell populations can be created and defined in a workspace for use in gene view scatterplots.

Figure 14 shows an example of how gating can be performed within a gene view scatterplot user interface to create a gene set.

Figure 15 shows an example of how cellular gene expression data can be augmented with gene sets as synthetic parameters.

Figure 16 shows an example of a cell view scatterplot where gene sets are presented as options in the parameter selection menus.

Figures 17A and B show an example user interface for viewing, editing, and creating new gene sets from other gene sets in a workspace.

Figure 18A-D show examples of how a user-selected third dimension can be overlaid on a scatterplot.

Figure 19 shows an example of how gating can be performed within the 3D scatterplots of Figures 18A-D.

Figures 20A-D show example reports can be created through the system.

Figure 21 shows an example process flow for execution by the system to switch between cell view and gene view modes.

Figure 22 shows an example of how the system can be operated to switch between cell view and gene view modes to support investigations and research into cell data.

Detailed Description of Example Embodiments:

Figure 1 discloses an example computer system 100 that can be used to support the innovative data processing and visualization techniques described herein. The example computer system 100 comprises a processor 102, memory 104, database 106, and display 108 that can be in communication with each other over an interconnect technology such as bus 110.

Processor 102 can take the form of any processor suitable for performing the operations described herein. For example, the CPU of a laptop or workstation would be suitable for use as processor 102. It should be understood that processor 102 may comprise multiple processors, including distributed processors that communicate with each other over a network to carry out tasks described herein (e.g., cloud computing processing resources). Memory 104 can take the form of any computer memory suitable for cooperating with processor 102 in the execution of the tasks described herein. It should be understood that memory 104 may take the form of multiple memory devices, including memory that is distributed across a network. Similarly, database 106 can take the form of any data repository accessible to the processor 102 (e.g., a file system on a computer, relational database, etc.), and it should be understood that database 106 may take the form of multiple distributed databases (e.g., cloud storage). Display 108 can take the form of a computer monitor or screen that is capable of generating the visualizations described herein.

The innovative data analysis and visualization techniques described herein can be performed on cellular gene expression data 112. Cellular gene expression data 112 can be generated by next-generation sequencing (e.g. for the measurement of RNA-Sequencing (RNASeq) and single cell RNA sequencing (scRNA-Seq) among other sequencing approaches). However, this is only an example, and other techniques for generating cellular gene expression data 112 may be employed. Additional examples include polymerase chain reaction approaches including digital droplet and reverse transcriptase. Still more examples include RNA measurement by flow cytometry, and microarrays, among others, that produce data files which contain the quantification of DNA and/or RNA, or through software programs that process the raw read data (primary and secondary analysis) to generate the gene expression data files. Yet another example is gene expression data derived from stochastic labeling of a sample. Examples of stochastically labeled gene expression data can

be found in U.S. Patent Nos. 9,567,645 and 8,835,358 and in patent application 15/715,028, filed September 25, 2017, and entitled "Measurement of Protein Expression Using Reagents with Barcoded Oligonucleotide Sequences", the entire disclosures of each of which are incorporated by reference.

Furthermore, the innovative data analysis and visualization techniques described herein can be applied to data generated from single cells by a variety of means. Single cell analysis may include the stochastic labeling of nucleic acids or proteins or any combination of proteins and nucleic acids. As an example, the innovative data analysis and visualization techniques described herein may be used to analyze quantitative features of protein density or gene expression or any combination thereof. The innovative data analysis and visualization techniques described herein can also provide for an improved visualization of quantitative data generated by stochastic labeling of a variety of proteins or nucleic acids in single cells. Quantitative values of biological populations (nucleic acids, proteins etc.) in individual cells may be compared with other individual cells, or between cell types or even between methods of generating the data. For example gene expression values may be visualized as a function of the method used of generating the data set. The innovative data analysis and visualization techniques described herein can also be used to compare quantitative data generated by a variety of means described here in ways that provide for the visualization of quantitative biological population data independent of the method of generating such data.

This data 112 can be characterized as a large multi-parameter data set which poses special technical challenges in terms of difficulty in creating meaningful visualizations, particularly when considered with respect to underlying biology so that biologically-relevant information is meaningfully presented to users in a visual manner. For example, the cellular gene expression data may comprise data for large numbers of individual cells and cell populations, with parameters for each cell or cell population that may stretch into 10,000-30,000 or more parameters. Cellular gene expression data 112 can be read out of files in database 106 and loaded into memory 104 as a plurality of data structures 116 to be manipulated by processor 102 during execution of an analysis and visualization program 114. Program 114 may comprise processor-executable computer code in the form of a plurality of processor-executable instructions that are resident on a non-transitory computer-readable storage medium such as memory 104.

Figure 2A depicts examples of cellular gene expression data sets where each cell (or cell population) is identified by a Cell ID and is associated with a plurality of parameters, each parameter having an ID and a value in relation to a Cell ID. As indicated, gene

expression data for cells is highly dimensional and the number of parameters for each cell may reach 10,000-30,000 or more parameters, either per cell or per population of cells. Examples of parameters in the cellular data include counts of gene expression in the subject cell for a large number of genes. Thus, Parameter 1 for Cell 1 may correspond to Gene 1 and its value can be a count of expressions for Gene 1 in Cell 1. Similarly, Parameter 2 for Cell 1 may correspond to Gene 2 and its value can be a count of expressions for Gene 2 in Cell 1. Figure 2B depicts a table view of example cellular gene expression data 112. Each row in the table 200 corresponds to a different cell (see the Cell column), and the various columns labeled Gene 1, Gene 2, etc. correspond to different genes and the table cells identify counts of gene expressions for the correspond genes in each subject cell. This table may also include parameters other than genes. For example, the cellular gene expression data 112 may include data values for parameters such as t-distributed stochastic neighbor embedding (tSNE), principal component analysis (PCA), linear discriminant analysis (LDA), etc. in each table cell, where these data values represent an analysis calculation whose value captures differences for individual cells across n parameters. The cellular gene expression data 112 can be stored in any of a number of formats (e.g., as CSV files, database tables (e.g., as relational data in a relational database), spare data representations, binary formats, and others).

Figure 3 depicts an example graph window (GW) that can be produced through execution of program 114 for presentation to a user via display 108, where the graph window visualizes Gene 1 vs. Gene 2 with respect to a cell population. This visualization can be referred to as a cell view of the cellular gene expression data 112. As explained below, gating on a selection of individual cells in the cell view creates a cell population. This graph window presents a scatterplot 300 of gene expressions for two user-selected genes 302 and 304 (TMEM216 and MMP2, respectively in this example) with respect to a cell population in a user-selected file 306. Each file may correspond to a single sample of cell populations, a single cell population (e.g. one population that has been sorted by flow cytometric active cell sorting and analyzed), or possibly multiple samples which have been concatenated (from different patients, or from a patient at different points in time). Each dot 308 in the scatterplot represents a cell in the cell population of the subject file 306. The scale of the X-axis and Y-axis in this example identify counts for the corresponding gene. Thus, the position of a dot 308 on the horizontal X-axis identifies a count of how many of the gene MMP2 are present in the cell corresponding to the subject dot 308, and the position of the dot 308 on the vertical Y-axis identifies a count of how many of the gene TMEM216 are present

in the cell corresponding to the subject dot 308. Thus, dots 308 in the upper right quadrant of the scatterplot 300 correspond to cells where both TMEM216 and MMP2 are highly expressed, while dots in the lower left quadrant of the scatterplot 300 correspond to cells where both TMEM216 and MMP2 are expressed at a low level. Likewise, the upper left quadrant corresponds to cells where TMEM216 is highly expressed but MMP2 is not, while the lower right quadrant corresponds to cells where MMP2 is highly expressed, but TMEM216 is not. The diagonal where y=x is where cells 308 are positioned if those cells equally express both selected genes. Thus, the distance of a cell 308 away from this diagonal in either direction indicates the extent of differential expression of the selected genes in a given cell. It is expected that for many cell populations, there will be large numbers of cells where the expression of the selected genes for those cells is at the zero level. This leads to a large clustering of dots 308 at the zero levels 310 and 312 on the X-axis and Y-axis respectively. Color coding can be used at 310 and 312 to indicate the density of cells with such zero-level gene expressions.

Figure 4 depicts an example process flow for execution as part of program 114 that describes how scatterplot 300 can be generated. At step 400, the processor creates a data structure in a memory workspace (see 116 in Figure 1). This data structure can be used for holding cell view data.

For a first axis, the processor selects a gene in the cellular gene expression data 112 based on user input (step 402). For example, the processor can respond to user input to select a gene column in table 200. Figure 5 shows how the user interface can present a user with a list of genes for each axis. To access a selection menu, a user can select a gene selector 302 or 304 for the Y and X axes respectively. If we assume the first axis in this example is the Y-axis, upon selection of 302, a parameter selection menu 500 is shown which presents a list of parameters available for selection with respect to the Y-axis. This list can be populated with parameters from the cellular gene expression data 112. As shown in Figure 5, the user selection 502 for the Y-axis gene is TMEM216. At step 404, the processor populates the data structure with a list of cells from the cellular gene expression data 112 (e.g., the cells in the cell column of table 200). Each listed cell is associated with its count value for the selected first axis gene.

For a second axis, the processor selects another gene in the cellular gene expression data 112 based on user input (step 406). For example, the processor can respond to user input to select another gene column in table 200. Figure 5 shows an example of a user interface where 304 is selected to access an X-axis parameter selection menu 506 (resulting in

selection 508 of MMP2). At step 408, the processor augments the cell list to add to each cell its associated count value for the selected second axis gene. Thus, at this point, the data structure comprises a list of cells associated with count pairs for the selected genes of the first and second axis; for example the list can comprise a set of vectors {Cell ID 1, Gene 1 Count, Gene 2 Count} for each cell in the cellular gene expression data 112.

At step 410, the process parses the list to find the maximum values for each selected gene (the highest counts). These maximum values are then used by the processor to define the appropriate scales for the X-axis and Y-axis in the scatterplot (step 412). For example, if the maximum value for the X-axis gene is 10, the X-axis scale can be from 0-10. At step 414, the processor draws the scatterplot based on the cell list and the defined scales using the cell's associated count values as X,Y coordinates in the scatterplot. The result is a scatterplot 300 as shown by Figure 3.

Returning to Figure 3, a user can access gate creation tools 320 in the user interface to create a gate through which child cell populations are created. For example, the user can access a tool 320 to draw a shape in the scatterplot 300 that encompasses a subset of the cells 308. Figure 6 depicts an example gate 600 drawn to capture the cells that have non-zero expressions of the two selected genes. The cells 308 falling inside the drawn shape 600 are gated into their own child cell population, and this cell population can be added to the workspace as a distinct object 602. To create the cell population, the processor can translate gate 600 into a plurality of boundary conditions for the gated cell population. For example, with respect to a rectangle shape as shown by gate 600, the boundary conditions can be all cells with (1) X-axis values between 1 and 10, and (2) Y-axis values between 1 and 8. The cell list data structure (e.g., the set of vectors {Cell ID 1, Gene 1 Count, Gene 2 Count}) can be traversed to find all Cell IDs meeting these criteria, and the data for these cell IDs can populate a new child cell population data structure in the workspace. Also, a corresponding scatterplot 300 of gene expressions for the child cell population can then be presented (where the cell population includes one or more cells 308). This gating can allow a user to focus on a biologically-interesting cluster of cells 308 in the scatterplot 300.

Also, the cell view mode of Figure 3 permits users to visually compare different samples, including child and parent cell populations using navigation tools 322. For example, through the back and next buttons in tools 322, a user can navigate to a scatterplot 300 for a next and previous sample in the workspace. Through the down button in tools 322, a user can navigate to a child cell population in an analysis hierarchy (e.g., to the child cell

population created via gate 600 of Figure 6). Also, through an up button (not shown), a user can navigate to a parent cell population in the analysis hierarchy.

Furthermore, when in cell view mode, it is expected that for many cell populations there will be large numbers of cells where the expression of the selected genes for those cells is at the zero level. This leads to a large clustering of dots 308 at the zero levels 310 and 312 on the X-axis and Y-axis respectively of scatterplot 300 shown by Figure 3. Color coding can be used at 310 and 312 to indicate the density of cells with such zero-level gene expressions. However, the inventors believe that the display can be enhanced to provide biologically-relevant information to users in certain circumstances by expanding the visualization of the zero levels for the two selected genes (or gene sets as explained below). To accomplish this, the user interface can include a "spread zeros" user control 700 (such as a check box, button, etc.) as shown in Figure 7. This control 700 can be provided on a menu for selecting the axis parameters, although if desired by a practitioner, control 700 could be positioned elsewhere such as somewhere on the user interface shown by Figure 3. The top half of Figure 7 shows the scatterplot when the zeros are not spread. The bottom half of Figure 7 shows the scatterplot when the spread zeros option is selected via input mechanism 700.

As shown by the bottom scatterplot in Figure 7, box 702 provides an expanded view of the cells that exhibit zero values with respect to the expression of the X-axis gene (MMP2 in this example). Box 704 provides an expanded view of the cells that exhibit zero values with respect to the expression of the Y-axis gene (TMEM216 in this example). The depth of the zero space along the X and Y axes can be defined as a function of how many dots/cells are at the zero levels for each gene, and the dots/cells can be spread across the zero-space defined by boxes 702 and 704 as a function of the density of cells at each zero location, although other distribution techniques could be used. As should be understood, the lower left quadrant of the bottom scatterplot of Figure 7 (defined by the overlap of boxes 702 and 704) comprises the cells that have zero expressions for both the X-axis and Y-axis genes. The lower right quadrant of this scatterplot comprises cells that have zero expressions of the Y-axis gene but positive expressions of the X-axis gene, and the upper left quadrant of this scatterplot comprises cells that have zero expressions of the X-axis gene but positive expressions of the Y-axis gene. The upper right quadrant of this scatterplot comprises the cells with positive expressions of both genes (effectively, the scatterplot shown in the top part of Figure 7). As can be seen from the bottom scatterplot of Figure 7, the cell population of the upper right quadrant is much sparser than the other quadrants. The inventors believe that

the ability to visualize the zero levels in this spread manner can provide users with biologically relevant information (such as an ability to assess gene combinations where unexpected distributions occur – e.g., where the upper right quadrant is not as sparse as would be generally expected).

The cell view scatterplot can also display cell information for selected parameters in the cellular gene expression data 112 other than genes. As indicated above, the cellular gene expression data 112 may include parameters from dimensionality reduction such as those resulting from tSNE, LDA, PCA, etc. and quality control parameters (above threshold, parameters relating to ribosomal RNA (rRNA) abundance, etc.) These parameters can be presented as options on the parameter selection menus 500 and 506. Figure 8 shows an example where selection 800 for the Y-axis is the parameter tSNE axis 2 and where selection 02 for the X-axis is the parameter tSNE axis 1. The resultant scatterplot of Figure 8 can be created via the Figure 4 process flow. A user can also gate desired cell populations in the scatterplot of Figure 8 (see gate 900 in Figure 9) as explained above with respect to Figure 6.

The cell view user interface of Figure 3 can also include a user control 330 (e.g., the "T" button of Figure 3) through which the user can alter the display of information in scatterplot 300. Upon selection of T button 330, the user interface of Figure 10 can be presented. Through this interface, a user can change the binning/display of scatterplot data interactively and live. The list at right in Figure 10 shows the available and selected parameters for the X-axis of the scatterplot. A histogram 1000 shows the binning of values for the selected parameter with respect to the subject cell population. Given the high prevalence of zeros, histogram 1000 shows a large spike at the zero level, and the scale of this example obscures the non-zero levels in the histogram. Through left/right arrows shown toward the top of Figure 10, a user can quickly preview displays for different samples. Through +/- controls in the middle of Figure 10, a user can easily alter the histogram zoom level.

Through scale controls 1002, the user can adjust the X-axis of the scatterplot in a variety of ways. For example, the X-axis scale can be defined to exhibit a liner scale or some other scale (such as a log2 scale) (see control 1004). Also, through min/max controls 1006, the user can define the minimum and maximum boundaries on the X-axis. For example, through these controls, the minimum value can be defined to be a value greater than zero, which would remove the zero spike from the histogram and re-present a newly scaled histogram where the distribution of non-zero values across the X-axis can be more clearly seen. Sliders 1008 can provide users with easy control over the transformation variables.

Also, it should be understood that additional transformation options can be provided, including user-supplied transforms whose variables can be adjusted by user input (see, e.g., US Pat App Pub 2016/0328249 entitled "Plugin Interface and Framework for Integrating External Algorithms with Sample Data Analysis Software", the entire disclosure of which is incorporated herein by reference).

A particularly innovative and powerful aspect of the inventive system disclosed herein is the ability to pivot the scatterplot display from a cell view mode to a gene view mode. Figure 11 shows an example of a scatterplot 1100 in the gene view mode. As used herein, a "gene view" visualization refers to a visualization where genes are the individual data points measured against the axis dimensions (e.g., genes as the dots in a scatterplot measured against two cell populations). With scatterplot 1100, the axis parameters are cell populations (see the X-axis parameter 1102 which identifies a population of B cells, and the Y-axis parameter 1104 which identifies a population of "not B" cells in this example). The dots 1106 in scatterplot 1106 represent specific genes (rather than individual cells as in scatterplot 300 of Figure 3). Population selection menu 1110 provides a list of cell populations available for selection to be used on the X-axis, and population selection menu 1112 provides a list of cell populations available for selection to be used on the Y-axis. As noted, in this example, selections 1114 and 1116 correspond to populations of B cells and "not B" cells respectively.

With this pivot, with reference to cellular gene expression data 112 such as table 200 from Figure 2, the table 200 can be pivoted such the genes become rows in the table and subsets of the cells are grouped into two cell populations that become the columns of the tables. Furthermore, computations can be performed on the gene counts in table 2 for the cells of the two cell populations to determine the values that will populate the cells of the pivoted table. Through control 1120 of Figure 11, a user can define the computations that are to be performed on the gene data to compute the values for the cells in the pivoted table. In the example of Figure 11, a normalized mean computation has been selected. However, it should be understood that other computational options can be available, such as normalized medians, normalized modes, straight means, straight medians, straight modes, etc. Accordingly, it should be understood that the computation performed on the pivoted table data can be any user-defined function that is deemed biologically-relevant to a user. The inventors note that it may be desirable for the computation to be based on averaging of some sort to account for potential discrepancies in the cell counts of the two cell populations and/or normalization to account for the variability between samples (e.g. normalization to a spike-in

of External RNA Controls Consortium (ERCC) controls which contain a known amount of RNA to be measured). For example, if the pivoted table values were straight counts of each gene in the two cell populations, and if there was a meaningful difference in the count of cells in the two cell populations, the aggregated gene count totals across the two cell populations would not be very informative in a comparative sense. However, if the cell populations were roughly similar in size, straight gene counts for each gene in the two cell populations might nevertheless be informative.

Figure 12 depicts an example pivot from a cell view to a gene view. Table 200 in Figure 12 shows gene expressions per cell for a number of cells in a sample. Each row corresponds to a different cell, and the columns correspond to expression counts for different genes in the associated cells. If subsets of these cells are grouped into two cell populations 1210 and 1212 as shown in Figure 12 (e.g., via gating in the cell view scatterplot 300), these two cell populations can be pivoted as shown to create the pivoted table 1200. In pivoted table 1200, the rows are the genes that were columns in table 200. The columns in table 1200 are the two cell populations 1210 and 1212. Each table cell is populated with a concatenation of the gene counts for the associated gene in the cells of each cell population. In the example of Figure 12, the computation being performed on these values is a straight averaging to compute means as shown in pivoted table 1200a. Once again, as noted above, it should be understood that other computations could be performed if desired by a practitioner. A gene view scatterplot of the type shown by scatterplot 1100 in Figure 11 can then be created from pivoted table 1200a using the same basic techniques described in Figure 4 (albeit using pivoted table 1200a rather than table 200).

Returning to Figure 11, the gene view scatterplot 1100 thus provides users with a powerful manner of visualizing differential gene expression between two cell populations over a list of genes that may be extremely large in size. The inventors believe that scatterplot 1100 represents a pioneering new way of visualizing cellular gene expression data in a manner that opens up a wide new array of investigatory options for practitioners, some examples of which are described below. The example gene view scatterplot 1100 of Figure 11 provides users with flexibility in specifying the x/y axis populations and dynamically updating gene expression graphs as populations change. The example gene view scatterplot 1100 also provides users with a plethora of graph options and resolutions. Importantly, the analysis does not stop with this graph; it continues with the ability to create new gene sets and dig deeper into the data as discussed below. Examining cellular heterogeneity and

differences between samples/patients revealed by particularly new single cell methods is not possible without this approach.

The diagonal where y=x is where genes 1106 are positioned if those genes are equally expressed in both cell populations according to the metric defined via 1120. Thus, the distance of a gene 1106 away from this diagonal in either direction indicates the extent of differential expression of the subject gene 1106 as between two selected cell populations. Given that it is expected that most genes 1106 will not be expressed (or will be only lightly expressed) in many cell populations, it is expected that there will typically be large cluster of genes 1106 in the lower left quadrant of scatterplot 1100.

Furthermore, since users can create multiple, hierarchically-related cell populations that exist as data objects in the workspace as described in connection with Figure 6, the user can select any of these cell populations while in the gene view mode. Furthermore, it should be understood that the workspace can be used to create Boolean expressions of cell populations (e.g., the OR combination of all cell populations that are "Not B Cells" to create complementary cell populations) for exploration in the gene view mode. Figures 13A and B show examples of such complementary Boolean cell populations in a workspace. Accordingly, a user can create a scatterplot 1100 that shows differential gene expression as between a population of B cells and a population of "not B" cells.

Another powerful and innovative aspect of the gene view mode described herein is an ability for users to gate genes while in the gene view mode to thereby create gene sets. Figure 14 shows an example where gate 1400 is created in the gene view scatterplot to capture a set of genes that the user deems worthy of further investigation. Gate 1400 can be drawn on the scatterplot using the techniques described above in connection with Figure 6. This gating will create a data structure in the workspace corresponding to the gene set defined by gate 1400 (where the gene set comprises one or more genes depending on how many genes 1106 are encompassed by the user-defined gate 1400). Such a gene set can then serve as a synthetic parameter in relation to the cellular gene expression data 112 that can be selected for visualization while in the cell view mode. In the example of Figure 14, gate 1400 is a trapezoidal shape that captures the genes 1106 that have a positive differential expression in B cells versus non-B cells. The extent of the positive differential expression is controlled by the user through the distance of the trapezoid's diagonal line away from the y=x diagonal of the scatterplot.

Figure 15 shows an example of how the cellular gene expression data 112 can be augmented with gene sets as a synthetic parameter (in the form of example augmented table

200). This table 200 has been augmented with two gene sets as parameters (see the Gene Set 1 and Gene Set 2 columns). In this example, Gene Set 1 consists of {Gene 2, Gene 3} and Gene Set 2 consists of {Gene 1, Gene 4}. The table cells for these gene sets (each corresponding to a different cell in the table rows) can then be populated with the sums of the gene counts for the constituent genes of the subject gene sets. Accordingly, a user can return to the cell view mode to create a cell view scatterplot where one or both of the X-axis and Y-axis parameters are gene sets. An example of this is shown by Figure 16. The parameter selection menus for each axis in Figure 16 include a section that lists available gene set options. The cell view scatterplot of Figure 16 thus shows a scatterplot of cells 308 where the X-axis parameter is the gene set labeled "B_GeneTable" and the Y-axis parameter is the gene set labeled "HighTHighB". Thus, it should be understood that the gating capabilities within the gene view scatterplot to create gene sets in combination with the ability use the created gene sets as a synthetic parameter used in a cell view scatterplot provide users with unprecedented capabilities for intelligently reducing the multi-parameter data space of cellular gene expression data. This combination allows the user to overcome the extremely difficult problem of identifying common genes in populations and identifying the differential genes in populations. In other words, whether the user is looking for what genes are in common, or which genes are different, the ability to create new gene sets and viewing/analyzing them as a single parameter allows the user to focus on the important relationships between cell populations and gene sets. Furthermore, as users create gene sets of biological interest based on various comparisons of cell populations, these gene sets can be shared with other users by passing these gene sets (as a defined list of encompassed genes) to other users so that those other uses can evaluate the gene sets with their samples of cell data. Based on such sharing and independent investigation across different cellular data sets, it is expected that greater insights into gene behavior with respect to cells can be gained.

Figure 17A shows an example view into a workspace that breaks the workspace down into various gene sets 1700, samples 1702, and cell populations 1704 that exist within the workspace. As shown, the section 1700 that lists gene sets can include various display fields that provide metadata about each subject gene set (e.g., a name, count of genes within the gene set, and a description of the gene set). The name and description can be editable by a user and will be helpful for informing users about pertinent characteristics of the gene set. Also, the gene sets can be listed in a manner that indicates any hierarchical relationships that exist within gene sets.

Gene set controls 1710 (shown in greater detail in Figure 17B) provide a user with an ability to create gene sets from other gene sets via Boolean operations. For example, via the union button, a user can create a new gene set that is the union of two or more selected gene sets within list 1700. Via the intersection button, a user can create a new gene set that consists of the genes that are present in both/all of two or more selected gene sets within list 1700. Via the complement button, a user can create a complementary gene set from two gene sets (Gene Set 1 and Gene Set 2) within list 1700 where a first complementary new gene set consists of all of the genes that are in Gene Set 1 but not Gene Set 2 and where a second complementary new gene set consists of all of the genes that are in Gene Set 2 but not Gene Set 1. Via the "All Comparisons" button, a user can create several new gene sets at the same time via each of the techniques described above (new gene sets via the union operation, the intersection operation, and the complementary operation). This flexible ability to create new gene sets from existing gene sets allows for the comparison of gene sets between treatment conditions, patients, experiments, etc. Accordingly, by using controls 1710, a user can create gene collections that comprise combinations of gene sets.

Another powerful and innovative aspect of the visualizations provided by program 116 include an ability to overlap a third dimension on the cell view and/or gene view scatterplots. For example, color coding (e.g., a heatmap) can be applied to the cells 308 or genes 1106 in a scatterplot to provide another dimension to the data presentation. An example of this is shown by Figures 18A-D. Figure 18A shows how third dimension controls 1800 can be used to define how the third dimension is presented within an example cell view scatterplot display (e.g., as a heatmap statistic). Through control 1804, the user can define the statistic to be used for the heatmap. Figure 18B shows the options that are available for this statistic 1804 in an example embodiment (e.g., median, mean, geometric mean, coefficient of variation (CV), robust CV, standard deviation (SD), robust SD, etc.). A heatmap statistic can be calculated as follows: Each dot in the graph represents one or more cells (usually multiple cells). For each dot, the selected statistic (e.g., mean, median, etc) is calculated for the cell(s). Thus if there are 400 dots in the graph, then 400 statistical values are calculated. For the 400 statistical values, the min and max are determined, and are used as the lower and upper bound to index into a color map (an array of color values whose gradient changes from one color to another) i.e the min value is mapped to index 0 in the color array, and the max value is mapped to the last index in the color map. Colors in the color map can then be applied to values between the min and max values in some fashion (such as a linear distribution of colors to values).

Figure 18C shows how a user can select a parameter 1808 from the cellular gene expression data for use as the third dimension. Parameter list 1810 can be populated with a list of the parameters for the cellular gene expression data (e.g., the columns in table 200). In this example, the user has selected the parameter HighTHighB as a third dimension to be overlayed via color coding on the cell view scatterplot, as shown by Figure 18D. This third dimension overlay provides a user with a further capability for meaningfully interpreting a scatterplot. For example, the user can use the color coding to identify cell populations of interest and gate such cell populations (see gate 1900 in Figure 19), thereby creating another new object in the workspace as a cell population for further investigation.

Furthermore, according to another aspect of the disclosed system, reports can be created in a report editor by dragging cell populations from the workspace into the report editor (see Figure 20A). As another example, cell populations can be overlayed by dragging one cell population on top of another to create 2D overlaps and heatmaps by gene set. Overlays of cell populations are shown in Figure 20A-D in two ways. In the scatter plot, the different populations are shown as dot plots of different colors in the same graph (and you can choose which population is laid on top of the other). In the heat map, the different populations are organized and appended horizontally as additional columns in the heat map, i.e. all cells of population 1 are rendered together and labeled, followed by all cells of population 2, etc. Figure 20B shows example heatmaps by gene sets and cell populations. Figure 20C shows an enlarged view of the leftmost portion of Figure 20B, and Figure 20D shows an enlarged view of the rightmost portion of Figure 20C.

Example Use Case:

As indicated above, the disclosed system provides a powerful mechanism for investigating cellular gene expression data 112 by switching between the cell view mode and gene view mode (or vice versa) while performing gating in those two viewing modes to focus on data of interest.

Figure 21 depicts an example process flow for execution as part of program 114 that describes how a user can switch between the cell view mode and gene view mode to support deep investigations of cellular gene expression data. At step 2100, the processor selects a sample from the workspace in response to user input. At step 2102, a decision is made regarding whether to operate in cell view mode or gene view mode. This decision can be made in response to user input. If a cell view mode is selected, the process flow proceeds to step 2104. If a gene view mode is selected, the process flow proceeds to step 2122. Steps

2104-2120 correspond to operation in the cell view mode, while steps 2122-2140 correspond to operation in the gene view mode.

At step 2104, the processor selects cell view axis parameters in response to user input (e.g., selection of parameters such as genes or gene sets). At step 2106, the processor generates a cell view scatterplot data structure from the cellular gene expression data 112 based on the axis parameters selected at step 2104. This step may involve selecting the columns corresponding to the selected parameters in table 200 to obtain a list of cells and their associated values for each selected parameter. At step 2108, the processor generates the cell view scatterplot 300 from the data within the cell view scatterplot data structure created at step 2106 for presentation to the user.

At step 2110, the processor receives a gate specification with respect to the cell view scatterplot in response to input from a user. This gating creates a cell population (step 2112), where the created cell population gets saved as a new data structure in the workspace. At this point, the user can choose whether to (1) work with a new sample (see step 2114 with progression back to step 2100), (2) define one or more new axis parameters with respect to the current sample while in the cell view mode (see step 2116 with progression back to step 2104), (3) define a new gate with respect to the current sample while in the cell view mode (see step 2118 with progression back to step 2110), or (4) switch to the gene view mode (see step 2120 with progression to step 2122).

At step 2122, the processor pivots the cellular genetic expression data 112 as discussed above. Then, at step 2124, the processor selects cell populations from the workspace in response to user input. At step 2126, the processor generates a gene view scatterplot data structure from the pivoted cellular gene expression data based on the cell populations selected at step 2124. This step may involve selecting the pivoted columns corresponding to the selected cell populations in a pivoted version of table 200 to obtain a list of genes and their associated metrics for each selected cell population. At step 2128, the processor generates the gene view scatterplot 1100 from the data within the gene view scatterplot data structure created at step 2126 for presentation to the user.

At step 2130, the processor receives a gate specification with respect to the gene view scatterplot in response to input from a user. This gating creates a gene set (step 2132), where the created gene set gets saved as a new synthetic parameter in the workspace for association with the cellular genetic expression data. The cellular genetic expression data can thus be augmented with new data values corresponding to the gene set created at step 2132. At this point, the user can choose whether to (1) work with a new sample (see step 2134 with

progression back to step 2100), (2) define one or more new cell populations with respect to the current sample while in the gene view mode (see step 2136 with progression back to step 2124), (3) define a new gate with respect to the current sample while in the gene view mode (see step 2138 with progression back to step 2130), or (4) switch to the cell view mode (see step 2140 with progression to step 2104).

Thus, Figure 21 shows how the system can be used by a user to quickly transition between cell view modes and gene view modes while creating cell populations and gene sets respectively via gating that can be used to aid visualizations after switching between modes.

As an example, a powerful and innovative mode of operation is shown by the example process flow of Figure 22 where a user interacts with program 114 via the process flow of Figure 21 to (1) create one or more cell populations within cell view modes of the display (step 2200), (2) pivot to a gene view mode that comparatively displays gene expressions across multiple cell populations (step 2202), (3) create one or more gene sets within gene view mode of the display (step 2204) which augments the cellular gene expression data with the new gene set(s) as a synthetic parameter (step 2206), (4) pivots back to a cell view scatterplot using one or more of these gene sets as axis parameter(s) for a cell view scatterplot (step 2208), and (5) iteratively repeats these operations as desired to drill down into biologically relevant relationships that might exist within the cellular gene expression data 112 (and where operations (1) and (3) may be aided with overlays of user-defined third dimensions on the cell view or gene view data display). For example, the disclosed system can serve as a powerful tool for research into precision/personalized medicine to perform work such as evaluating cell data from numerous patients to find patients with genes that correlate well with survival and therapy response to various cancers or other illnesses/pathologies.

For example, through the tools provided herein, a user might be able to analyze cell populations to identify differentially expressed gene sets that are correlated to better chances for survival with respect to a particular Cancer X (which we can label as "Survival Gene Set"). At the same time, a user might be able to analyze cell populations to identify differentially expressed gene sets that are correlated to poor chances for survival with respect to Cancer X (which we can label as "Not Survival Gene Set"). Further still, a user might be able to analyze cell populations to identify differentially expressed gene sets that are correlated to responding well to Therapy Y for Cancer X (which we can label "Therapy Responsive Gene Set"). Then, these gene sets can be used as synthetic parameters in the cell view mode of the system to find cell populations in patients that are genetically predisposed

to respond well to treatment by Therapy Y in order to survive Cancer X. For example, the cell view scatterplot can use the Survival Gene Set as one axis parameter (e.g., X-axis parameter) and the Not Survival Gene Set as the other axis parameter (e.g., Y-axis parameter), while using the Therapy Responsive Gene Set as the third dimensional overlay. The resultant scatterplot can show cell populations that will correlate well with both survival and therapy responsiveness (as well as cell populations that do not correlate well with survival or therapy responsiveness).

While the invention has been described above in relation to its example embodiments, various modifications may be made thereto that still fall within the invention's scope. For example, while the example scatterplots shown herein present the X-axis as a horizontal axis and the Y-axis as a vertical axis, it should be understood that some practitioners may find it desirable to tilt the scatterplots. An example would a scenario where the diagonal y=x is deemed biologically important. In such a case, the scatterplot might be tilted so that the y=x diagonal is presented as a horizontal or vertical line rather than a 45 degree line to thereby help focus users on how far data might lie away from the y=x line. Accordingly, it should be understood that these and other modifications to the invention will be recognizable upon review of the teachings herein.

WHAT IS CLAIMED IS:

1. A method of visualizing a multi-parameter data set comprising:

a processor generating a scatterplot of the multi-parameter data set across a first and second axis, the data set comprising a plurality of data items, each data item being associated with a plurality of parameters and exhibiting data values for each of the associated parameters, and wherein the first and second axis correspond to user-selected parameters within the data set, the scatterplot comprising a plurality of dots, each dot corresponding to a data item from the data set and being positioned on the scatterplot at a position along the first and second axes at a location corresponding to the data values for the parameters corresponding to the axes.

2. The method of claim 1 wherein the multi-parameter data set comprises cellular gene expression data, the data items comprising a plurality of cells, wherein the parameters comprise a plurality of genes, and wherein the data values comprise data indicative of counts for expressions of the corresponding to genes in each cell.

3. The method of any of claims 1-2 wherein the scatterplot comprises a cell view scatterplot.

4. The method of claim 3 wherein the generating step comprises a processor generating the cell view scatterplot from the cellular gene expression data based on user specification of at least one gene set as an axis parameter for the cell view scatterplot.

5. The method of any of claims 3-4 further comprising:

a processor gating a group of dots in the cell view scatterplot to define a cell population; and

a processor creating a data object in a workspace, the data object being representative of the defined cell population.

6. The method of claim 5 further comprising a processor repeating the gating and creating steps to create a plurality of data objects in the workspace that represent a plurality of cell populations.

7. The method of any of claims 5-6 further comprising:

a processor combining a plurality of cell populations according to Boolean operations to generate new cell populations for inclusion as data objects in the workspace.

8. The method of any of claims 3-7 further comprising:

altering the cell view scatterplot to spread dots corresponding to zeros across a defined zero space of the scatterplot.

9. The method of any of claims 3-8 further comprising:

a processor pivoting from the cell view scatterplot to a gene view scatterplot that plots a plurality of different genes across a plurality of cell populations.

10. The method of claim 9 further comprising:

a processor gating a group of dots in the gene view scatterplot to define a gene set; and

a processor creating a synthetic parameter data object in a workspace, the synthetic parameter data object being representative of the defined gene set in the workspace.

11. The method of claim 10 further comprising:

a processor augmenting the cellular gene expression data with the defined gene set as a synthetic parameter associated with a cell in the cellular gene expression data.

12. The method of any of claims 10-11 further comprising:

a processor generating a cell view scatterplot where the defined gene set is selected as one of the axis parameters.

13. The method of claim 12 further comprising:

a processor gating a group of dots in the cell view scatterplot based on the defined gene set to define another cell population; and

a processor creating a data object in the workspace that is representative of the defined another cell population.

14. The method of any of claims 10-13 further comprising:

a processor combining a plurality of gene sets according to Boolean operations to generate new gene sets for inclusion as data objects in the workspace.

15. The method of any of claims 2-14 wherein the scatterplot comprises a gene view scatterplot from the cellular gene expression data, and wherein the generating step comprises a processor generating the gene view scatterplot based on user specification of a plurality of cell populations.

16. The method of claim 15 further comprising:

a processor pivoting from a cell view scatterplot of the cellular gene expression data to the gene view scatterplot in response to the user specification of a plurality of cell populations.

17. The method of any of claims 15-16 further comprising:

a processor generating a gene set by gating a plurality of genes in the gene view scatterplot.

18. The method of any of claims 16-17 further comprising:

a processor pivoting to a cell view scatterplot from the gene view scatterplot based on a user specification of a gene set for use as an axis parameter in the cell view scatterplot.

19. The method of any of claims 2-18 wherein the cellular gene expression data comprises a plurality of stochastic labels derived from stochastic labeling of a sample.

20. The method of any of claims 1-19 further comprising:

a processor overlaying a third dimension of parameter data values on a scatterplot in response to user input.

21. The method of any of claims 1-20 further comprising:

a processor adjusting how a scatterplot is displayed based on user input.

22. A method comprising:

a processor gating a region of a cell view scatterplot of cellular gene expression data to create a cell population, wherein the cellular gene expression data comprises a plurality of

data values for a plurality of parameters, wherein the cell view scatterplot comprises first and second axes corresponding to a first parameter and a second parameter of the cellular gene expression data;

a processor pivoting to a gene view scatterplot of the cellular gene expression data, wherein the gene view scatterplot comprises at least one axis corresponding to the created cell population; and

a processor gating a region of the gene view scatterplot of the cellular gene expression data to create a gene set.

23. The method of claim 22 further comprising:

a processor augmenting the cellular gene expression data with the created gene set such that the created gene set is available as a synthetic parameter of the cellular gene expression data.

24. The method of any of claims 22-23 further comprising:

a processor pivoting to another cell view scatterplot of the cellular gene expression data, wherein the another cell view scatterplot comprises at least one axis corresponding to the created gene set.

25. The method of any of claims 22-24 further comprising;

a processor selecting parameters of the cellular gene expression data for use as axes of the scatterplots in response to user input.

26. The method of any of claims 22-25 further comprising:

performing the method steps to create a first gene set corresponding to genes that are differentially expressed in a plurality of cell populations and correlate to a relatively better chance for survival with respect to a cancer type; and

performing the method steps to create a second gene set corresponding to genes that are differentially expressed in a plurality of cell populations and correlate to a relatively poorer chance for survival with respect to the cancer type.

27. The method of claim 26 further comprising:

a processor pivoting to another cell view scatterplot of the cellular gene expression data, wherein the another cell view scatterplot comprises (1) a first axis corresponding to the first gene set and (2) a second axis corresponding to the second gene set.

28. The method of claim 27 further comprising:

a processor gating a region of the another cell view scatterplot of the cellular gene expression data to create a cell population that correlates to a relatively better chance for survival with respect to the cancer type.

29. The method of claim 28 further comprising:

a processor augmenting the cellular gene expression data with the cell population that correlates to a relatively better chance for survival with respect to the cancer type such that the cell population that correlates to a relatively better chance for survival with respect to the cancer type is available as a synthetic parameter of the cellular gene expression data.

30. The method of any of claims 27-29 further comprising:

a processor gating a region of the another cell view scatterplot of the cellular gene expression data to create a cell population that correlates to a relatively poorer chance for survival with respect to the cancer type.

31. The method of claim 30 further comprising:

a processor augmenting the cellular gene expression data with the cell population that correlates to a relatively poorer chance for survival with respect to the cancer type such that the cell population that correlates to a relatively poorer chance for survival with respect to the cancer type is available as a synthetic parameter of the cellular gene expression data.

32. The method of any of claims 26-31 further comprising:

performing the method steps to create a third gene set corresponding to genes that are differentially expressed in a plurality of cell populations and correlate to a relatively better therapy responsiveness with respect to the cancer type.

33. The method of claim 32 further comprising:

a processor pivoting to another cell view scatterplot of the cellular gene expression data, wherein the another cell view scatterplot comprises (1) a first axis corresponding to the first gene set and (2) a second axis corresponding to the second gene set; and

a processor overlaying a third dimension of the cellular gene expression data on the another cell view scatterplot, wherein the third dimension corresponds to the third gene set.

34. The method of claim 33 further comprising:

a processor gating a region of the another cell view scatterplot of the cellular gene expression data with the third gene set overlay to create a cell population that correlates to both a relatively better chance for survival with respect to the cancer type and a relatively better therapy responsiveness with respect to the cancer type.

35. The method of claim 34 further comprising:

a processor augmenting the cellular gene expression data with the cell population that correlates to both a relatively better chance for survival with respect to the cancer type and a relatively better therapy responsiveness with respect to the cancer type such that the cell population that correlates to both a relatively better chance for survival with respect to the cancer type and a relatively better therapy responsiveness with respect to the cancer type is available as a synthetic parameter of the cellular gene expression data.

36. The method of any of claims 26-35 further comprising:

performing the method steps to create a third gene set corresponding to genes that are differentially expressed in a plurality of cell populations and correlate to a relatively poorer therapy responsiveness with respect to the cancer type.

37. The method of claim 36 further comprising:

a processor pivoting to another cell view scatterplot of the cellular gene expression data, wherein the another cell view scatterplot comprises (1) a first axis corresponding to the first gene set and (2) a second axis corresponding to the second gene set; and

a processor overlaying a third dimension of the cellular gene expression data on the another cell view scatterplot, wherein the third dimension corresponds to the third gene set.

38. The method of claim 37 further comprising:

a processor gating a region of the another cell view scatterplot of the cellular gene expression data with the third gene set overlay to create a cell population that correlates to a relatively better chance for survival with respect to the cancer type but a relatively poorer therapy responsiveness with respect to the cancer type.

39. The method of claim 38 further comprising:

a processor augmenting the cellular gene expression data with the cell population that correlates to a relatively better chance for survival with respect to the cancer type but a relatively poorer therapy responsiveness with respect to the cancer type such that the cell population that correlates to a relatively better chance for survival with respect to the cancer type but a relatively poorer therapy responsiveness with respect to the cancer type is available as a synthetic parameter of the cellular gene expression data.

40. The method of any of claims 22-39 further comprising:

a processor overlaying a third dimension of the cellular gene expression data on at least one of the scatterplots in response to a user selection of a parameter of the cellular gene expression data for use as the third dimension.

41. The method of any of claims 22-40 further comprising:

a processor augmenting the cellular gene expression data with the created cell population such that the created cell population is available as a synthetic parameter of the cellular gene expression data.

42. The method of any of claims 22-41 wherein the pivoting step comprises:

a processor performing a statistical computation on a plurality of parameter data values in the cellular gene expression data to create a plurality of pivoted data values; and

a processor populating a pivoted view of the cellular gene expression data with the pivoted data values.

43. A method comprising:

a processor generating a gene view scatterplot of cellular gene expression data based on user specification of a plurality of cell populations.

44. The method of claim 43 further comprising:

a processor pivoting from a cell view scatterplot of cellular gene expression data to the gene view scatterplot in response to the user specification of a plurality of cell populations.

45. The method of any of claims 43-44 further comprising:

a processor generating a gene set by gating a plurality of genes in the gene view scatterplot.

46. The method of any of claims 43-45 further comprising:

a processor pivoting to a cell view scatterplot from the gene view scatterplot based on a user specification of a gene set for use as an axis parameter in the cell view scatterplot.

47. A method comprising:

a processor generating a cell view scatterplot of cellular gene expression data based on user specification of at least one gene set as an axis parameter for the cell view scatterplot.

48. A method comprising:

a processor iteratively pivoting from a cell view scatterplot to a gene view scatterplot in response to user input while performing gating within the cell view and gene view scatterplots in response to user input to create cell populations and gene sets respectively, the cell populations and gene sets for use as axis parameters in the gene view scatterplots and cell view scatterplots respectively.

49. An apparatus comprising:

a memory configured to store a multi-parameter data set; and

a processor for cooperation with the memory, the processor configured to perform the method of any of claims 1-48.

50. The apparatus of claim 49 further comprising:

a display in cooperation with the processor, wherein the display is configured to graphically present the scatterplots of any of claims 1-48 to a user.

51. A computer program product comprising:

a plurality of processor-executable instructions that are resident on a non-transitory computer readable storage medium, wherein the instructions are configured, upon execution by a processor, to cause the processor to perform the method of any of claims 1-48.

52. The computer program product of claim 51 wherein the instructions, upon execution by the processor, are further configured to instruct a display to graphically present the scatterplots of any of claims 1-48 to a user.

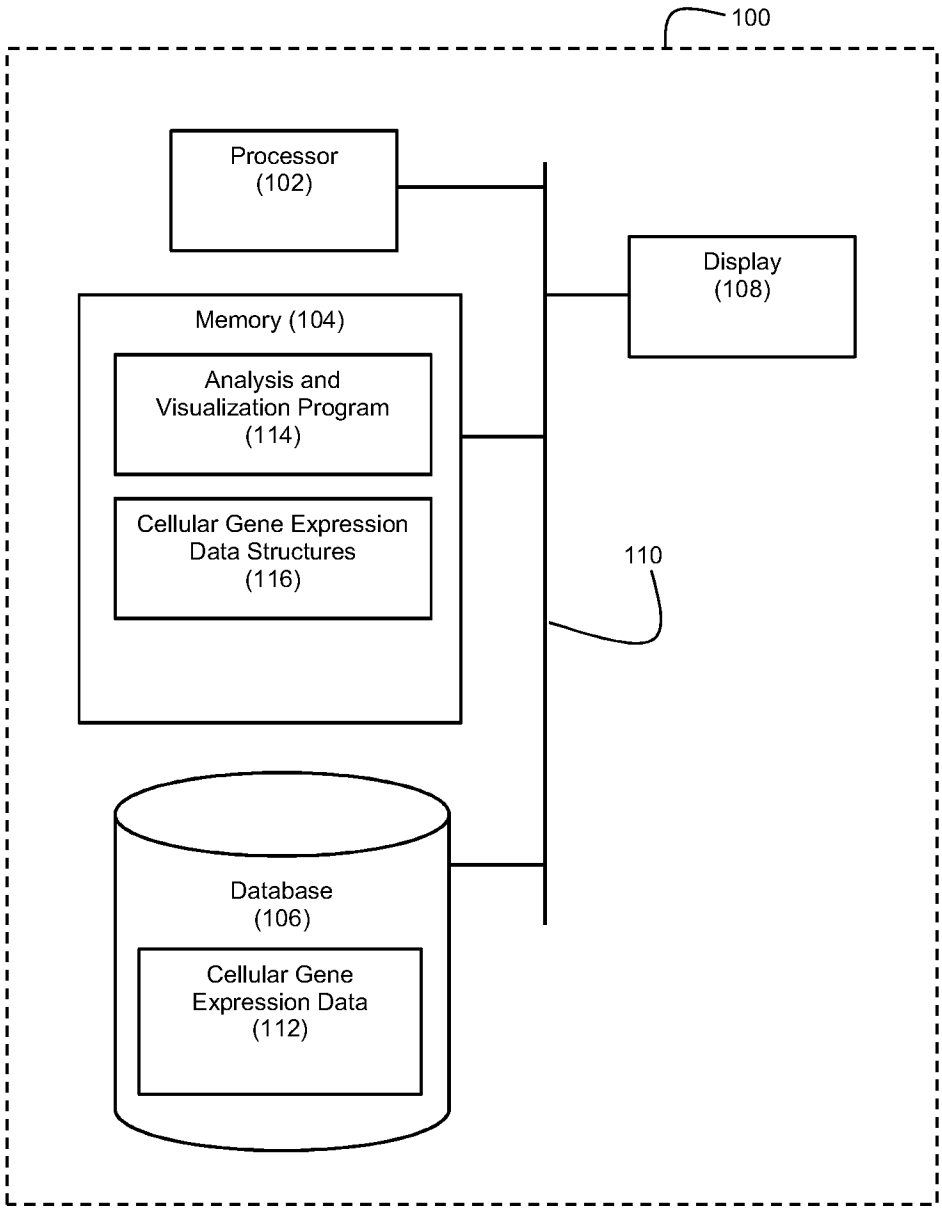

100
Processor
(102)
Display
(108)
Memory (104)
Analysis and
Visualization Program
(114)
Cellular Gene Expression
Data Structures
(116)
110
Database
(106)
Cellular Gene
Expression Data
(112)

Figure 1

{Cell ID 1, Parameter 1 (ID, Value), Parameter 2 (ID, Value), ... , Parameter n (ID, Value)}
{Cell ID 2, Parameter 1 (ID, Value), Parameter 2 (ID, Value), ... , Parameter n (ID, Value)}

.....

Figure 2A

200

Cell	Gene 1	Gene 2	Gene 3	.....
ID1	0	6	0	...
ID2	1	4	0	...
ID3	3	5	2	...

.....

Figure 2B

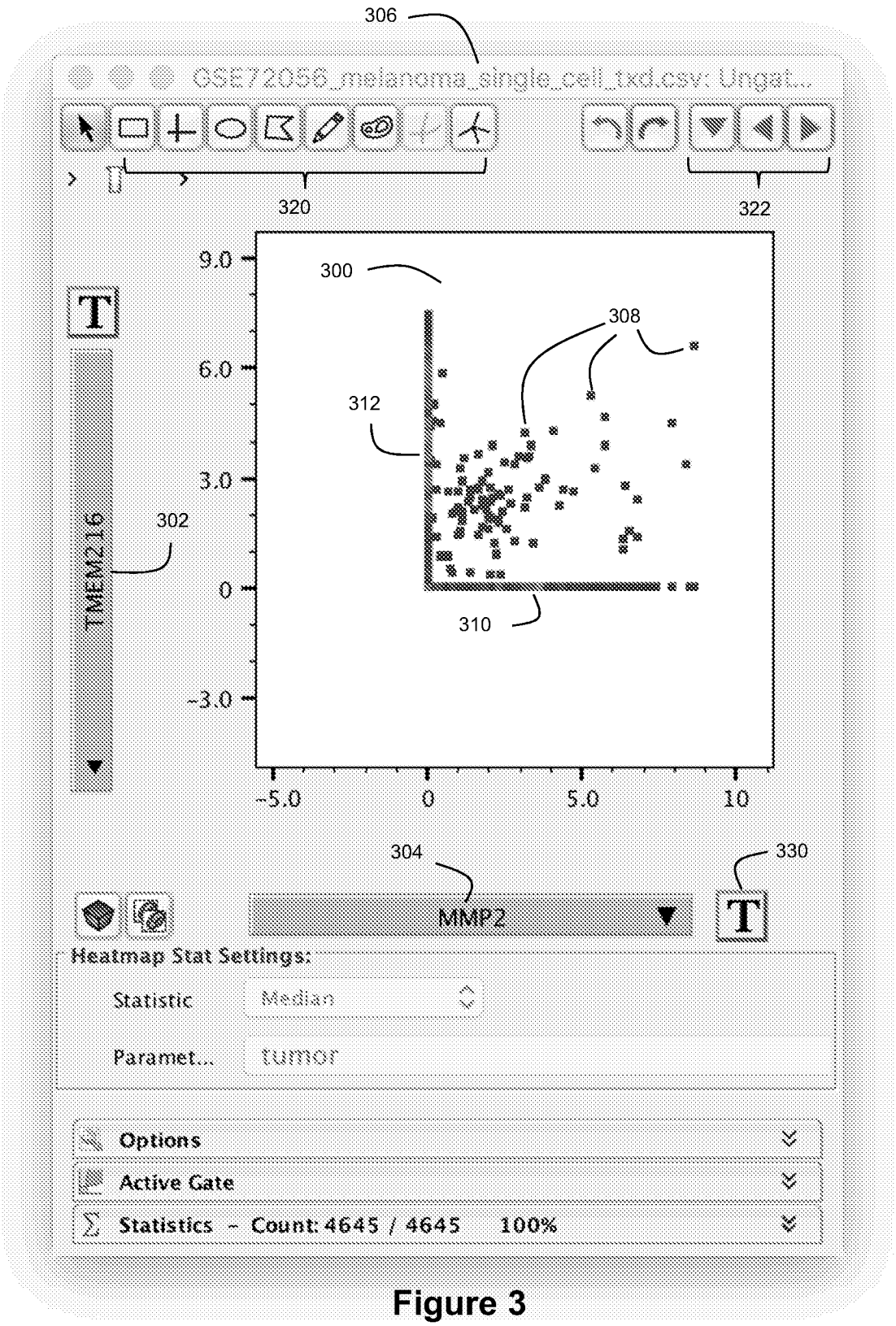

306
GSE72056_melanoma_single_cell_txd.csv: Ungat...
320
322
9.0
6.0
3.0
0
-3.0
300
308
312
302
310
TMEM216
-5.0
0
5.0
10
304
330
MMP2
Heatmap Stat Settings:
Statistic
Median
Paramet...
tumor
Options
Active Gate
Statistics - Count: 4645 / 4645 100%

Figure 3

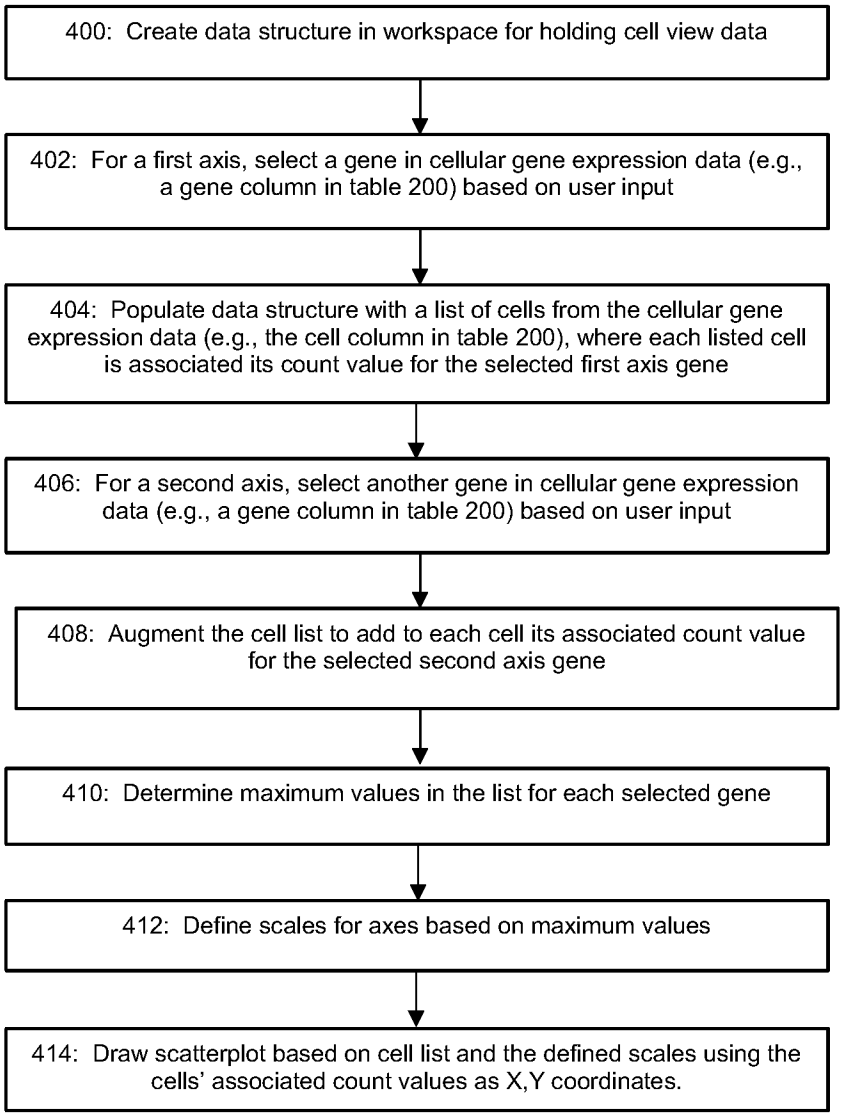

400: Create data structure in workspace for holding cell view data
402: For a first axis, select a gene in cellular gene expression data (e.g., a gene column in table 200) based on user input
404: Populate data structure with a list of cells from the cellular gene expression data (e.g., the cell column in table 200), where each listed cell is associated its count value for the selected first axis gene
406: For a second axis, select another gene in cellular gene expression data (e.g., a gene column in table 200) based on user input
408: Augment the cell list to add to each cell its associated count value for the selected second axis gene
410: Determine maximum values in the list for each selected gene
412: Define scales for axes based on maximum values
414: Draw scatterplot based on cell list and the defined scales using the cells' associated count values as X,Y coordinates.

Figure 4

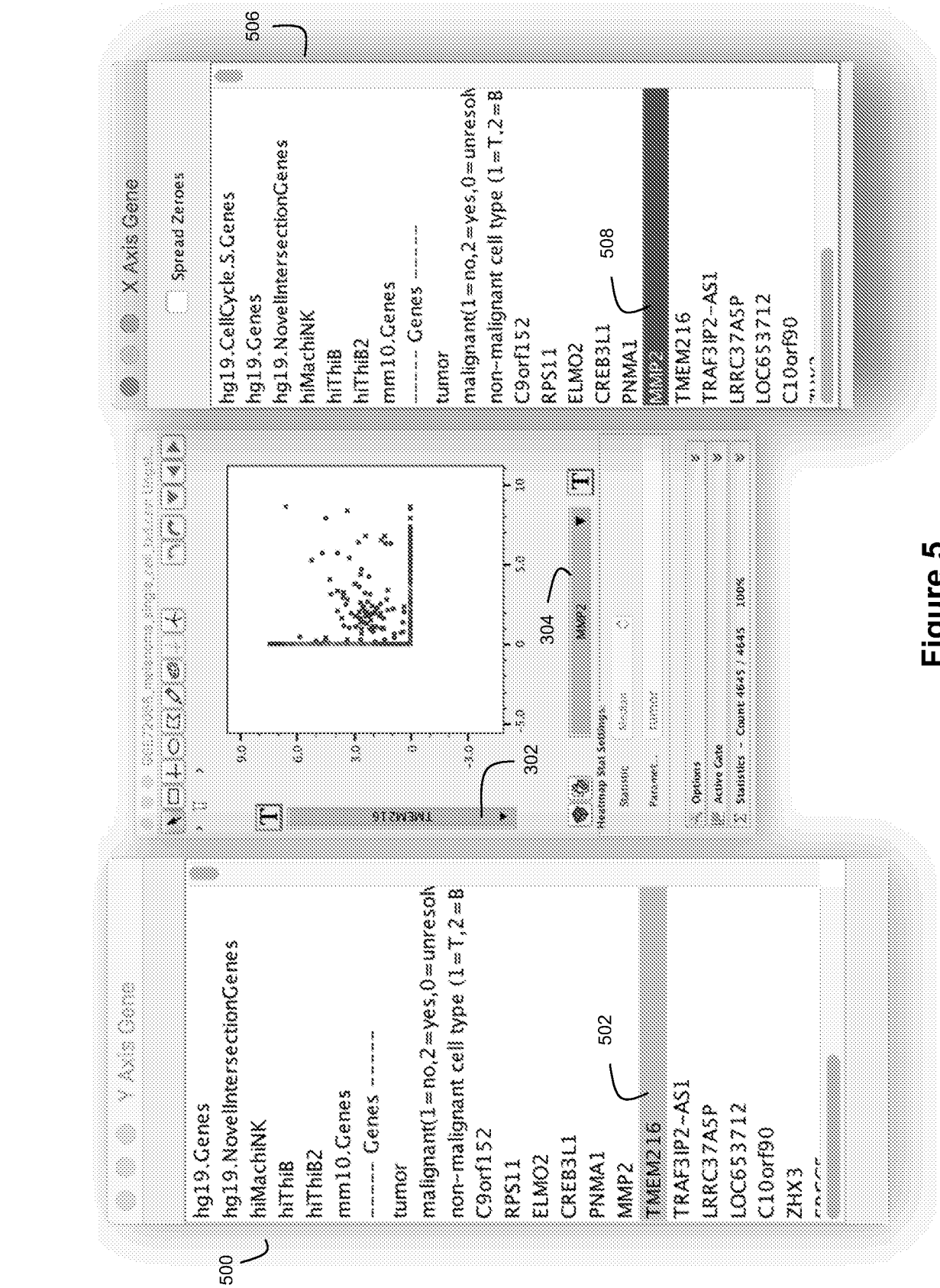

Y Axis Gene
hg19.Genes
hg19.NovelIntersectionGenes
hiMachiNK
hiThiB
hiThiB2
mm10.Genes
------ Genes ------
tumor
malignant(1=no,2=yes,0=unresoh
non-malignant cell type (1=T,2=B
C9orf152
RPS11
ELMO2
CREB3L1
PNMA1
MMP2
TMEM216
TRAF3IP2-AS1
LRRC37A5P
LOC653712
C10orf90
ZHX3
500
502
X Axis Gene
Spread Zeroes
hg19.CellCycle.S.Genes
hg19.Genes
hg19.NovelIntersectionGenes
hiMachiNK
hiThiB
hiThiB2
mm10.Genes
------ Genes ------
tumor
malignant(1=no,2=yes,0=unresoh
non-malignant cell type (1=T,2=B
C9orf152
RPS11
ELMO2
CREB3L1
PNMA1
MMP2
TMEM216
TRAF3IP2-AS1
LRRC37A5P
LOC653712
C10orf90
506
508
302
304

Figure 5

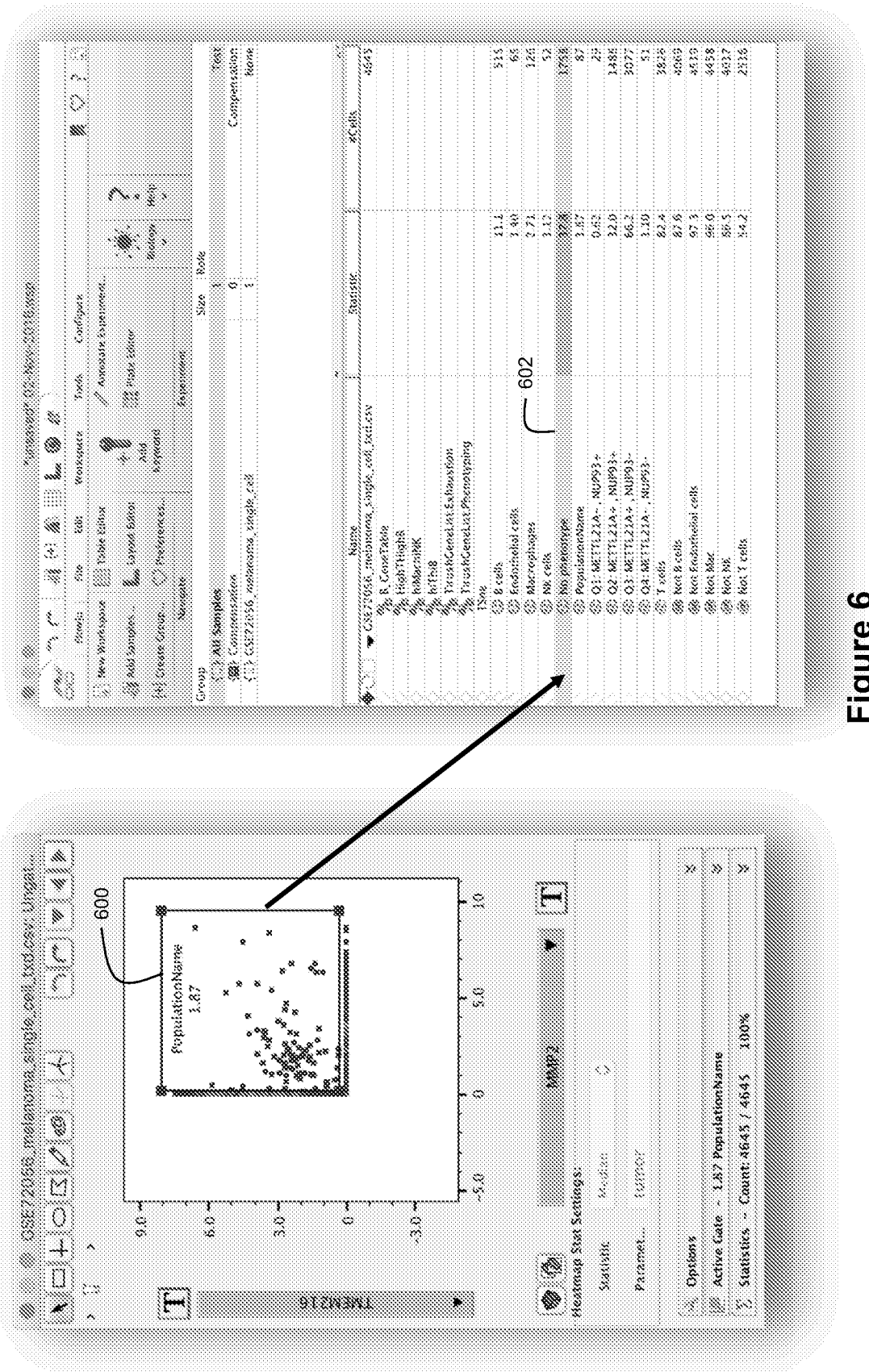


Figure 6

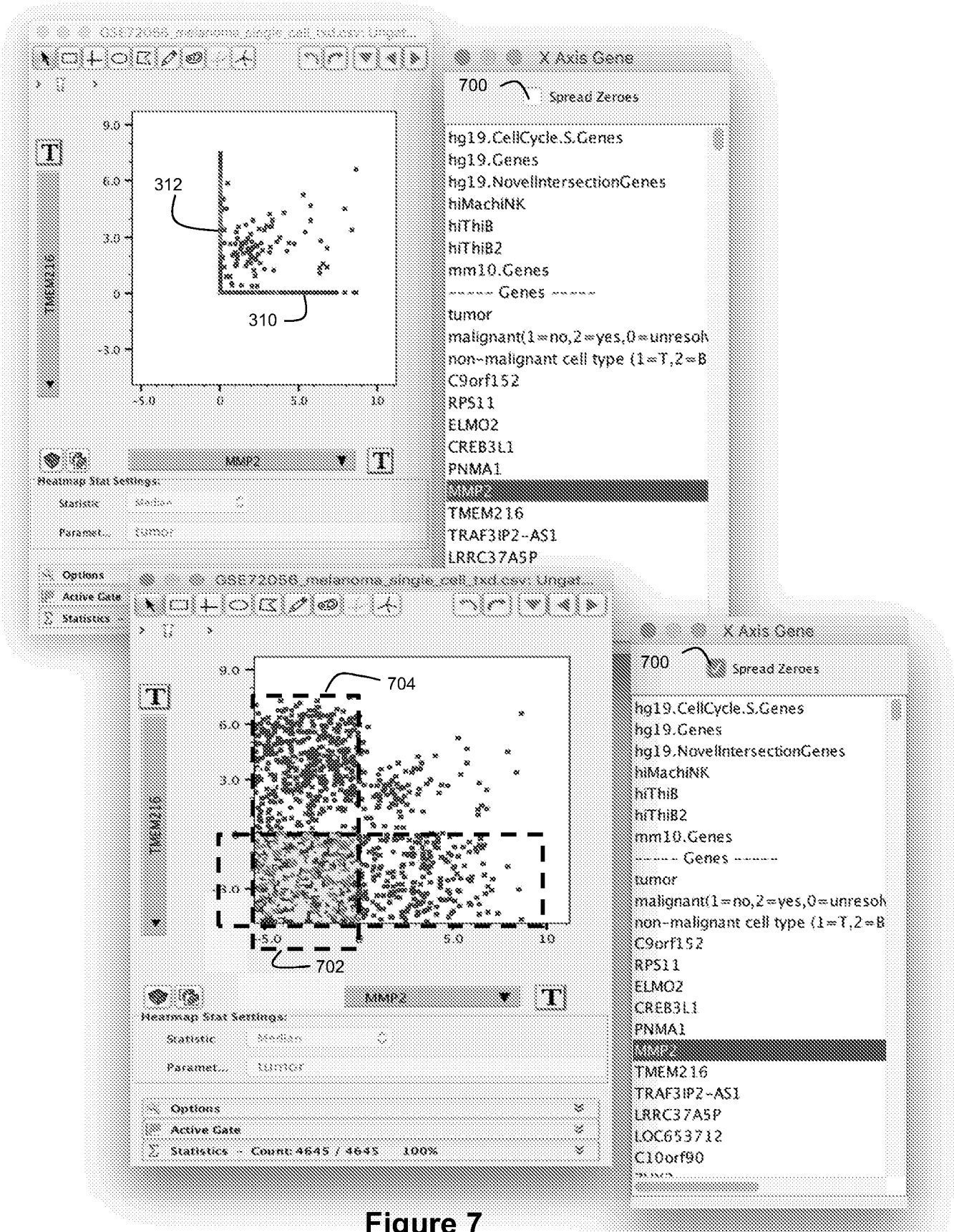

X Axis Gene
700
Spread Zeroes
hg19.CellCycle.S.Genes
hg19.Genes
hg19.NovelIntersectionGenes
hiMachiNK
hiThiB
hiThiB2
mm10.Genes
------ Genes ------
tumor
malignant(1=no,2=yes,0=unresol
non-malignant cell type (1=T,2=B
C9orf152
RPS11
ELMO2
CREB3L1
PNMA1
MMP2
TMEM216
TRAF3IP2-AS1
LRRC37A5P
LOC653712
C10orf90
312
310
704
702
TMEM216
MMP2
Heatmap Stat Settings:
Statistic
Median
Paramet...
tumor
Options
Active Gate
Statistics
Statistics - Count: 4645 / 4645 100%

Figure 7

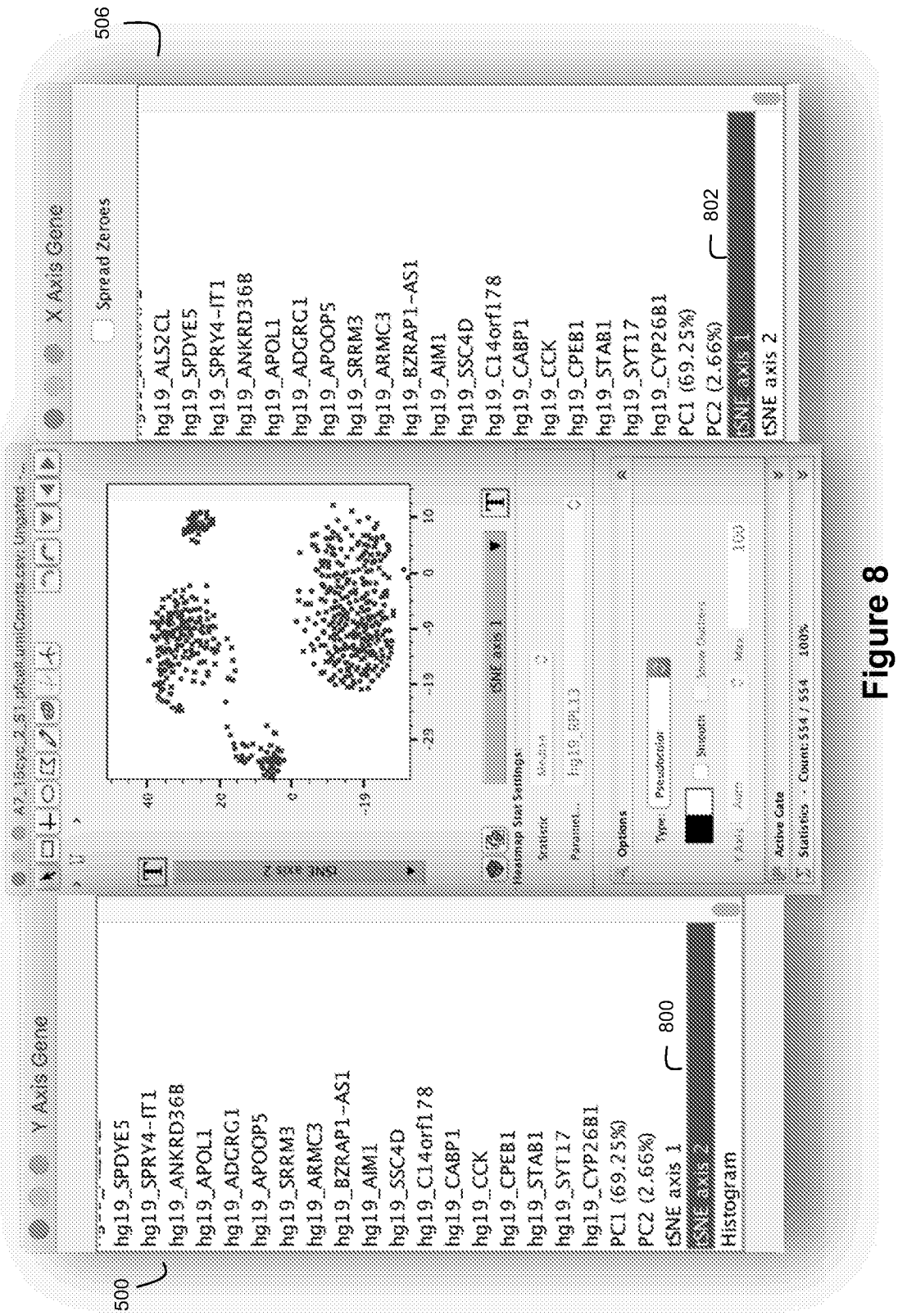

500
506
800
802
Y Axis Gene
X Axis Gene
Spread Zeroes
hg19_SPDYE5
hg19_SPRY4-IT1
hg19_ANKRD36B
hg19_APOL1
hg19_ADGRG1
hg19_APOOP5
hg19_SRRM3
hg19_ARMC3
hg19_BZRAP1-AS1
hg19_AIM1
hg19_SSC4D
hg19_C14orf178
hg19_CABP1
hg19_CCK
hg19_CPEB1
hg19_STAB1
hg19_SYT17
hg19_CYP26B1
PC1 (69.25%)
PC2 (2.66%)
tSNE axis 1
tSNE axis 2
Histogram
Options
Active Gate
Statistics - Count: 554 / 554 100%

Figure 8

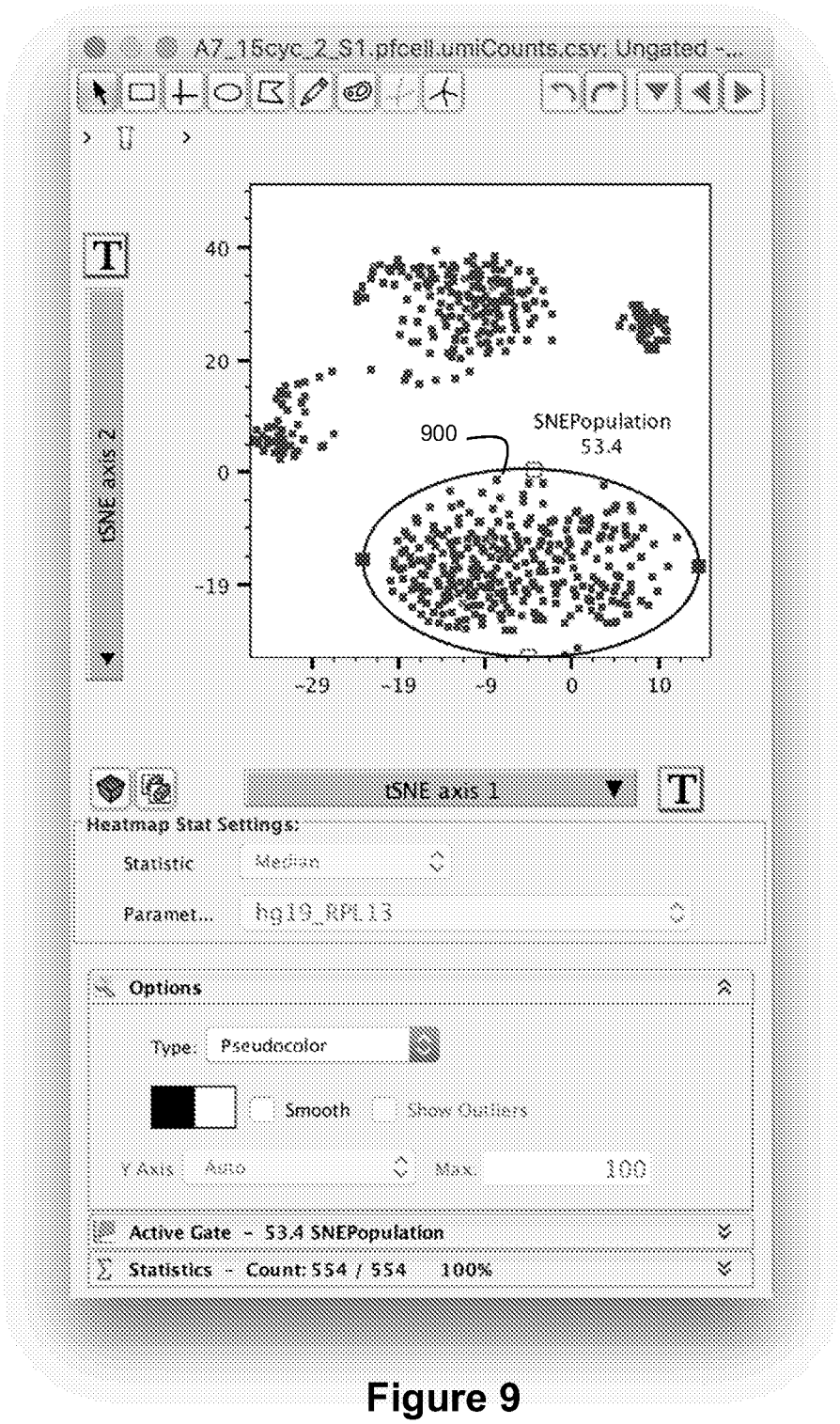

A7_15cyc_2_S1.pfcell.umiCounts.csv: Ungated -...
tSNE axis 2
40
20
0
-19
900
SNEPopulation
53.4
-29
-19
-9
0
10
tSNE axis 1
Heatmap Stat Settings:
Statistic
Median
Paramet...
hg19_RPL13
Options
Type:
Pseudocolor
Smooth
Show Outliers
Y Axis:
Auto
Max:
100
Active Gate - 53.4 SNEPopulation
Statistics - Count: 554 / 554 100%

Figure 9

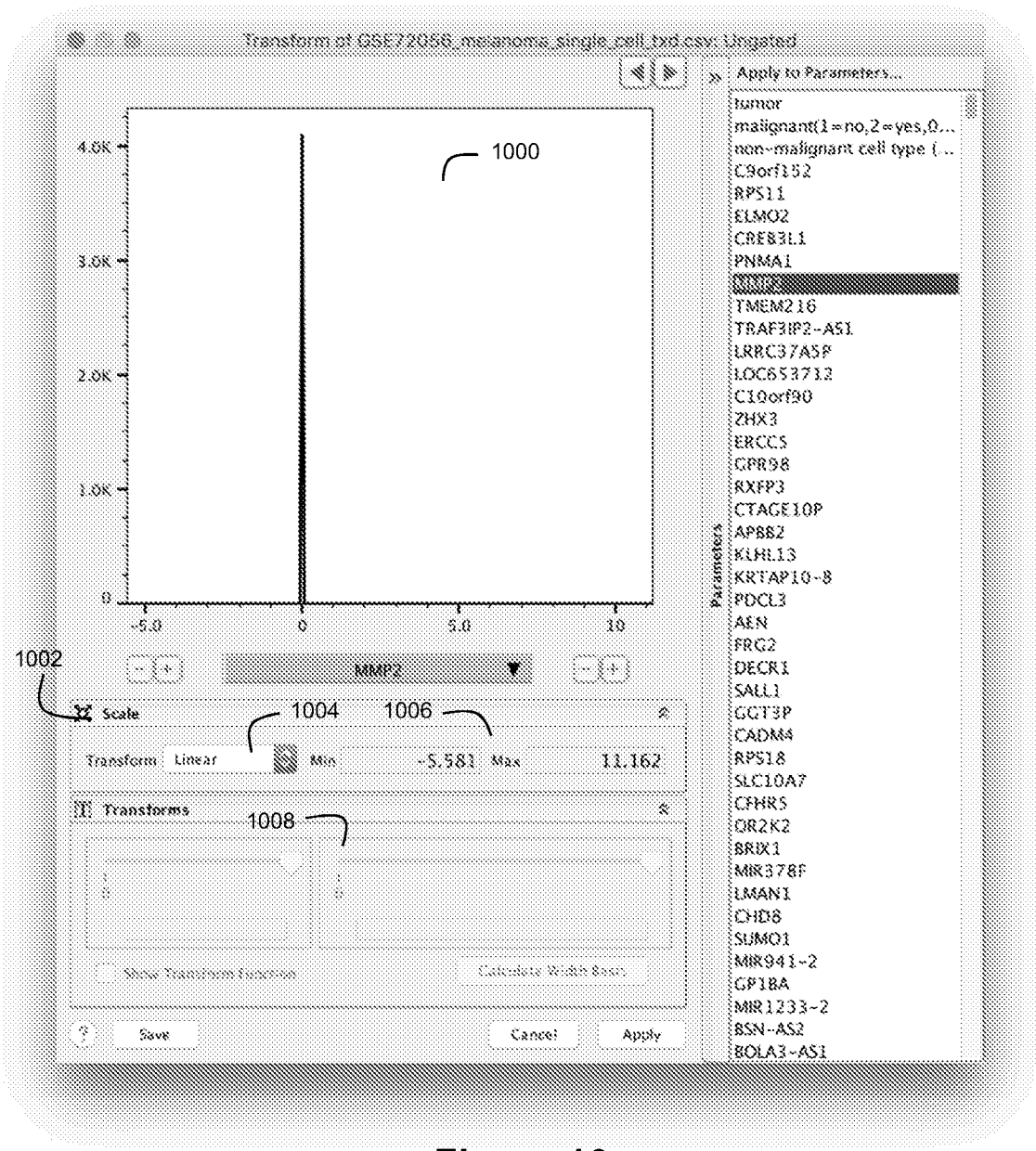

Transform of GSE72056_melanoma_single_cell_txd.csv: Ungated
Apply to Parameters...
tumor
malignant(1=no,2=yes,0...
non-malignant cell type (...
C9orf152
RPS11
ELMO2
CREB3L1
PNMA1
MMP2
TMEM216
TRAF3IP2-AS1
LRRC37A5P
LOC653712
C19orf90
ZHX3
ERCC5
GPR98
RXFP3
CTAGE10P
APBB2
KLHL13
KRTAP10-8
PDCL3
AEN
FRG2
DECR1
SALL1
GGT3P
CADM4
RPS18
SLC10A7
CFHR5
OR2K2
BRIX1
MIR378F
LMAN1
CHD8
SUMO1
MIR941-2
GP1BA
MIR1233-2
BSN-AS2
BOLA3-AS1
Parameters
4.0K
3.0K
2.0K
1.0K
0
-5.0
0
5.0
10
1000
1002
1004
1006
1008
Scale
Transform
Linear
Min
-5.581
Max
11.162
Transforms
Save
Cancel
Apply

Figure 10

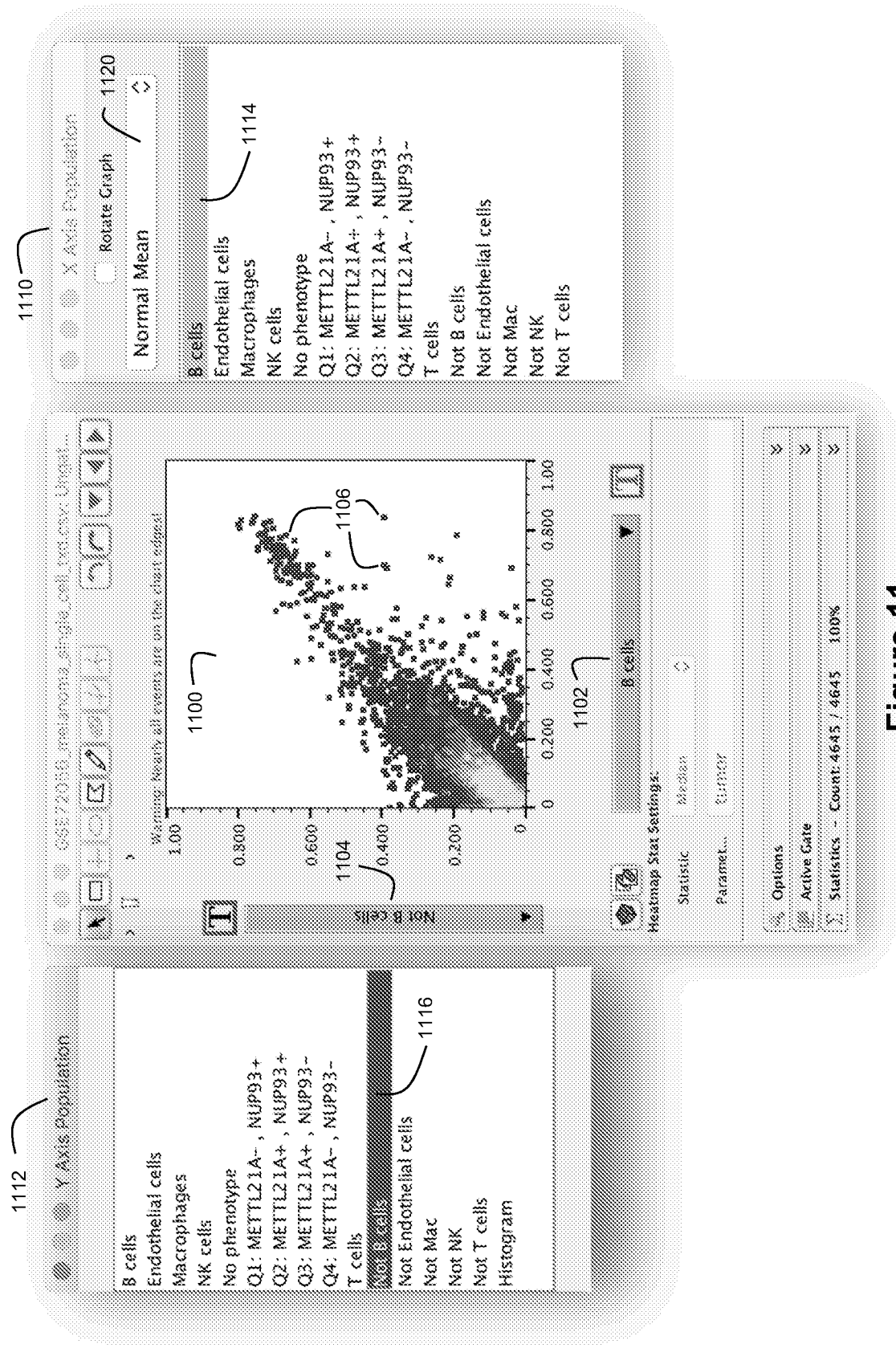
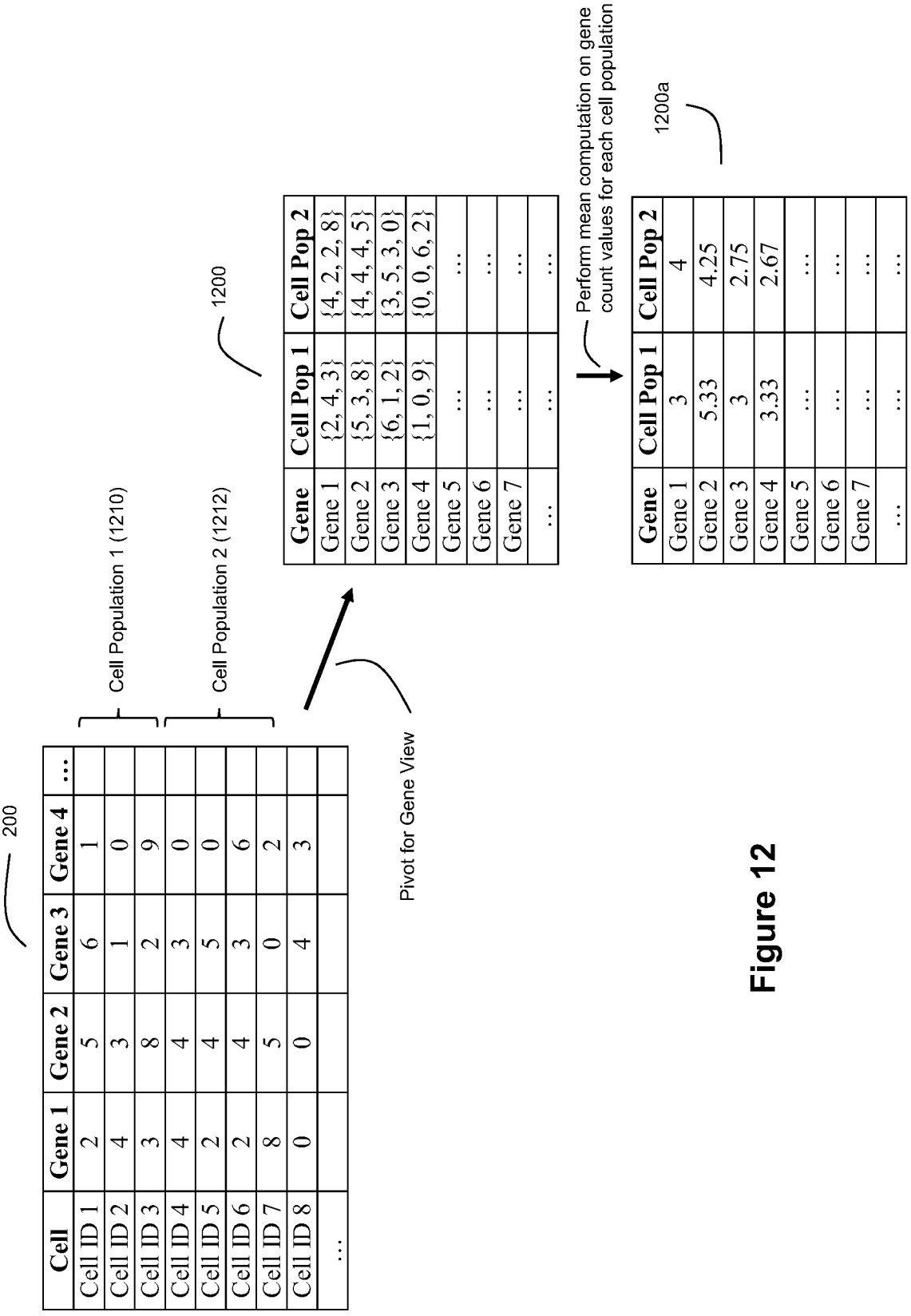

1110
1112
1120
1114
1116
1100
1102
1104
1106
X Axis Population
Rotate Graph
Normal Mean
B cells
Endothelial cells
Macrophages
NK cells
No phenotype
Q1: METTL21A- , NUP93+
Q2: METTL21A+ , NUP93+
Q3: METTL21A+ , NUP93-
Q4: METTL21A- , NUP93-
T cells
Not B cells
Not Endothelial cells
Not Mac
Not NK
Not T cells
Histogram
Y Axis Population
Heatmap Stat Settings
Statistic
Options
Active Gate
Statistics - Count: 4645 / 4645 100%

Figure 11



Cell	Gene 1	Gene 2	Gene 3	Gene 4	…
Cell ID 1	2	5	6	1	
Cell ID 2	4	3	1	0	
Cell ID 3	3	8	2	9	
Cell ID 4	4	4	3	0	
Cell ID 5	2	4	5	0	
Cell ID 6	2	4	3	6	
Cell ID 7	8	5	0	2	
Cell ID 8	0	0	4	3	
…					

Gene	Cell Pop 1	Cell Pop 2
Gene 1	{2, 4, 3}	{4, 2, 2, 8}
Gene 2	{5, 3, 8}	{4, 4, 4, 5}
Gene 3	{6, 1, 2}	{3, 5, 3, 0}
Gene 4	{1, 0, 9}	{0, 0, 6, 2}
Gene 5	…	…
Gene 6	…	…
Gene 7	…	…
…	…	…

Gene	Cell Pop 1	Cell Pop 2
Gene 1	3	4
Gene 2	5.33	4.25
Gene 3	3	2.75
Gene 4	3.33	2.67
Gene 5	…	…
Gene 6	…	…
Gene 7	…	…
…	…	…

Figure 12

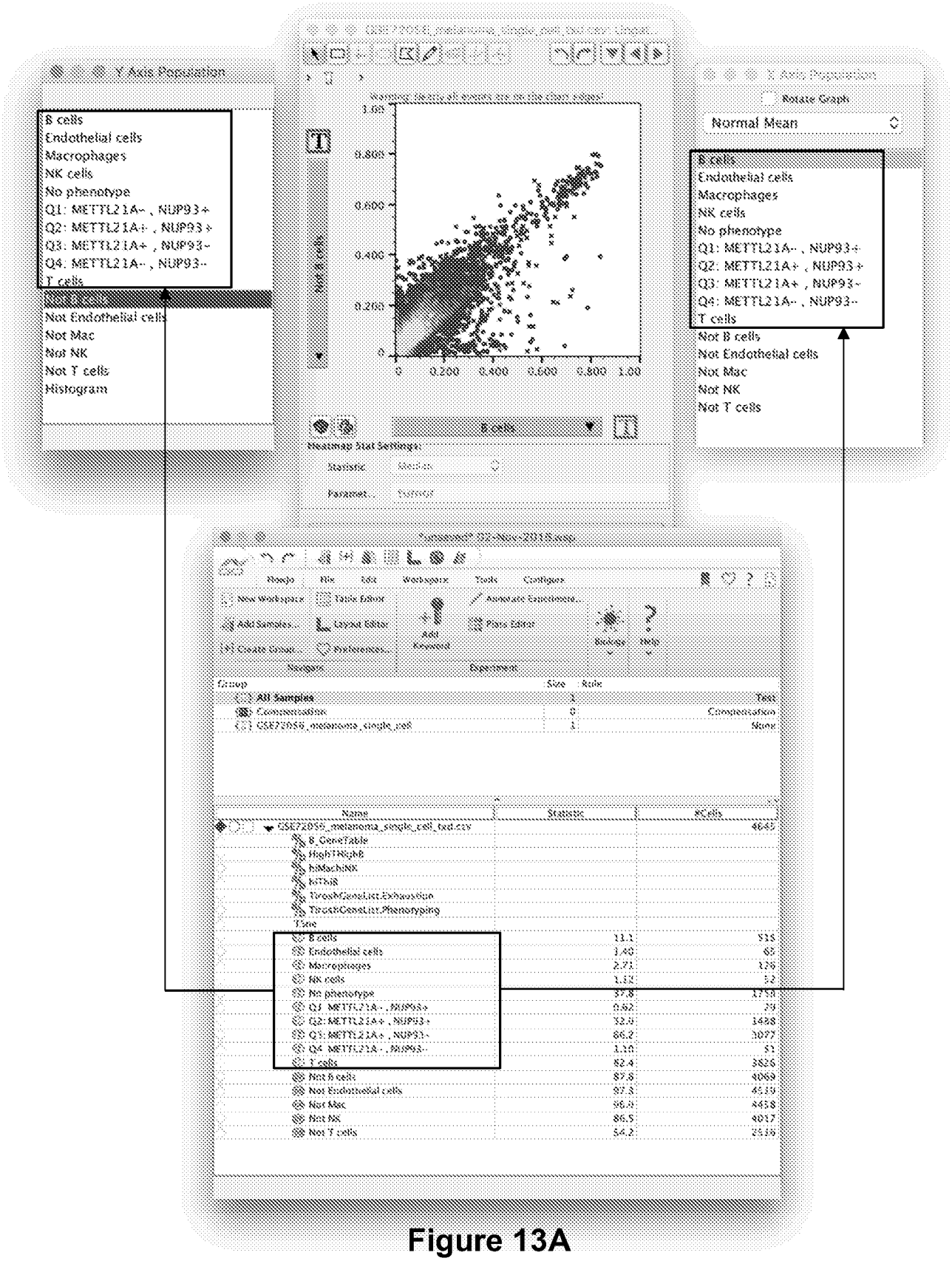

Y Axis Population
B cells
Endothelial cells
Macrophages
NK cells
No phenotype
Q1: METTL21A- , NUP93+
Q2: METTL21A+ , NUP93+
Q3: METTL21A+ , NUP93-
Q4: METTL21A- , NUP93-
T cells
Not B cells
Not Endothelial cells
Not Mac
Not NK
Not T cells
Histogram
Rotate Graph
Normal Mean
1.000
0.800
0.600
0.400
0.200
0
0.200
0.400
0.600
0.800
1.00
Heatmap Stat Settings:
Statistic
Format...
Name
Statistic
#Cells
B cells
Endothelial cells
Macrophages
NK cells
No phenotype
Q1: METTL21A- , NUP93+
Q2: METTL21A+ , NUP93+
Q3: METTL21A+ , NUP93-
Q4: METTL21A- , NUP93-
T cells
Not B cells
Not Endothelial cells
Not Mac
Not NK
Not T cells

Figure 13A

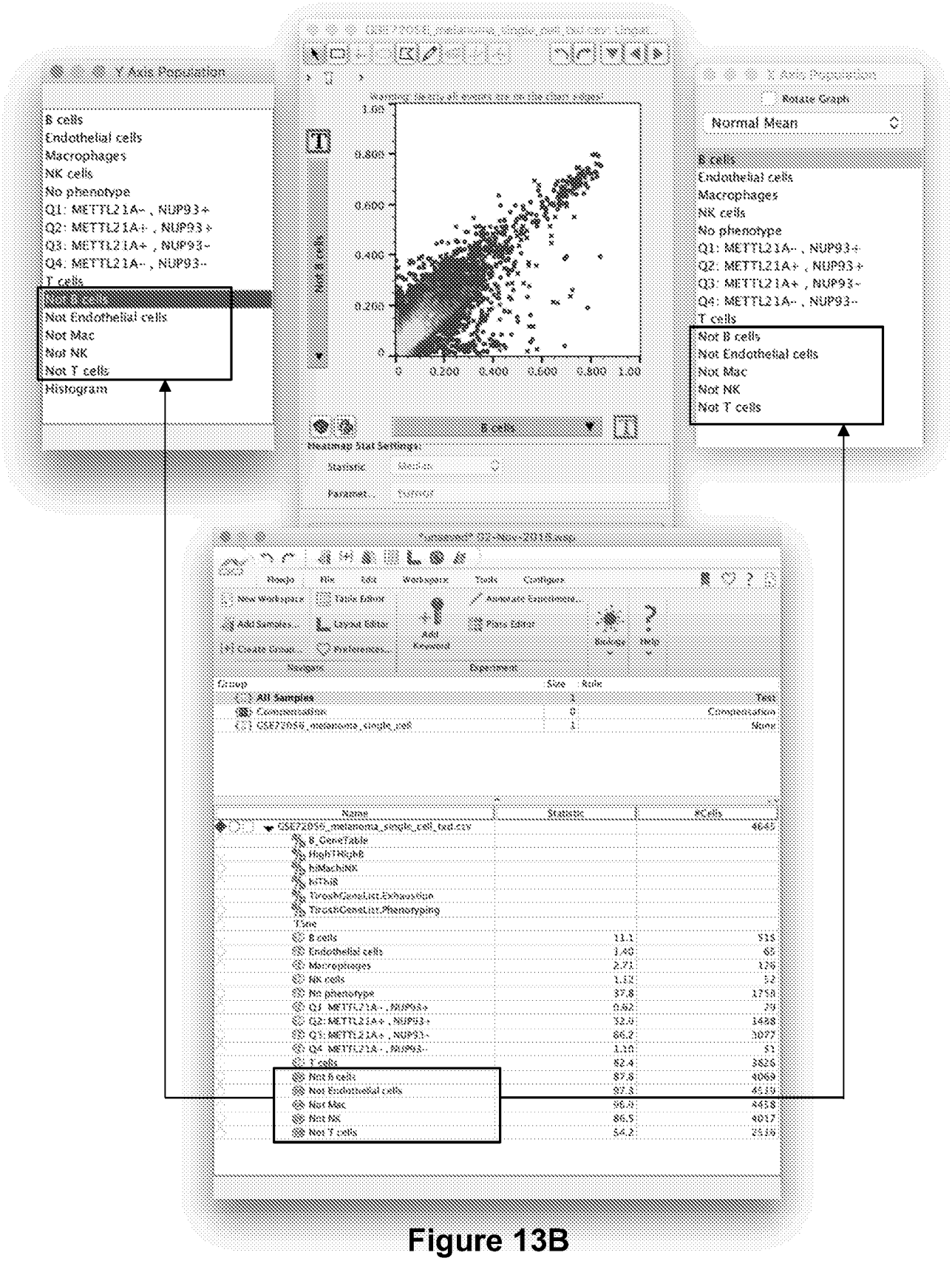

Y Axis Population
B cells
Endothelial cells
Macrophages
NK cells
No phenotype
Q1: METTL21A- , NUP93+
Q2: METTL21A+ , NUP93+
Q3: METTL21A+ , NUP93-
Q4: METTL21A- , NUP93-
T cells
Not B cells
Not Endothelial cells
Not Mac
Not NK
Not T cells
Histogram
Rotate Graph
Normal Mean
Heatmap Stat Settings:
Statistic
Format...
Name
Statistic
#Cells

Figure 13B

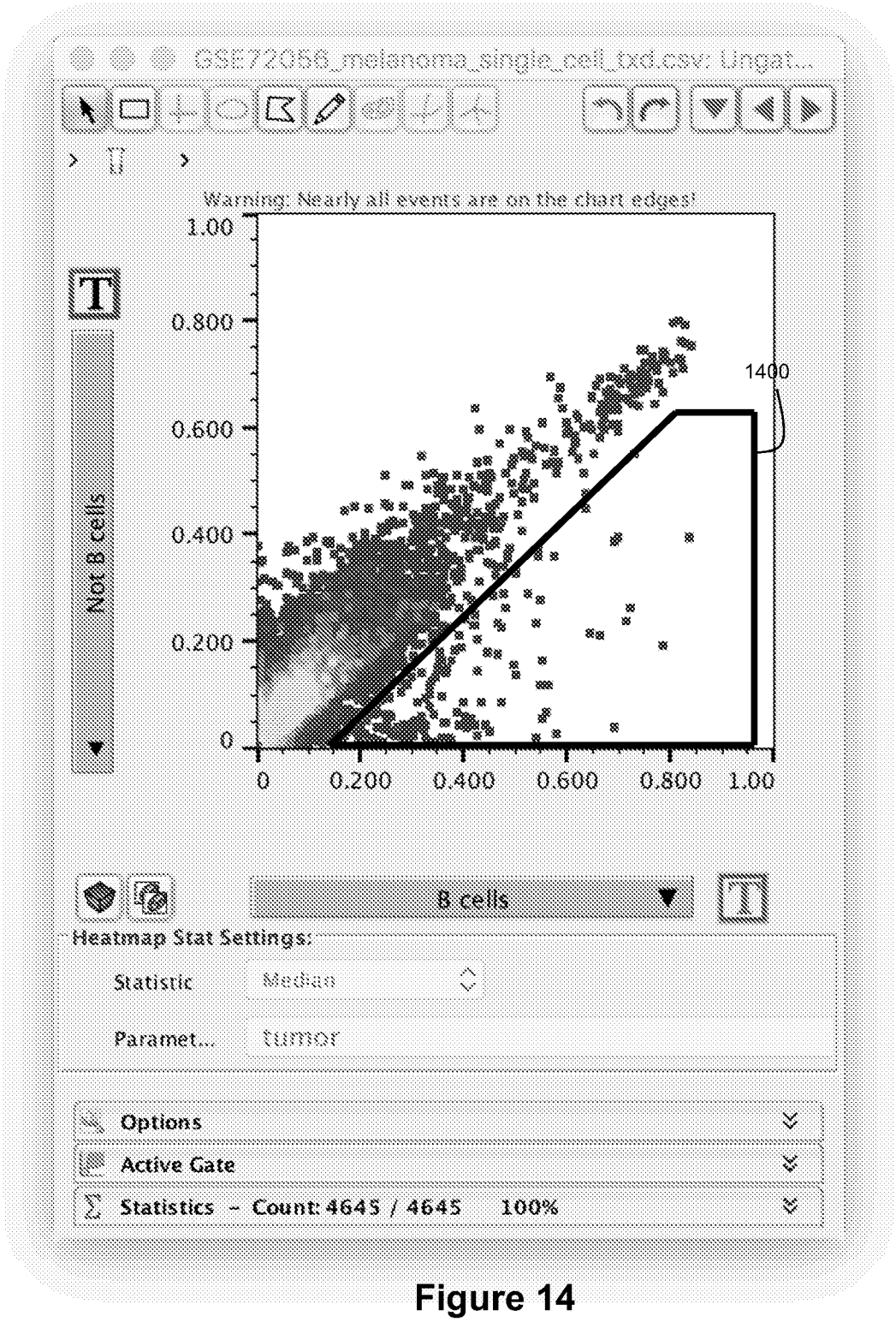
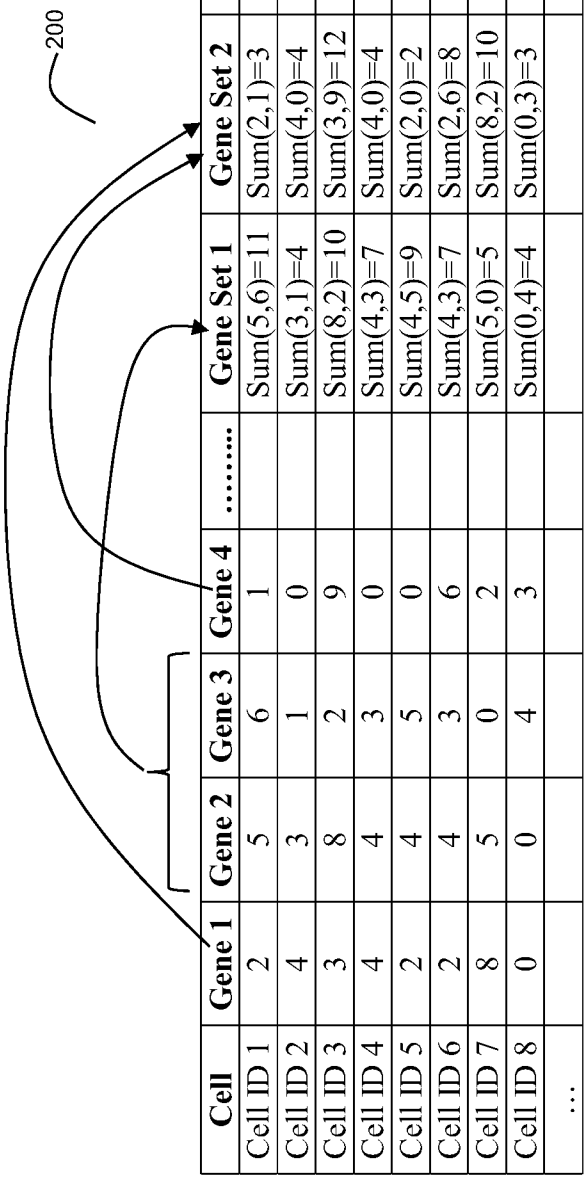

GSE72056_melanoma_single_cell_txd.csv: Ungat...
Warning: Nearly all events are on the chart edges!
1.00
0.800
0.600
0.400
0.200
0
Not B cells
1400
0
0.200
0.400
0.600
0.800
1.00
B cells
Heatmap Stat Settings:
Statistic
Median
Paramet...
tumor
Options
Active Gate
Statistics - Count: 4645 / 4645 100%

Figure 14



Cell	Gene 1	Gene 2	Gene 3	Gene 4	……..	Gene Set 1	Gene Set 2	…
Cell ID 1	2	5	6	1		Sum(5,6)=11	Sum(2,1)=3	
Cell ID 2	4	3	1	0		Sum(3,1)=4	Sum(4,0)=4	
Cell ID 3	3	8	2	9		Sum(8,2)=10	Sum(3,9)=12	
Cell ID 4	4	4	3	0		Sum(4,3)=7	Sum(4,0)=4	
Cell ID 5	2	4	5	0		Sum(4,5)=9	Sum(2,0)=2	
Cell ID 6	2	4	3	6		Sum(4,3)=7	Sum(2,6)=8	
Cell ID 7	8	5	0	2		Sum(5,0)=5	Sum(8,2)=10	
Cell ID 8	0	0	4	3		Sum(0,4)=4	Sum(0,3)=3	
…								

Figure 15

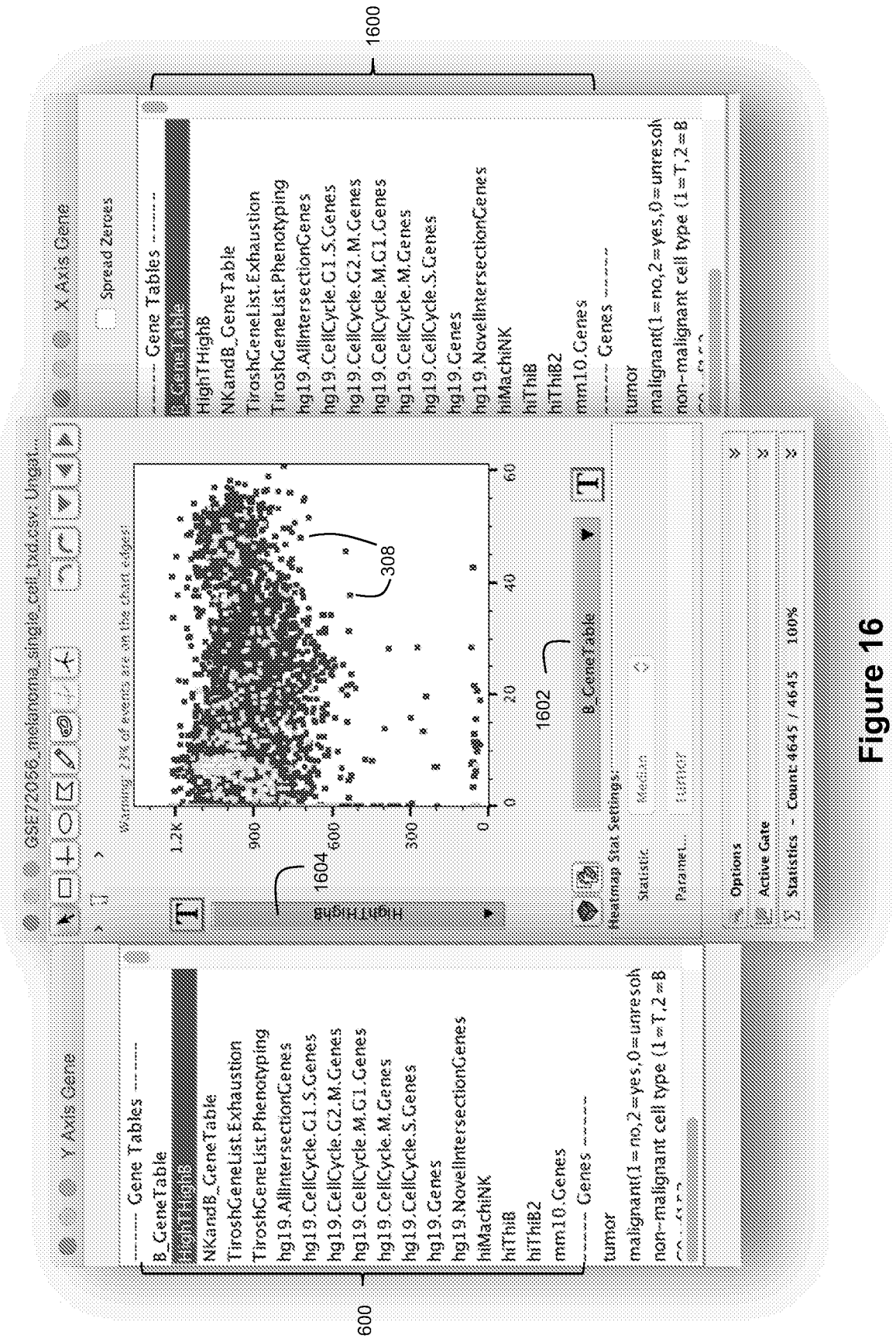

1600
1600
1604
1602
308
Y Axis Gene
X Axis Gene
Gene Tables
B_GeneTable
NKandB_GeneTable
TiroshGeneList.Exhaustion
TiroshGeneList.Phenotyping
hg19.AllIntersectionGenes
hg19.CellCycle.G1.S.Genes
hg19.CellCycle.G2.M.Genes
hg19.CellCycle.M.G1.Genes
hg19.CellCycle.M.Genes
hg19.CellCycle.S.Genes
hg19.Genes
hg19.NovelIntersectionGenes
hiMachiNK
hiThi8
hiThi82
mm10.Genes
tumor
malignant(1=no,2=yes,0=unresoh
non-malignant cell type (1=T,2=B
Spread Zeroes
HighTHighB
Heatmap Stat Settings:
Statistic
Median
Parameter
Options
Active Gate
Statistics - Count: 4645 / 4645 100%

Figure 16

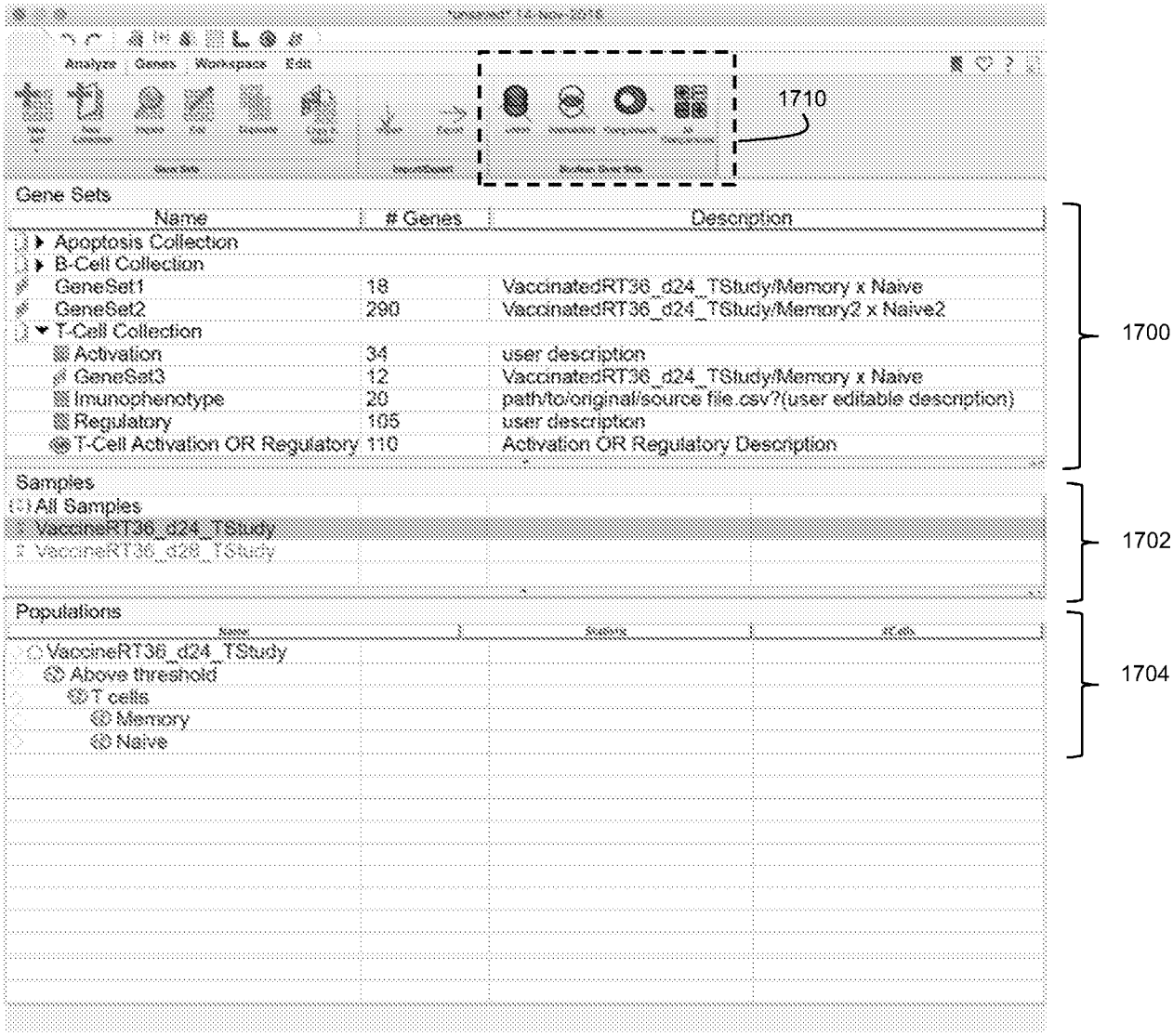


Figure 17A

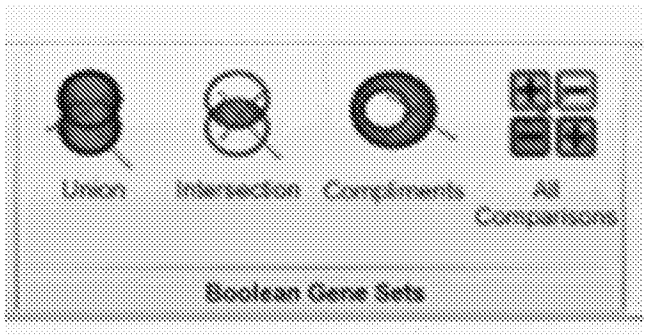


Figure 17B

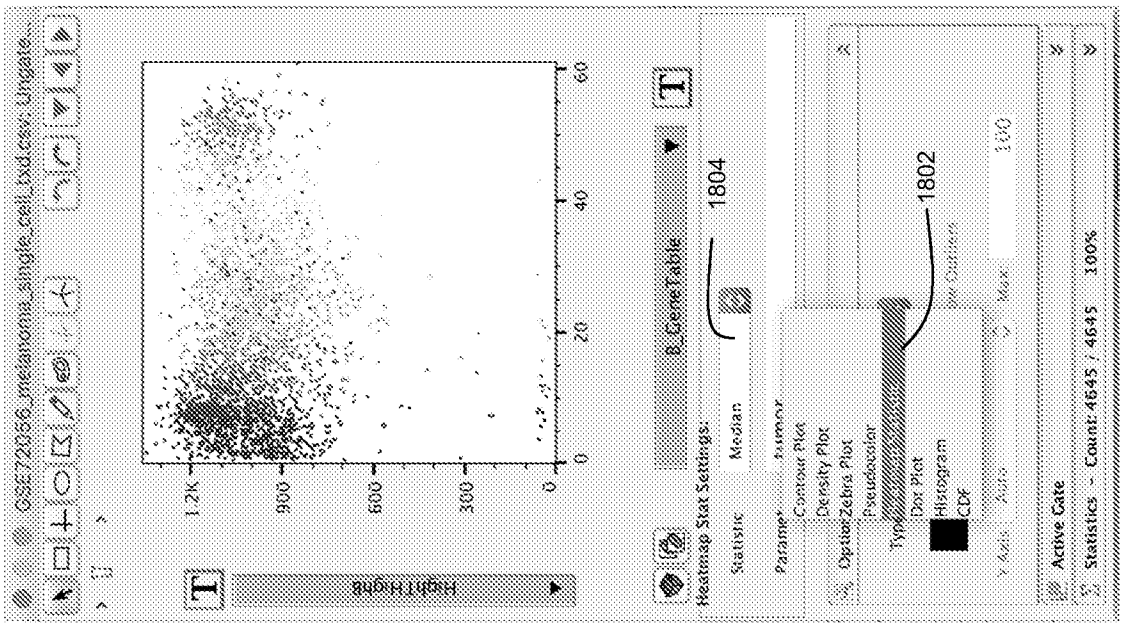


Figure 18A

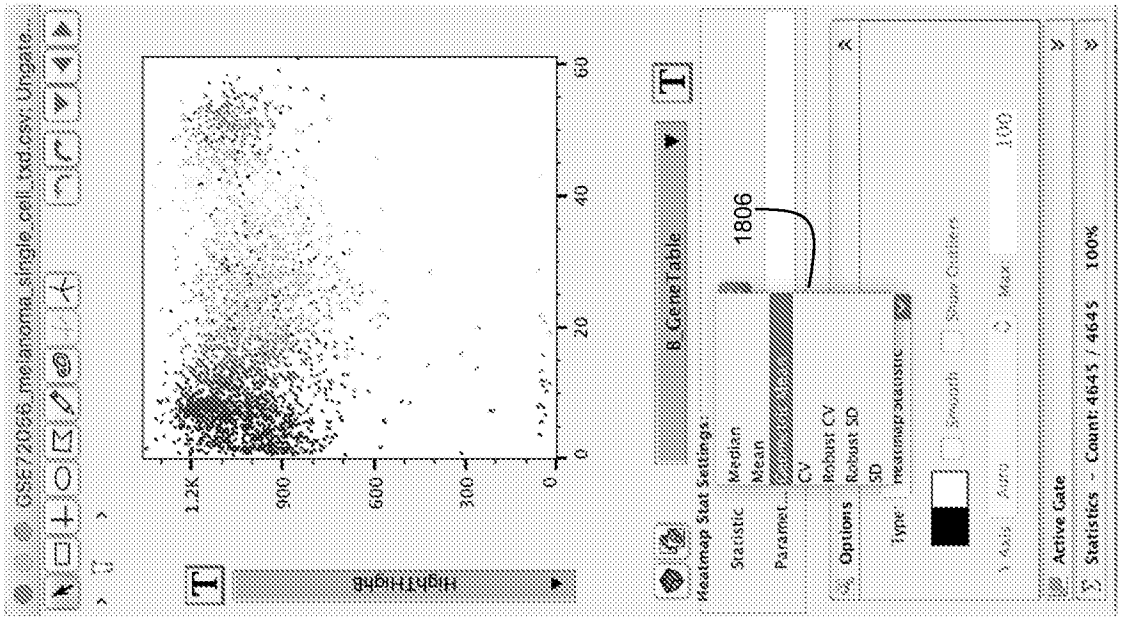


Figure 18B

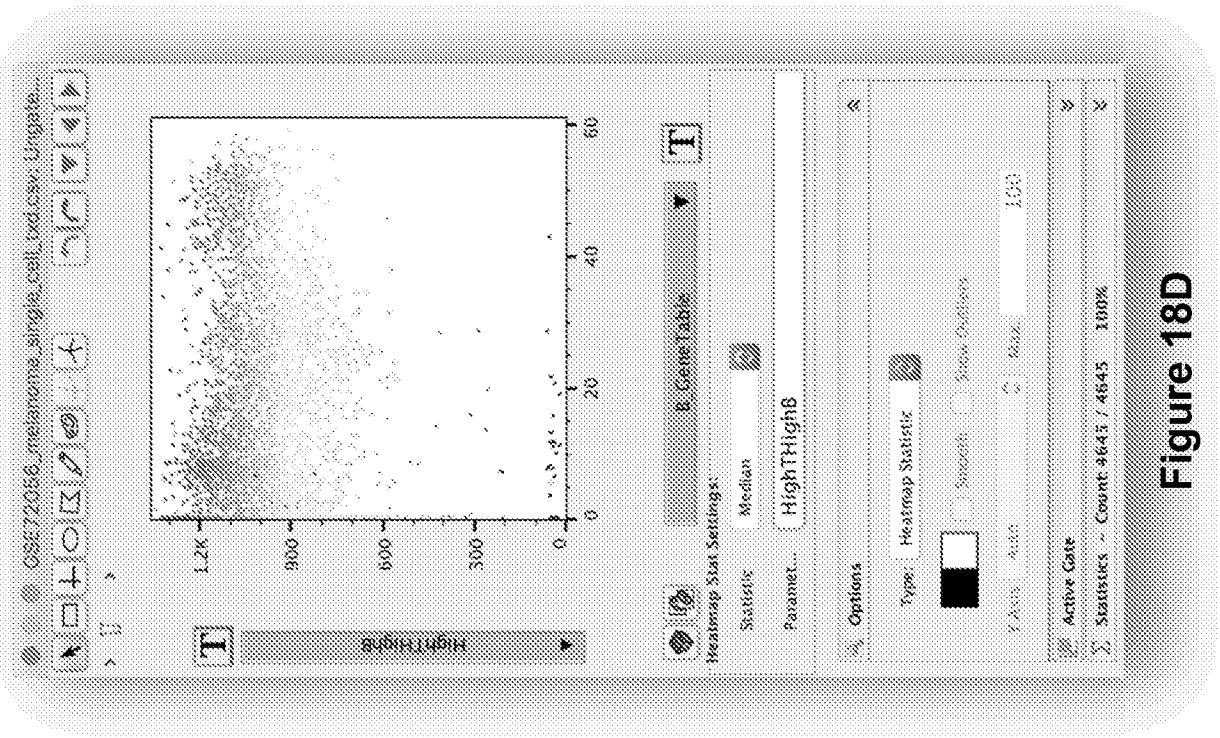


Figure 18D

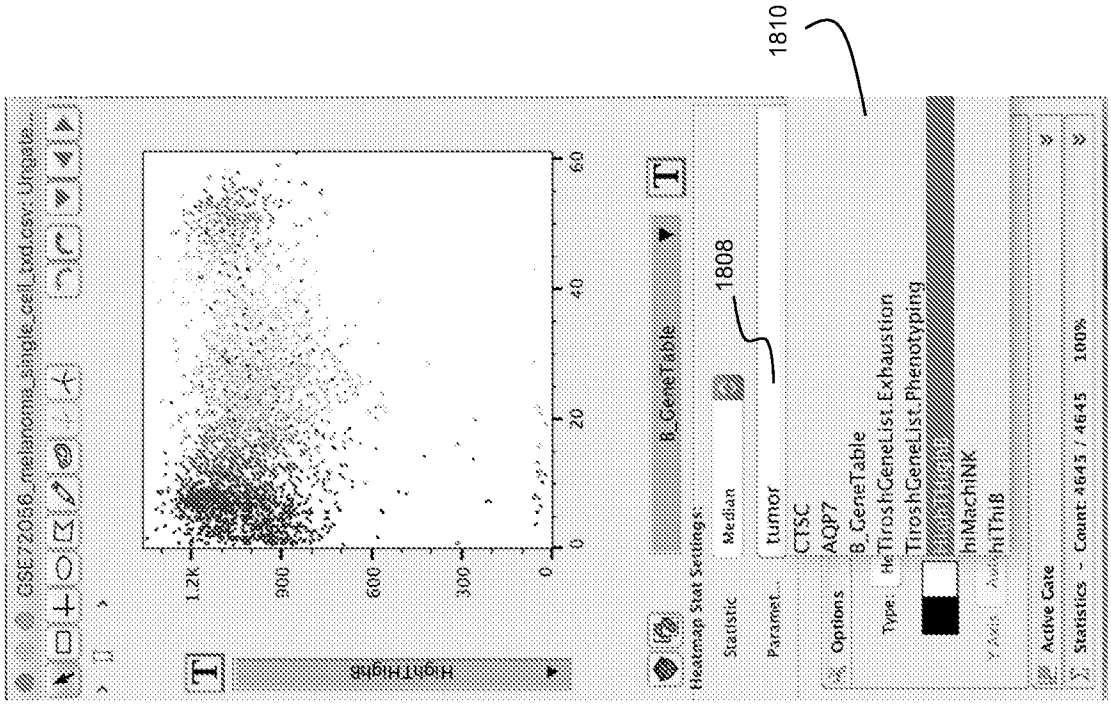


Figure 18C

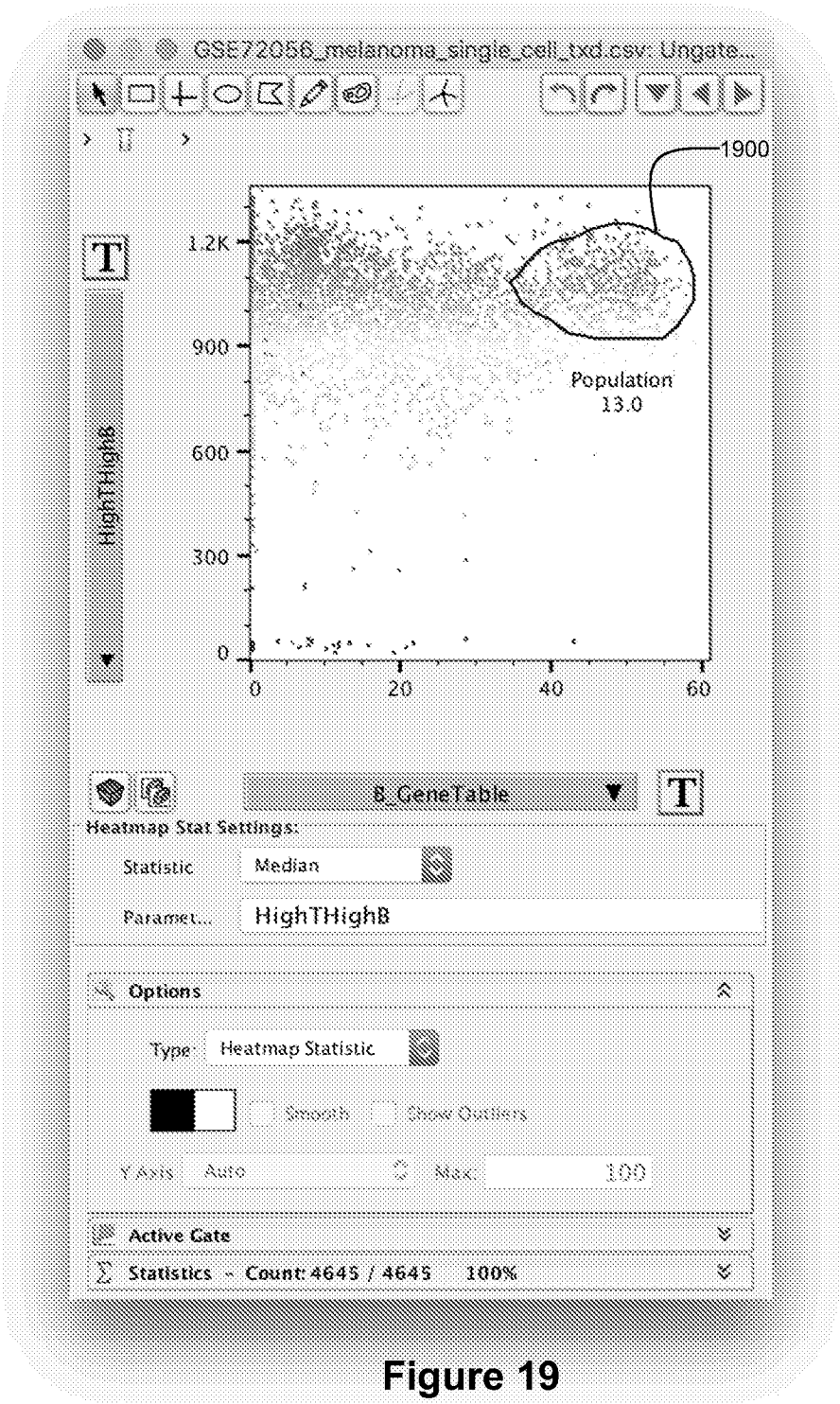

GSE72056_melanoma_single_cell_txd.csv: Ungate...
1900
1.2K
900
600
300
0
HighTHighB
Population
13.0
0
20
40
60
B_GeneTable
Heatmap Stat Settings:
Statistic
Median
Paramet...
HighTHighB
Options
Type
Heatmap Statistic
Smooth
Show Outliers
Y Axis
Auto
Max
100
Active Gate
Statistics - Count: 4645 / 4645 100%

Figure 19

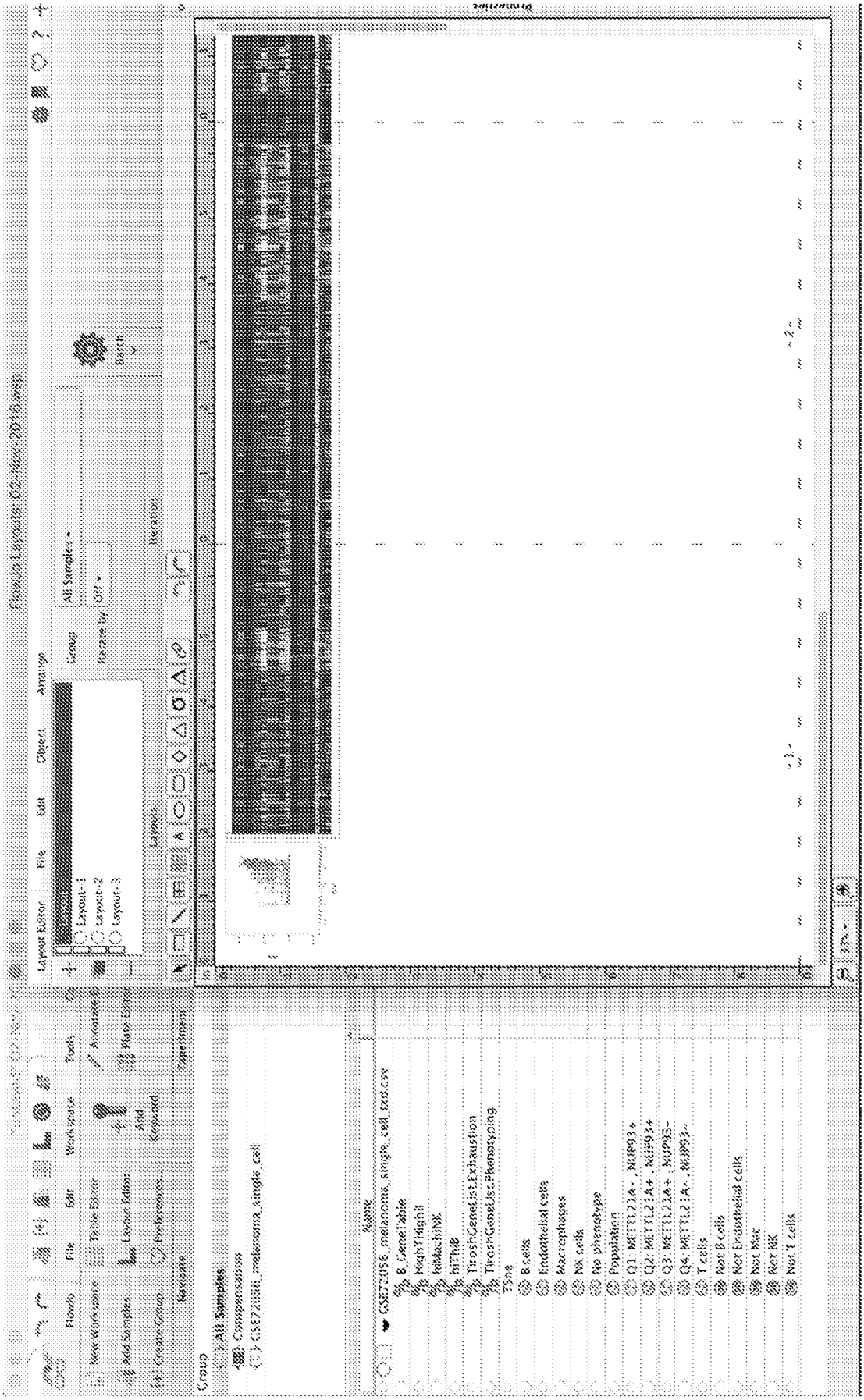


Figure 20A

Figure 20B

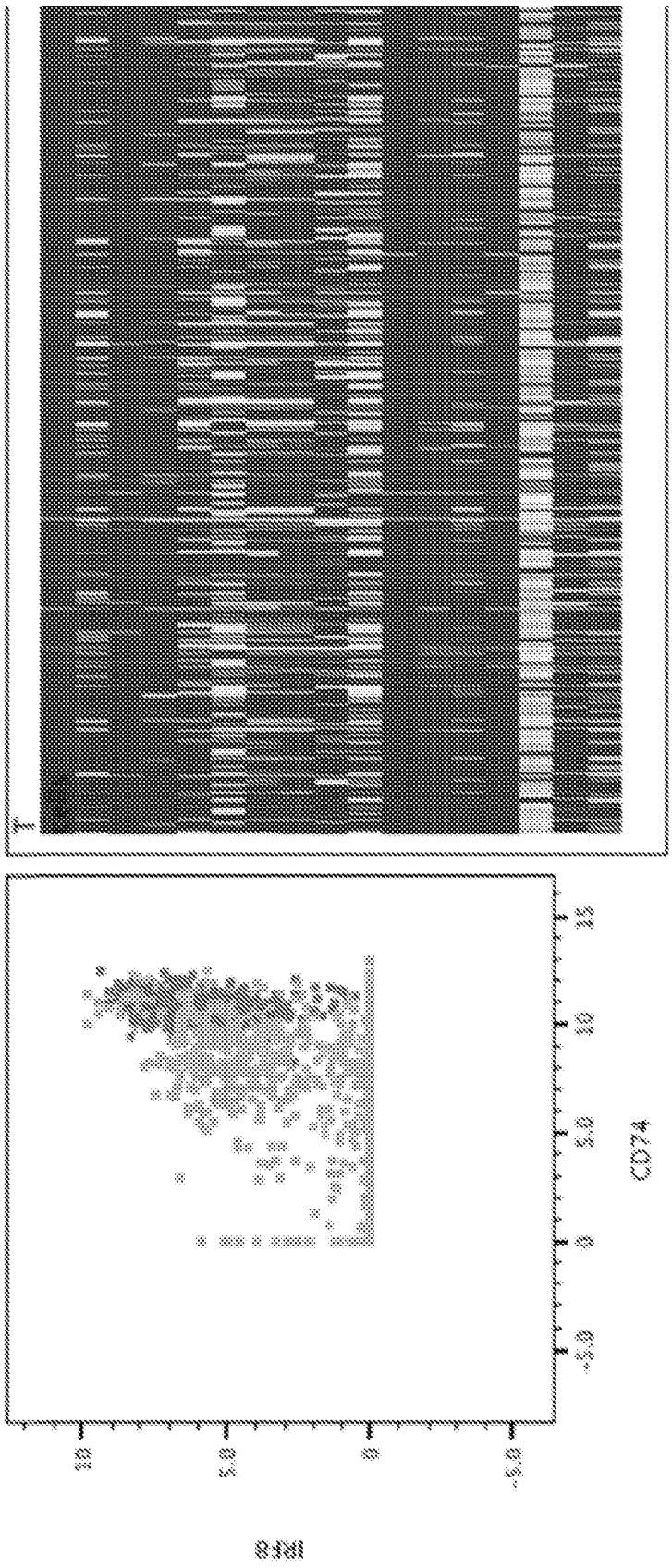


Figure 20C

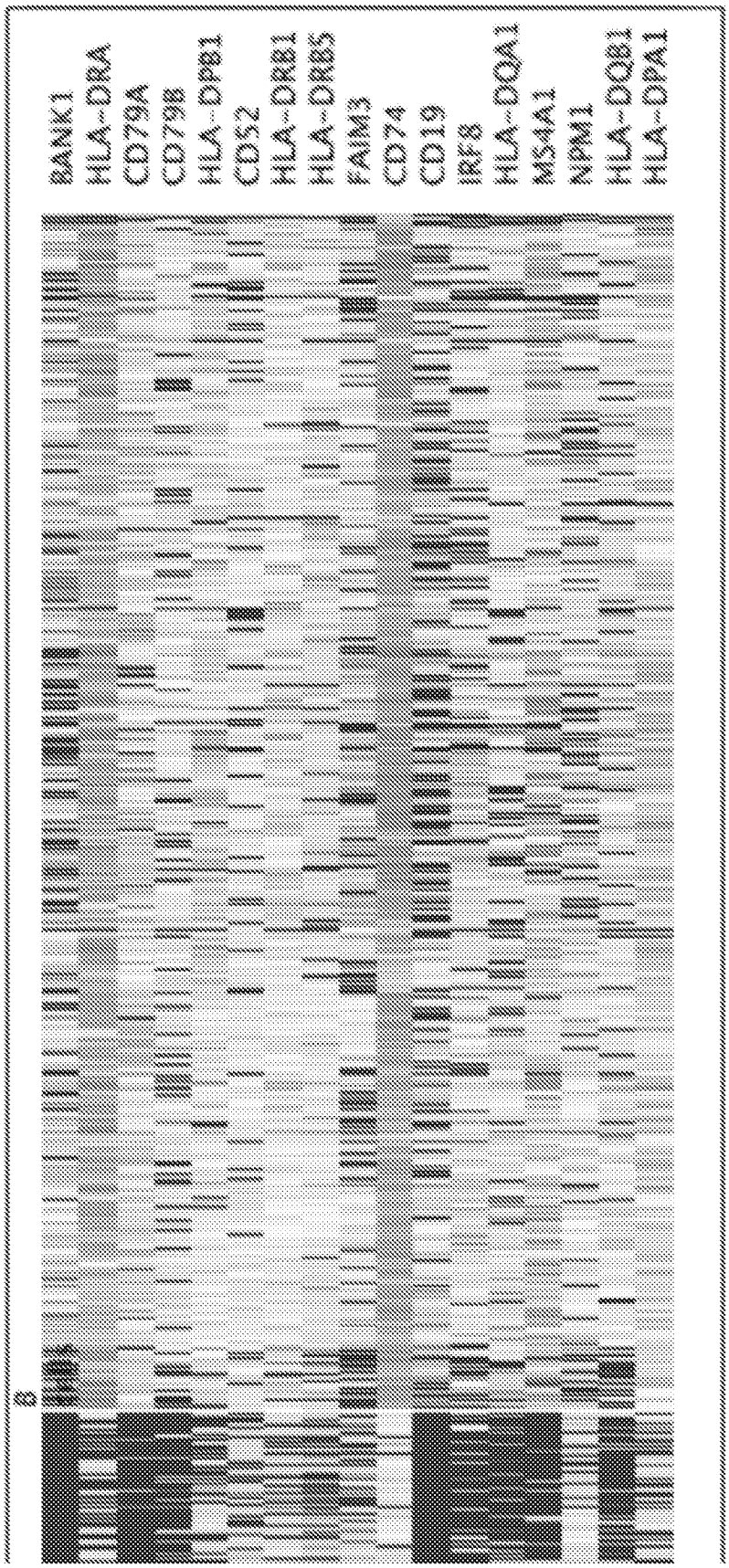

BANK1
HLA-DRA
CD79A
CD79B
HLA-DPB1
CD52
HLA-DRB1
HLA-DRB5
FAIM3
CD74
CD19
IRF8
HLA-DQA1
MS4A1
NPM1
HLA-DQB1
HLA-DPA1

Figure 20D

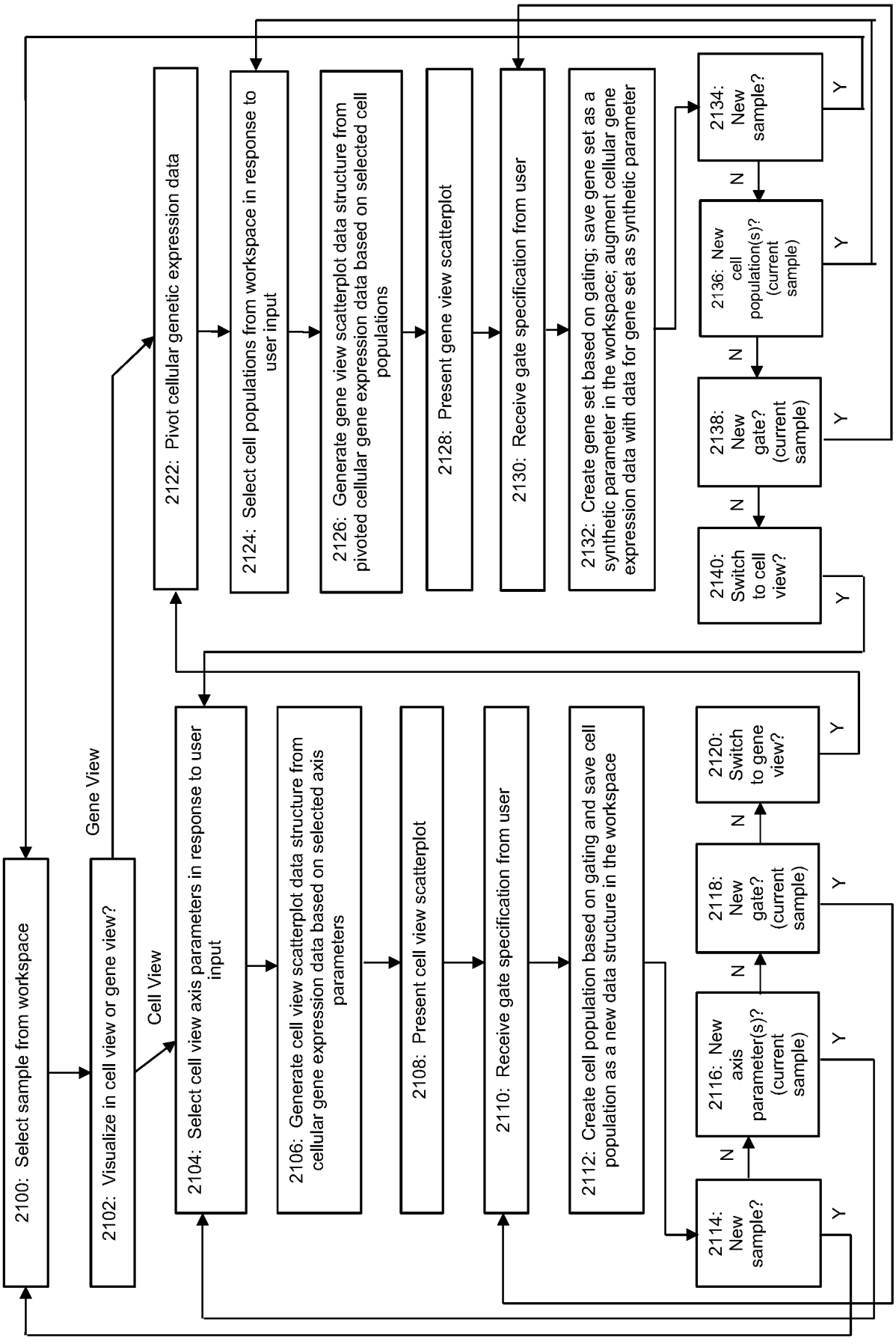

2100: Select sample from workspace
2102: Visualize in cell view or gene view?
Gene View
Cell View
2104: Select cell view axis parameters in response to user input
2106: Generate cell view scatterplot data structure from cellular gene expression data based on selected axis parameters
2108: Present cell view scatterplot
2110: Receive gate specification from user
2112: Create cell population based on gating and save cell population as a new data structure in the workspace
2114: New sample?
2116: New axis parameter(s)? (current sample)
2118: New gate? (current sample)
2120: Switch to gene view?
2122: Pivot cellular genetic expression data
2124: Select cell populations from workspace in response to user input
2126: Generate gene view scatterplot data structure from pivoted cellular gene expression data based on selected cell populations
2128: Present gene view scatterplot
2130: Receive gate specification from user
2132: Create gene set based on gating; save gene set as a synthetic parameter in the workspace; augment cellular gene expression data with data for gene set as synthetic parameter
2134: New sample?
2136: New cell population(s)? (current sample)
2138: New gate? (current sample)
2140: Switch to cell view?
Y
N

Figure 21

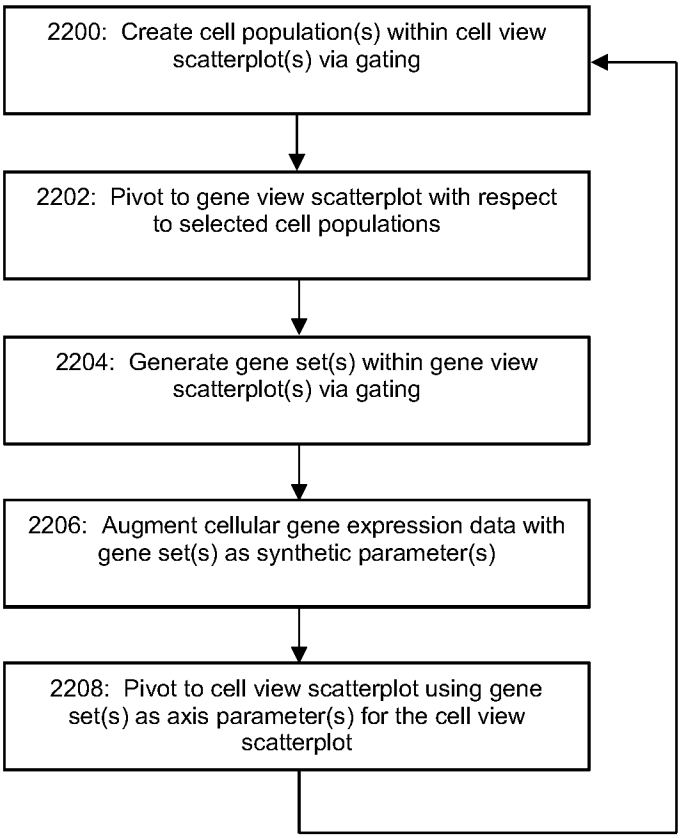

2200: Create cell population(s) within cell view scatterplot(s) via gating
2202: Pivot to gene view scatterplot with respect to selected cell populations
2204: Generate gene set(s) within gene view scatterplot(s) via gating
2206: Augment cellular gene expression data with gene set(s) as synthetic parameter(s)
2208: Pivot to cell view scatterplot using gene set(s) as axis parameter(s) for the cell view scatterplot

Figure 22

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US17/65987

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G01N 15/14, 33/53, 33/48; C12N 15/10, 5/071; C12Q 1/68 (2018.01)

CPC - G01N 15/14, 33/5308, 33/48; C12N 15/1072, 5/06; C12Q 1/6881, 1/6883

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X – Y	US 2010/0070904 A1 (ZIGON, R et al.) 18 March 2010; figure 2C; paragraphs [0039], [0048], [0058]-[0059], [0062], [0065], [0075]-[0076], [0110], [0121], [0134], [0141]	1, 3/1 --- 2, 3/2, 4/3/1-2, 22-23, 24/22-23, 43-44, 45/43-44, 47-48
Y	US 2016/0130574 A1 (MERCK SHARP AND DOHME CORP.,) 12 May 2016; abstract; figures 2A-2B, 3A; paragraphs [0006], [0013]	2, 3/2, 4/3/1-2, 22-23, 24/22-23, 43-44, 45/43-44, 47-48
Y	(BULCKE, TV et al.) SynTReN: a Generator of Synthetic Gene Expression Data for Design and Analysis of Structure Learning Algorithms. BMC Bioinformatics. 26 January 2006; Vol. 7, No. 43; pages 1-12; abstract; page 6, column 2, paragraph 4 – page 7, column 1, paragraph 1; page 8, column 1, paragraph 2; DOI: 10.1186/1471-2105-7-43	23, 24/23

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 February 2018 (09.02.2018)	23 FEB 2018
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US17/65987

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-21, 25-42, 46, 49-52
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.