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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 September 2023 (14.09.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/172923 A2

(51) International Patent Classification:
G16B 20/00 (2019.01)

(21) International Application Number:
PCT/US2023/063877

(22) International Filing Date:
07 March 2023 (07.03.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/317,636 08 March 2022 (08.03.2022) US

(71) Applicant: **BIOSKRYB GENOMICS, INC.** [US/US]; 2810 Meridian Parkway Suite 110, Durham, North Carolina 27713 (US).

(72) Inventors: **CARTER, Robert**; 2810 Meridian Parkway Suite 110, Durham, North Carolina 27713 (US). **WEIGMAN, Victor**; 2810 Meridian Parkway Suite 110, Durham, North Carolina 27713 (US). **CULLER, Corey**; 2810 Meridian Parkway Suite 110, Durham, North Carolina 27713 (US). **ZAWISTOWSKI, Jon**; 2810 Meridian Parkway Suite 110, Durham, North Carolina 27713 (US). **WEST, Jay, A.A.**; 2810 Meridian Parkway Suite 110, Durham, North Carolina 27713 (US).

(74) Agent: **REED, Sean**; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, California 94304 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

(54) Title: SYSTEMS AND METHODS RELATING TO BIOINFORMATICS

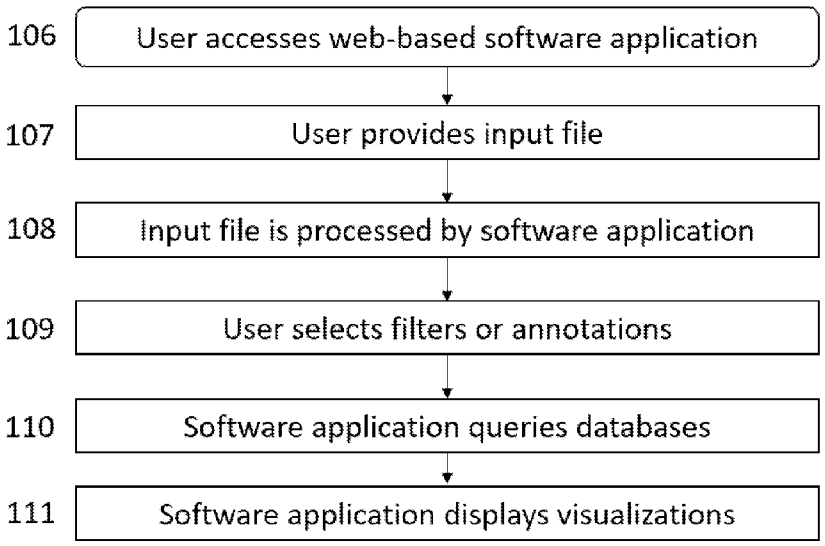


FIG. 1A

(57) Abstract: Provided herein are systems, platforms, and methods for processing and visualizing genomics data. Further provided herein are systems, platforms, and methods for rapid generation of biomarker results. Further provided herein are systems, platforms, and methods for identifying genomic correlates of diseases.

WO 2023/172923 A2

[Continued on next page]

LU, LV, MC, ME, 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).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

SYSTEMS AND METHODS RELATING TO BIOINFORMATICS

CROSS REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/317,636, filed on March 8, 2022, which is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE

[0002] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BACKGROUND

[0003] High-throughput genetic sequencing methods have provided large data sets of genetic information (such as single cell whole-genome sequencing data) for applications such as genetic engineering, agriculture optimization, and diagnosis or treatment of human disease. However, existing methods frequently suffer from slow processing or visualization speeds which may hinder analysis. A need exists for systems and methods which address these challenges.

BRIEF SUMMARY

[0004] Provided herein are systems for visualizing molecular data from biological samples.

[0005] Provided herein are systems for visualizing biological data comprising: a device comprising at least one processor and instructions executable by the at least one processor to provide a first application configured to perform operations comprising: i. accessing one or more datasets comprising biological data; ii. generating a visual representation of the one or more datasets, wherein the data is generated rapidly enough for live user interpretation. Further provided herein are systems wherein the visual representation is generated in no more than a minute. Further provided herein are systems wherein the first application comprises one or more modules. Further provided herein are systems wherein the one or more modules comprise one or more of plot, tree, cells, counting, or sequencing metrics modules. Further provided herein are systems wherein the sequencing metrics module comprises quality metrics. Further provided herein are systems wherein the biological data comprises one or more of molecular data. Further provided herein are systems wherein the molecular data comprises one or more of genomic data, proteomics data, transcriptomics data, and methylomics data. Further provided herein are

systems wherein the molecular data comprises two or more of genomic data, proteomics data, transcriptomics data, and methylomics data. Further provided herein are systems wherein the genomic data comprises one or more genomic variants. Further provided herein are systems wherein the genomic variants comprise one or more of somatic, germline, SNP, or indels. Further provided herein are systems wherein one or more datasets corresponds to a project. Further provided herein are systems wherein a project comprises a virtual collection of samples Further provided herein are systems wherein the system comprises at least 10 datasets. Further provided herein are systems wherein the system comprises at least 1000 datasets. Further provided herein are systems wherein the system comprises at least 100,000 datasets. Further provided herein are systems wherein the one or more datasets are each representative of a single cell. Further provided herein are systems wherein the one or more datasets represent at least 30 cells. Further provided herein are systems wherein the one or more datasets represent at least 100 cells. Further provided herein are systems wherein the one or more datasets represent at least 1000 cells Further provided herein are systems wherein the one or more datasets represent at least a 100,000 cells Further provided herein are systems wherein the one or more datasets comprise at least 1 biomarker. Further provided herein are systems wherein the one or more datasets comprise at least 100 biomarkers. Further provided herein are systems wherein the system comprises no more than four computer processors. Further provided herein are systems wherein the system comprises at least four processors. Further provided herein are systems wherein the system comprises at least 16 processors. Further provided herein are systems wherein the system comprises a platform for data security comprising an access control for one or more users; a security framework; and molecular data from an individual or biological species Further provided herein are systems wherein the back end module is configured to perform user authentication and registration. Further provided herein are systems wherein access to genomic data is limited to authorized users who are a part of authorized organizations. Further provided herein are systems wherein the system is accessible by at least 2 users concurrently. Further provided herein are systems wherein the system can be accessible by at least 20 users concurrently. Further provided herein are systems wherein the system can be accessible by at least 200 users concurrently. Further provided herein are systems wherein the security framework comprises an NIST cybersecurity framework. Further provided herein are systems wherein the platform complies with HIPAA standards relating to the individual. Further provided herein are systems wherein the system is configured for a user to define one or more analytical workflows. Further provided herein are systems wherein the one or more analytical workflows operate on molecular data defined in a project. Further provided herein are systems wherein the one or more analytical workflows are executed by one or more pipeline modules.

Further provided herein are systems wherein the pipeline module is configured to perform complex genomic analytical transformations. The system of claims 37, wherein the pipeline module is configured to run genomics analysis tools to extract biomarkers from one or more sequencing data. Further provided herein are systems wherein the system is configured for a user to queue and/or order pipelines to be run on a dataset. Further provided herein are systems wherein the system is configured for a user to queue and/or order pipelines to be run on a dataset that does not require manual intervention. Further provided herein are systems wherein the system comprises a pipeline module configured to identify biomarkers present in the genome. Further provided herein are systems wherein biomarkers comprise one or more of single nucleotide variants, insertions and deletions. Further provided herein are systems wherein variants across every sample of a project are aggregated to a single table. Further provided herein are systems wherein the aggregation comprises joint genotyping. Further provided herein are systems wherein the pipeline module is configured to extract annotations and variants from an input file into a database. Further provided herein are systems wherein the database comprises COSMIC, ClinVar, Ensembl, or other genomic annotation databases. Further provided herein are systems wherein the pipeline module is configured to identify structural biomarkers present in a genome. Further provided herein are systems wherein structural biomarkers comprise variations in chromosomal copy number. Further provided herein are systems wherein structural biomarkers comprises one or more of transversions, inversions, tandem repeats, translocation, and element insertion. Further provided herein are systems wherein the pipeline module is configured to identify genomic arrangement information specific to classes of genes. Further provided herein are systems wherein the genomic arrangement information comprises one or more of major histocompatibility complexes, microsatellite instability, tumor mutational burden and immune repertoire. Further provided herein are systems wherein the pipeline module is configured to one or more run transcriptome analysis tools. Further provided herein are systems wherein transcriptome analysis tools are configured to extinct counts of isoforms from one or more sequencing files. Further provided herein are systems wherein transcriptome analysis tools are configured to identify fusion events from one or more sequencing files Further provided herein are systems wherein transcriptome analysis tools are configured to identify nucleotide variation in expressed genes from one or more sequencing files. Further provided herein are systems wherein the pipeline module is configured to run methylation analysis tools from one or more sequencing files. Further provided herein are systems wherein the methylation analysis tools are configured to extract methylation status from one or more sequencing files. Further provided herein are systems wherein methylation status comprises one or more of regions of methylated cytosine, such as mC, hmC and hhmC from one

or more sequencing files. Further provided herein are systems wherein the pipeline module is configured to run proteomics analysis tools from one or more sequencing files. Further provided herein are systems wherein the proteomics analysis tools provide proteomic data. Further provided herein are systems wherein the proteomic data comprises sequence counts of identified native conformationally altered proteins. Further provided herein are systems comprising a platform for visualizing molecular data comprising: a device comprising at least one processor and instructions executable by the at least one processor, wherein the instructions comprise: a front end module comprising a graphical user interface; a security layer configured to reduce unauthorized access; a back end module; and one or more pipeline modules configured to generate a visual representation. Further provided herein are systems wherein the visual representation comprises at least one plot and genomics scale. Further provided herein are systems wherein the visual representation displays one or more biomarkers. Further provided herein are systems wherein the one or more biomarkers comprises specific molecular states associated with biological states of a sample. Further provided herein are systems wherein the system is configured for a user to select one or more applications. Further provided herein are systems wherein the one or more applications are configured for visually representing specific types of molecular data. Further provided herein are systems wherein the molecular data comprises one or more of genome, transcriptome, methylome and proteome. Further provided herein are systems wherein the graphical user interface is configured for a user to select one or more samples processed through a pipeline on the platform to be represented in a visualization application. Further provided herein are systems wherein the back end module is configured to encode data for one or more visualizations. Further provided herein are systems wherein visualization modules comprises a query engine. Further provided herein are systems wherein the query engine comprises one or more of Vaex and AG Grid. Further provided herein are systems wherein the back end module is configured to perform one or more calculations or filtering commands. Further provided herein are systems wherein one or more calculations or filtering commands are performed on biomarkers. Further provided herein are systems wherein one or more calculations or filtering commands are performed on a project or visualization application. Further provided herein are systems wherein a visualization application receives (as input) an output generated from one or more pipeline modules performed on the system. Further provided herein are systems wherein a visualization application is configured for a user to view biomarkers. Further provided herein are systems wherein the biomarkers are identified through the platform pipelines against one or more of an annotation and genomic reference tables. Further provided herein are systems wherein a visualization application comprises of one or more modules which represent one or more characteristics of a biological data type. Further

provided herein are systems wherein user is able to directly manipulate visual representations in a visualization application. Further provided herein are systems wherein the visual representation comprises one or more of feature selection, feature filtering, importing gene lists, and external biomarker results. Further provided herein are systems wherein configurations or manipulations of a visual application are exported or imported. Further provided herein are systems wherein configurations or manipulations control a layout of other visualization applications within the platform. Further provided herein are systems wherein chromosomal coordinates are leveraged to highlight various information across multiple visualization applications. Further provided herein are systems wherein biological groupings, label or group samples within a visualization module within an application. Further provided herein are systems wherein biological groupings comprises cellular phenotypes. Further provided herein are systems wherein a user is able to view biomarker information within genomic context, leveraging one or more genomic loci, along with associated annotation. Further provided herein are systems wherein the system depicts one or more modules within a visualization application for transcription. Further provided herein are systems wherein a visualization application depicts one or more of heatmaps, principal components analysis, differential expression, and alternative transcriptional mapping. Further provided herein are systems wherein a visualization application is configured for a user to view changes in copy number or structural changes within genome scale. Further provided herein are systems wherein a visualization application is configured for a user to view annotated information of changes in copy number or structural changes within the genome scale. Further provided herein are systems wherein the system is configured for syncopation of molecular data. Further provided herein are systems wherein syncopation of molecular data comprises managing and querying for one or more analysis tools. Further provided herein are systems wherein molecular data comprises data objects derived from the same cell. Further provided herein are systems wherein syncopated molecular information is visualized together on the same application within the platform. Further provided herein are systems wherein syncopated molecular information is visualized together using one or more of a Circos plots, genome viewers, and graph networks. Further provided herein are systems wherein the system is configured to associate molecular information with biological groupings. Further provided herein are systems wherein the system is configured to associate molecular information with biological groupings using one or more statistical methods. Further provided herein are systems wherein one or more statistical comprises differential expression (transcriptomic/proteomic), genome-wide association studies (genomic) and quantitative trait loci calculations using molecular information. Further provided herein are systems wherein the system is configured to export information in the system contained in one or more objects.

Further provided herein are systems wherein the one or more objects comprise files, images, pipeline and visualization configurations, biomarker lists or projects. Further provided herein are systems wherein information is exported as one more files within a project through a file explorer embedded directly on the platform. Further provided herein are systems wherein configurations of visualization applications is exported. Further provided herein are systems wherein configurations of visualization applications is exported in a format readable by the platform's visualization applications. Further provided herein are systems wherein the system is configured for export of images of visual modules or applications which preserve manipulations performed by the user. Further provided herein are systems wherein the system is configured for specific output directories. Further provided herein are systems wherein a user has rights and access to place information in the output directory. Further provided herein are systems wherein the system is configured for export of a project or pipeline to specified output locations. Provided herein are methods for performing the functions of a system described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0007] **FIG. 1A** depicts a workflow schematic of a system for input, processing, and visualization genomic data.

[0008] **FIG. 1B** depicts an architectural schematic of a system for processing and visualization genomic data.

[0009] **FIG. 1C** depicts a non-limiting example of a user interface for viewing project summaries. Headers (left to right) include Project [name], Project Size, Number of Cells, Analysis Status, Pipeline(s) performed, BioSkryb Product Lot ID, Initiating User, Date. The horizontal Ellipsis is used as a menu button for several project-related activities.

[0010] **FIG. 2A** depicts a browser for visualizing genomic data with a full set of variants in a selected region of a genome.

[0011] **FIG. 2B** depicts a browser for visualizing the genomic variation information that can be pulled from multiple visualization applications that report back genomic coordinates.

[0012] **FIG. 2C** depicts an interface exploring selected variants, including annotations. From top to bottom: Features (gene name, gene Id, gene type, strand, Tdl, and Hgnc Id); Predictions

(Left: SIFT, FATHMM, PROVEAN, MetaSVM, MetaLR), (Middle: Phylo 100-way Vertebrate) (Right: Phylo 30-way mammal); Evidence (Cosmic Genomic ID, Primary site, primary histology); and Population.

[0013] **FIG. 2D** depicts a browser for visualizing genomic data with a full set of variants in a selected region of a genome.

[0014] **FIG. 2E** depicts a browser for visualizing the genomic variation information that can be pulled from multiple visualization applications that report back genomic coordinates.

[0015] **FIG. 3A** depicts an interface for visualization of somatic mutations in the context of lineage ("tree" module). The interface provides rapid switching from plots, tree, cells, quality metrics, and more metrics visualization modules.

[0016] **FIG. 3B** depicts an interface for visualization of filtered somatic mutations ("tree" module) that occurred ancestral to one or more other data sets.

[0017] **FIG. 4A** depicts an interface for visualization of a cell population ("cells" module), including a 2 dimensional plot of cell similarity (left) and cell statistics for each individual cell (right). Statistic fields include name, number of variations by type (somatic, germline, SNPs, indels, high, medium, and low confidence). Statistics for 10 cells are shown for example only; any number of cells may be visualized.

[0018] **FIG. 4B** depicts an interface for visualization of sequencing quality control metrics ("quality metrics" module). Smaller plots shown (left to right) include chromosome M population, percent pass/fail reads aligned, WGS mean coverage, WGS PCT EXC DUPE; bottom graph: WGS coverage by sample.

[0019] **FIG. 5A** depicts a circos plot depicting variations in human genomic data. 6,416,550 somatic, 7,398,087 germline, 13,814,637 SNP, and 0 indel variations are shown.

[0020] **FIG. 5B** depicts a circos plot with a region of chromosome 3 enlarged in the inset.

[0021] **FIG. 6** depicts a plot of existing and Vaex methods for generating circus plots. The y-axis is labeled from 0 to 12,000 milliseconds at 1,000 millisecond intervals.

[0022] **FIG. 7** depicts an interface showing a circos plot generated to simultaneously visualize and compare genomic data from multiple cells (groups 1 and group 2). Filters may be applied such as threshold number of variants, viewing of specific chromosomes or cells, as well as filters regarding variant type (SNP, Indel, Germline, Somatic, Cosmic, Coding change, Clinvar) and features (gene, gene type, prediction).

[0023] **FIG. 8** depicts an expandable tree structure which enables the visualization of variable parameters (i.e. CNV, SNV, express gene, methylation site) that is compatible with filters included in the bioinformatic analysis system.

[0024] **FIG. 9** depicts a plot of run times in minutes vs. number of cells filtered (without annotations). 31 cells of WGS were processed with 4 virtual CPUs. The y-axis is labeled runtime from 0 to 3000 milliseconds. The x-axis is labeled number of cells from -5 to 35 cells.

[0025] **FIG. 10** depicts a plot of run times in minutes vs. number of cells filtered (with annotation). 31 cells of WGS were processed with 4 virtual CPUs. All single-cell WGS data, approximately 108 million rows of data, was rendered and visualized in approximately 1.2 seconds, allowing scale up to much greater numbers of cells using parallel processes. The y-axis is labeled runtime from 0 to 3000 milliseconds. The x-axis is labeled number of cells from -5 to 35 cells.

[0026] **FIG. 11** depicts a non-limiting example of a computing device; in this case, a device with one or more processors, memory, storage, and a network interface.

[0027] **FIG. 12** depicts a non-limiting example of a web/mobile application provision system; in this case, a system providing browser-based and/or native mobile user interfaces.

[0028] **FIG. 13** depicts a non-limiting example of a cloud-based web/mobile application provision system; in this case, a system comprising an elastically load balanced, auto-scaling web server and application server resources as well synchronously replicated databases.

[0029] **FIGS. 14A-14B** shows the initial user registration forms any new user will need to complete prior to gaining access to the system. **FIG. 14A** depicts a sign in or create account form (create account tab shown). **FIG. 14B** depicts a terms and conditions form.

[0030] **FIGS. 15A-15C** shows the administration console, which enables the platform administrator to onboard new users, organizations and workspaces. Company-specific details can be supplied as benefits proper tracking needed by the company. Data ingress and egress details can be provided for a workspaces to ensure proper security and data transfer information. User access within and across groups is also able to be set through the console. The organizations tab is shown in **FIG. 15A**. The edit workspace interface is shown in **FIG. 15B**. The change user group interface is shown in **FIG. 15C**.

[0031] **FIGS. 16A-16D** depicts how a user is able to create projects and select samples to include from which shared access point provided through the administration console. **FIG. 16A** depicts an interface for starting new project types, such as genomic ("Resolve DNA") or multiomic (ResolveOME) projects. **FIG. 16B** depicts an interface for creating a new project including project name, product lot ID, genome, sequencing library preparation, and BSSH/shared data options. **FIG. 16C** depicts an interface for adding samples to a new project, including a search function for basespace sequence hub project name. **FIG. 16D** depicts another example of an interface for adding samples to a new project, including a biosample name search query tool.

[0032] **FIGS. 17A-17D** depict how pipelines are customized on samples within a project and how to select configurations that a user determines. **FIG. 17A** depicts an interface prompting a user to launch a pipeline immediately after FASTQ files are downloaded. **FIG. 17B** depicts a menu of pipelines for user selection. **FIG. 17C** depicts pipeline parameters and modules for user selection. **FIG. 17D** depicts an interface displaying pipeline parameters.

[0033] **FIGS. 18A-18C** depict how a user selects secondary or tertiary analyses within a project for which they have administrative privileges. Details and tracker for the steps needed to launch a tertiary analysis are provided. **FIG. 18A** depicts an interface of samples (biosamples) with FASTQ validation status, size, total number of reads, read length, upload date, and lot ID. **FIG. 18B** depicts an interface for selection of secondary pipelines including name, version, number of biosamples, status, size, duration, initiator, and launch date. **FIG. 18C** depicts an interface for tertiary pipelines available for user selection including name, version, and a description of various pipelines.

[0034] **FIGS. 19A-19D** depict one of the visualization applications showing how variants are identified through pipelines used on a project. Mater variant tables, highlighting all alternate alleles in a project can be filtered based on numerous characteristics and downloaded. **FIG. 19A** depicts an interface for hiding filters, applying filters, and resetting filters. Parameters displayed (left side) include gene names, amino acids, chromosome position, variant quality thresholds, mapping quality, reference datasets (clinvar, cancer hallmarks, COSMIC, any state). The right side depicts a list of variants obtained after filtering, and displays chromosome, position, dbSNP ID, reference, alternate, mapping quality, clinvar, and gnomAD values. Genotypes may also be displayed per sample, genotype stats, or empty columns may be removed. **FIG. 19B** depicts the entire interface shown on the left of **FIG. 19A**, including gene names, amino acids, chromosome position, variant quality thresholds, mapping quality, reference datasets (clinvar, cancer hallmarks, COSMIC, any state), protein feature types (gene variant, intergenic region, transcript), polymorphic thresholds, and protein projection. **FIG. 19C** depicts further columns from the right side of **FIG. 19A**, including variant prevalence, genotype stats, and genotype per sample. **FIG. 19D** depicts an interface comprising a command line providing the current filter set and a history interface for storing previously used filter(s).

[0035] **FIG. 20** depicts the copy number visualization application showing multiple samples' profiles with ability to highlight genomic region of interest. These regions are represented in gene-level tables that can be filtered and downloaded.

[0036] **FIGS. 21A-21E** depict initial expression visualization application modules. Users can view and add sample phenotypes that can be used to label other modules within the application. Samples can then be identified as outliers and removed from other visualization

modules presented in the application. Differential expression can be performed on the application based on groups identified by the platform. **FIG. 21A** depicts an interface for importing or exporting visualization session settings. **FIG. 21B** depicts a visualization for comparing multiple samples. **FIG. 21C** depicts a heat map visualization of gene data, including cell types. User controlled parameters include number of genes, group annotation, show/hide group legend/sample names, sample to sample normalization (TMM or Raw), gene normalization (Raw, Log odds, or z-score), column sort order (dendrogram or groups), row sort order (+/- dendrogram), genes (top MAD genes, top genes, upload gene set, pathway). **FIG. 21D** depicts an interface for changing visualization parameters and statistical analysis for differential gene expression. Parameters include method (Deseq2 or Llmma), design formula, contrast, minimum number of counts, test (Wald Significance or likelihood ratio), fit type (parametric, local, mean, glmGamPol), false discovery rate method (Bonferroni or BH), and contrast chategories. **FIG. 21E** depicts a volcano plot visualization generated using the systems described herein.

[0037] **FIGS. 22A-22B** depict expression modules that showcase projections of samples' expression across multiple representations: PCA and UMAP. Users are able to customize facets of the projections, along with labelling samples based on sample phenotypes. **FIG. 22A** depicts a PCA quality control interface and visualization with filters for a user to select: sample to sample normalization (TMM or Raw), gene normalization (Raw, log odds, z-score), filters to show/hide sample names and size as library size, and color coding samples (by phase, progenitor, TGCA_tumor, tissue, TGCA_tissue). **FIG. 22B** depicts a UMPA interface and visualization including filters for sample to sample normalization (TMM or Raw), gene normalization (Raw, log odds, z-score), number of neighbors, minimal distance, type of metric function, and option for 2D or 3D UMAP visualization.

[0038] **FIGS. 23A-23B** depict a representation of pipeline outputs from multiple molecular types through a circos plot. User is able to customize samples, chromosomal regions and molecular data types represented on the plot. Regions on the plot are selectable and important regions can be exported to be used to filter other visualization modules on the platform. **FIG. 23A** depicts an interface for generating a circos plot. Users may apply filters to select specific samples, tracks (variants such as SNP/Indel/CNV), and chromosomes for visualization. **FIG. 23B** depicts an exemplary circos plot generated by the systems described herein.

DETAILED DESCRIPTION

[0039] Provided herein are systems and methods for processing and visualization of biological data (e.g., biomarkers). Further provided herein are systems and methods described herein result in faster processing and real-time visualization of biological data (such as a genome, transcriptome, proteome). Further provided herein are computer interfaces for visualizing and manipulating genomic data from cell populations.
The methods and systems described herein in some instances automates many of the required functions formerly requiring labor intensive processes as well dedicated personnel to curate, analyze and interpret complex genomic data.

[0040] Described herein are computer-implemented systems for visualization of biological data. In some instances, a system comprises one or more platforms. In some instances, the data comprises genomic, transcriptomic, proteomic, methylation and epigenomic data. Further described herein are computer-implemented systems comprising one or more modules. Further described herein are computer-implemented systems comprising: at least one memory storing computer-executable instructions; and at least one processor configured to access the at least one memory and execute the computer-executable instructions, wherein the computer-executable instructions comprise one or more of a frontend, a backend, and a pipeline module. In some instances, an exemplary arrangement of modules is shown in **FIG. 1B**. In some instances, modules are accessed from a cloud-based database or interface. In some instances, an exemplary workflow for genomics data visualization is shown in **FIG. 2A**. Methods and systems described herein in some instances comprise one or more steps of accessing a web-based software application; providing or otherwise linking an input file (such as a file comprising whole genomes sequencing, RNA, or other biological information); processing the file; applying one or more filters or annotations to the data in the file; querying one or more databases; and displaying a visualization of the filtered and/or annotated data. In some instances, a system comprises a device comprising at least one processor and instructions executable by the at least one processor to provide a first application configured to perform operations comprising: accessing one or more datasets comprising biological data; generating a visual representation of the one or more datasets, wherein the data is generated rapidly enough for live user interpretation. Biological data in some instances comprises molecular data. In some instances, systems are configured for the processing and visualization of one or more datasets.

[0041] Systems provided herein may comprise one or more applications. In some instances, a system comprises a first application, second application, third application, or more applications. In some instances, a first application comprises one or more modules. In some

instances, a module comprises one or more of a plot, tree, cells, counting, and sequencing metrics modules. In some instances, an application is configured to receive user parameters/data, process data, and/or visualize data.

[0042] Systems may comprise one or more datasets. In some instances, one or more datasets are associated with a project. In some instance a project comprises a plurality of samples. In some instances, a system comprises at least 1, 2, 5, 10, 20, 25, 50, 75, 100, 150, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, or at least 100,000 datasets. In some instances, a system comprises 10-100,000, 100-100,000, 500-100,000, 1000-100,000, 1000-50,000, 1000-10,000, 5000-100,000, 5000-50,000, 5000-10,000, 10,000-100,000, 10,000-50,000 or 75,000 to 500,000 datasets. In some instances one or more datasets are each representative of a single cell. In some instances, a dataset represents at least 1, 2, 5, 10, 20, 25, 50, 75, 100, 150, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, or at least 100,000 cells. In some instances, a dataset represents 10-100,000, 100-100,000, 500-100,000, 1000-100,000, 1000-50,000, 1000-10,000, 5000-100,000, 5000-50,000, 5000-10,000, 10,000-100,000, 10,000-50,000 or 75,000 to 500,000 cells. In some instances, a dataset comprises one or more biomarkers. In some instances, a dataset comprises at least 1, 2, 5, 10, 20, 25, 50, 75, 100, 150, 200, 500, 1000, 2000, 5000, 10000, 20000, 50000, or at least 100,000 biomarkers. In some instances, a dataset comprises 10-100,000, 100-100,000, 500-100,000, 1000-100,000, 1000-50,000, 1000-10,000, 5000-100,000, 5000-50,000, 5000-10,000, 10,000-100,000, 10,000-50,000 or 75,000 to 500,000 biomarkers.

[0043] The systems and methods described herein may comprise a frontend module. In some instances, the frontend module comprises a Vue.js application that provides the user interface and visualizations for the systems and methods described herein. In some instances, the frontend makes requests to the backend to query data. In some instances, a frontend comprises computer-executable instructions for one or more of: displays complex visualizations such as the circos plot, phylogenic tree, etc. (e.g., as navigable tabs); displays quality metrics; visualizes filters and filtering interactions; and presents data tables for cell information. In some instances, a web version of IGV is integrated into the frontend.

[0044] The systems and methods described herein may comprise a backend module. In some instances, the backend comprises a Flask framework application and provides one or more backend features of for the methods and systems described herein. In some instances, the backend is written in Python. In some instances, a backend comprises computer-executable instructions for one or more of: user authentication and registration; data computations and filtering; access of a Vaex open-source library for speeding up data interactions; interacting with a database and HDF5 files to process data requests; presenting and encoding data for visualizations; and presenting data for IGV.

[0045] A user may define one or more analytical workflows to the system to perform. After a use selects one or more analytical workflows, the system in some instances executes the analytical workflow using one or more pipelines.

[0046] The systems and methods described herein may comprise a pipeline module. In some instances, the pipeline comprises a computationally-intensive workflow that runs genomics analysis tools to extract signatures of biomarkers from sequencing files and loads them into a database. In some instances, the methods and systems described herein comprise one or more pipeline modules. In some instances, pipeline modules comprise multi-omics, such as WGS/exome, methylation, proteome, proteome bacterial, or RNA-seq/transcriptome. In some instances, pipeline comprises one or more sub-modules. In some instances, a pipeline comprises one or more data files. In some instances, a pipeline comprises one or more of sequencing input files, sub-pipeline modules, and summary files.

[0047] Pipelines may be configured for whole genome or exome sequencing data. In some instances, a WGS/exome pipeline is configured to input one or more FASTQ files. In some instances, a WGS/exome pipeline comprises one or more of alignment, haplotype caller, joint-genotyping, heterozygous site detector (Pipeline used for the analysis of cell lines without *a priori* knowledge of reference heterozygous variant sites), statistics, ADO, and CNV are needed to drive insights from sequencing data. In some instances the files contain sequence(ing) information/data. In some instances, files comprise sequence data from the clusters that pass filter on a flow cell. In some instances the files comprise FASTQ files. In some instances the database comprises a PostgreSQL database. In some instances, the databases are accessed from a backend module. rises computer-executable instructions for one or more of: accepts a sequencing information file as input (e.g., FASTQ); running joint genotyping to produce VCF file and linking variants to COSMIC, ClinVar, or other variant list. In some instances a VCF file contains the variants called from multiple samples (cells) all together and represent high confidence variants distributed across the cells. These variants in some instances represent changes in nucleotides observed in a cell in relation to the reference genome. In some instances, these variants are placed along the genome using genomic coordinates (e.g. chr1 base 18903). Such a configuration having a specific location for a variant allows in some instances association of information complied in databases to this given variant.

[0048] Pipelines may be configured for multi-omics analysis. In some instances, multi-omics comprises two or more types of biological information (or biological data). In some instances biological data comprises molecular data. In some instances, molecular data comprises two or more of transcript (transcriptome), genomic, proteomic, methylome, or other form of sample analysis. In some instances, molecular data comprises three or more of transcript

(transcriptome), genomic, proteomic, methylome, or other form of sample analysis. In some instances, molecular data comprises four or more of transcript (transcriptome), genomic, proteomic, methylome, or other form of sample analysis. In some instances, methods described herein display and/process multi-omics data. Molecular data in some instances is obtained from a single cell. Molecular data in other instances is obtained by evaluation of a population of cells. In some instances, methods described herein display transcript and genomic data. In some instances, methods described herein utilize transcript, genomic data, and proteomics data. In some instances, methods described herein utilize transcript, genomic data, and methylome data. In some instances, a pipeline is configured to identify one or more biomarkers (e.g., those present in multiomic data).

[0049] Pipeline modules in a system may be configured to perform complex genomic analytical transformations. In some instances, a pipeline module is configured to run genomics analysis tools to extract biomarkers from one or more sequencing data. In some instances, the system is configured for a user to queue and/or order pipelines to be run on a dataset. In some instances, a system is configured for a user to queue and/or order pipelines to be run on a dataset that does not require manual intervention. In some instances, a pipeline module is configured to identify biomarkers present in the genome. In some instances biomarkers comprise one or more of single nucleotide variants, insertions and deletions. In some instances variants across every sample of a project are aggregated to a single table. In some instances the aggregation comprises joint genotyping. In some instances the pipeline module is configured to extract annotations and variants from an input file into a database. In some instances a database comprises COSMIC, ClinVar, Ensembl, or other genomic annotation databases. In some instances a pipeline module is configured to identify structural biomarkers present in a genome. In some instances structural biomarkers comprise variations in chromosomal copy number. In some instances structural biomarkers comprises one or more of transversions, inversions, tandem repeats, translocation, and element insertion. In some instances a pipeline module is configured to identify genomic arrangement information specific to classes of genes. In some instances a genomic arrangement information comprises one or more of major histocompatibility complexes, microsatellite instability, tumor mutational burden and immune repertoire. In some instances pipeline module is configured to one or more run transcriptome analysis tools. In some instances transcriptome analysis tools are configured to extinct counts of isoforms from one or more sequencing files. In some instances transcriptome analysis tools are configured to identify fusion events from one or more sequencing files. In some instances transcriptome analysis tools are configured to identify nucleotide variation in expressed genes from one or more sequencing files. In some instances the pipeline module is configured to run methylation analysis tools from one or more sequencing

files. In some instances the methylation analysis tools are configured to extract methylation status from one or more sequencing files. In some instances methylation status comprises one or more of regions of methylated cytosine, such as mC, hmC and hhmC from one or more sequencing files. In some instances a pipeline module is configured to run proteomics analysis tools from one or more sequencing files. In some instances a proteomics analysis tools provide proteomic data. In some instances a proteomic data comprises sequence counts of identified native conformationally altered proteins.

[0050] In some instances an alignment pipeline comprises one or more of a compressed alignment file describing the alignment information of the reads in the project against a given reference (e.g. hg38), a .bam file) and an index file of the alignment file). In some instances, the pipeline comprises a .bam file.

[0051] In some instances a haplotype caller pipeline comprises one or more of a genomic variant call format (GVCF) file containing the detected variants for a given sample) and an indexer file associated with the GVCF file.

[0052] In some instances a joint-genotyping pipeline comprises one or more of a genomic variant call format (GVCF) file containing the joint variant calling of multiple samples) and an indexer file associated with the Joint-Genotyped GVCF file.

[0053] In some instances a heterozygous site detector pipeline comprises one or more of a genomic variant call format (GVCF) file containing the called variants with high degree of prevalence across a dataset and high confidence; and an indexer file associated with the GVCF file.

[0054] In some instances a statistics pipeline comprises one or more of a tabulator-separated value table describing whole genome sequence (WGS) level statistics estimated from the aligned reads (e.g. 1X, 5X, 10X coverage, etc.); and a tabulator-separated value table showing exome-panel specific statistics (e.g., On, OFF, Near target events).

[0055] In some instances a ADO pipeline comprises one or more of a tabulator-separated value table showing allele frequencies of N number of queried heterozygous sites. This table is in some instances used to estimate WGS allele balance.

[0056] In some instances a CNV pipeline comprises one or more of a tabulator-separated value table describing, for a sample, the estimated copy number for bins of size N across the whole genome; and tabulator-separated value table describing, for a sample, the type of event (insertion, deletion) for all bins of size N across the genome.

[0057] Pipelines may be configured for bacterial sequencing data. In some instances, a bacterial pipeline is configured to input a FASTQ file. In some instances, a bacterial pipeline comprises one or more of: a compressed FASTQ files containing trimmed and filtered high

quality sequences; a tabulator-separated value table describing taxonomic assignation of each read to a given species using a database, such as Kraken's database); a fasta file describing the genome assembly, at the level of contigs, constructed from the reads in the dataset; fasta file describing the genome-assembly, at the level of scaffolds, constructed from the reads in the dataset; a BAM file describing the alignments of the reads in reference to the assemble genome (e.g., contigs). In some instances, a bacterial pipeline comprises one or more summary files. In some instances, summary files comprise one or more of: a Tabulator-separated value table describing the taxonomic assignment of contigs in an assembly based on the proportion of reads mapped to them; a tabulator-separated value table showing the estimated completeness of a given assembly based on a set of phylogenetic marker genes.

[0058] Pipelines may be configured for RNA-seq data. In some instances, an RNA-seq pipeline is configured to accept one or more of a compressed alignment file describing the alignment information of the reads in the project against a given reference (e.g. hg38); an index file of the compressed alignment file; a compressed alignment file describing the alignment information of the reads in the project against a RNA-Seq specific index for a given reference and an index file for the alignment file. In some instances, an RNA-seq pipeline comprises one or more summary files. In some instances, summary files comprise one or more of a tabulator-separated table describing the matrix of counts of the genomic features (e.g. exons in a gene) across samples; a tabulator-separated table describing the number of unique splice-junction overlaps; a tabulator-separated table describing overall alignment metrics (e.g. number of genes with counts, etc.); and a tabulator-separated table showing the estimated ratio of exon-non exon alignment events.

[0059] Systems and methods described herein may comprise filters for visualizing data. In some instances, filters comprise one or more of: Germline mutation, Somatic mutation, Copy number variation, Single nucleotide variation, Insertions and deletions, Tumor Mutation Burden (TMB) Analysis, Catalog of somatic mutation in cancer (Cosmic)4, ClinVar, and Predicted Coding Change.

[0060] Further described herein are computer-implemented systems comprising: at least one memory storing computer-executable instructions; and at least one processor configured to access the at least one memory and execute the computer-executable instructions to perform: receiving a query, wherein the query comprises genomic data from one or more samples; querying a database; wherein the database comprises a plurality of genomic data and a plurality of phenotype data; generating, using at least the genomic data, a genome summary, the genome summary comprising genes and gene variants of the cohort; determining a graphical

representation of the genome summary; and sending the graphical representation to a display device.

[0061] Also described herein are computer-implemented systems comprising: at least one memory storing computer-executable instructions; and at least one processor configured to access the at least one memory and execute the computer-executable instructions to generate a graphical user interface (GUI) that accepts a query from a user comprising sequencing information from one or more samples and presents to the user genome information (**FIG. 1C**). In some instances, a GUI comprises a project browser or dashboard such as **100**. In some instances, a GUI comprises search menus for one or more of project **102 or LotID** or sorting functions that can rank size **104** initiator **105**, analysis date **106**. In some instances a GUI comprises a list of previous and current projects **102**. In some instances, projects can have a menu **107** that enables users to modify name, add sample collections, export to delivery server or remove from platform. In some instances, projects and data are shared among a group of users across workspaces or organizations **108**. In some instances, GUI is facilitated by a frontend. In some instances, the front end enables operators or queries to a backend to reflect actions of a user to manipulate and transform information.

[0062] Computer-implemented systems may comprise a genome browser. In some instances a genome browser is configured to display sections of a genome and/or variants (**FIGS. 2A-2B**). In some instances, a genome browser comprises an IGV (integrated genome viewer). In some instances, in the IGV window the bin size is selectable from the entire genome down to the individual base. In some cases a user can request specific samples to be viewed that were analyzed within a project. In some cases a user can query positions by providing dbSNP ids. In this view individual mutations in some instances are viewed to determine the alternative allele or base change. In some instances, each mutation is selectable, further detailing the nature of the modification and presenting it to the user. In some cases, amino acid information is conveyed for user to gather information about potential impact to coding sequence.

[0063] Computer-implemented systems may comprise an interface for annotating variants. This is an important step to empower interpretation of downstream coding changes in protein structure and function. Variant information in some instances comprises one or more of features (name, gene id, gene type, strand, TdI, HgncId), predictions (SIFT/sorting intolerant from tolerant, LFT / likelihood ratio test, FATHMM, PROVEAN/ protein variation effect analyzer, MetaSVM, MetaLR), conservation among species (e.g., vertebrates, mammals, etc.); evidence (pathology-related data from databases such as COSMIC), and biological population (**FIG. 2C**). In some instances, a variant annotation interface assesses the degree of conservation among (100) vertabrates and (30) mammals. In some instances this display is helpful in the

investigation of de-novo variant alleles which are not annotated by ClinVar, Cosmic, Genecards or Ensembl. The comparison allows the determination of conservation of alleles found in the sample compared to the same allele found in an alternative species. Conserved alleles are right shifts, where the conservation is high, where alleles which have low conservation are shifted left. As an example, in the Phylo 30-way mammal plot the allele is highly conserved across all 30 mammals indicating the gene is highly conserved and likely to be important for the health of all mammalian species. Having assessed the potential for the mutation to be pathogenic, if annotated the user in some instances navigates to a variety of external databases (e.g. GeneCards, Ensembl, Clinvar and COSMIC) by simply selecting the hyperlink for that specific database.

[0064] Variants in some instances are annotated as one or more of Germline mutation, Somatic mutation, Copy number variation (CNV), Single nucleotide variation (SNV), Insertions and deletions, Catalog of somatic mutation in cancer (Cosmic), ClinVar, and Predicted Coding Change. Additional resources are also accessed in some instances, such as GeneCards, Essembl, CinVar and Cosmic. In some instances, variants comprise complex markers such as those obtained using Tumor Mutation Burden (TMB) Analysis.

[0065] Computer-implemented systems may comprise an interface for tracing variant lineages. In some instances, lineages comprise somatic, ancestral, or reference lineages. Lineage trees in some instances are generated from specific chromosomes, and graphically display variants in a chart format (**FIGS. 3A-3B**).

[0066] Computer-implemented systems may comprise an interface for analyzing cells. In some instances, samples comprise one or more cells. Cells in some instances are searched, or summary information about each cell is displayed such as cell name, variants detected (somatic, germline, SNPs, and indels (**FIG. 4A**). In some instances, metrics high, medium, and low are used to describe confidence of variant calls for each cell. In some instances, inter-cell distances are graphed.

[0067] Computer-implemented systems may comprise an interface for visualizing sequencing metrics (e.g., Picard metrics) (**FIG. 4B**). Metrics include but are not limited to chromosome M population, percent pass/fail reads aligned, WGS mean coverage, and WGS percent excluded duplicate reads. Each metric in some instances is also displayed on an individual per-cell basis.

[0068] Computer-implemented systems (platforms) may comprise an interface for visualizing genomic data. In some instances, a platform for visualizing molecular data comprises one or more of a device comprising at least one processor and instructions executable by the at least one processor, wherein the instructions comprise: a front end module comprising a

graphical user interface; a security layer configured to reduce unauthorized access; a back end module; and one or more pipeline modules. In some instances a visualization platform comprises one or more visualization applications. In some instances one or more visualization applications are configured to execute one or more pipelines (e.g., visualization pipelines). In some instances visualization applications are controlled by a user.

[0069] A visualization system (platform) may be configured to generate a visual representation of a dataset. In some instances, a visual representation comprises at least one plot and genomics scale. In some instances, a visual representation displays one or more biomarkers. In some instances, one or more biomarkers comprise specific molecular states associated with biological states of a sample. In some instances, a system is configured for a user to select one or more applications. In some instances, In some instances, one or more applications are configured for visually representing specific types of molecular data. In some instances, molecular data comprises one or more of genome, transcriptome, methylome and proteome. In some instances, a graphical user interface is configured for a user to select one or more samples processed through a pipeline on the platform to be represented in a visualization application. In some instances, a back end module is configured to encode data for one or more visualizations. In some instances, visualization modules comprises a query engine. In some instances, a query engine comprises one or more of Vaex and AG Grid. In some instances, a back end module is configured to perform one or more calculations or filtering commands. In some instances, one or more calculations or filtering commands are performed on biomarkers. In some instances, one or more calculations or filtering commands are performed on a project or visualization application. In some instances, a visualization application receives (as input) an output generated from one or more pipeline modules performed on the system.

[0070] A visualization application may be configured for a user to view biomarkers. In some instances, biomarkers are identified through the platform pipelines against one or more of an annotation and genomic reference tables. In some instances, a visualization application comprises of one or more modules which represent one or more characteristics of a biological data type. In some instances, a user is able to directly manipulate visual representations in a visualization application. In some instances, a visual representation comprises one or more of feature selection, feature filtering, importing gene lists, and external biomarker results. In some instances, configurations or manipulations of a visual application are exported or imported. In some instances, In some instances, configurations or manipulations control a layout of other visualization applications within the platform. In some instances, chromosomal coordinates are leveraged to highlight various information across multiple visualization applications. In some instances, biological groupings, label or group samples within a visualization module within an

application. In some instances, biological groupings comprises cellular phenotypes. In some instances, a user is able to view biomarker information within genomic context, leveraging one or more genomic loci, along with associated annotation. In some instances, a system depicts one or more modules within a visualization application for transcription. In some instances, a visualization application depicts one or more of heatmaps, principal components analysis, differential expression, and alternative transcriptional mapping. In some instances, a visualization application is configured for a user to view changes in copy number or structural changes within genome scale. In some instances, a visualization application is configured for a user to view annotated information of changes in copy number or structural changes within the genome scale.

[0071] A system may be configured for syncopation of molecular data. In some instances, syncopation of molecular data comprises managing and querying for one or more analysis tools. In some instances, molecular data comprises data objects derived from the same cell. In some instances, syncopated molecular information is visualized together on the same application within the platform. In some instances, syncopated molecular information is visualized together using one or more of a Circos plots, genome viewers, and graph networks.

[0072] A system may be configured to associate molecular information with biological groupings. In some instances, a system is configured to associate molecular information with biological groupings using one or more statistical methods. In some instances, one or more statistical comprises differential expression (transcriptomic/proteomic), genome-wide association studies (genomic) and quantitative trait loci calculations using molecular information. In some instances, a system is configured to export information in the system contained in one or more objects. In some instances, one or more objects comprise files, images, pipeline and visualization configurations, biomarker lists or projects.

[0073] Data may be visualized using plots. In some instances, data may be visualized using a circos plot (**FIG. 5A**) Circos plots in some instances comprise additional variant information, such as number of somatic, germline, SNP or indel variants. Variants in some instances visualized at the chromosome level (**FIG. 5B**). In some instances, a circos plot comprises a lineage tree (**FIG. 7**). In some instances, a user interface is configured to apply one or more filters to the circos plot. In some instances, two or more groups of cells or samples are compared (optionally filtered by number of variants). In some instances, views of one or more chromosomes are displayed or hidden. In some instances, data from one or more cells is hidden or displayed. In some instances, variant filters comprise one or more of variant type (SNP, indel), origin (somatic vs. germline), annotation (COSMIC, CLINVAR, coding change), or features. In some instances features comprise name, gene id, gene type, strand, TdI, HgncId. In

some instances, variant filters comprise predictions (SIFT, FATHMM, PROVEAN, MetaSVM, and MetaLR). Upon selection of a region or chromosome within a cell's genome, a pop-up window is in some instances presented to the user which includes a genome viewing frame (e.g., IGV) plot. This window can be configured in terms of genome window bin size allowing the visualization of the entire chromosome to the individual bases across that genome, which can be completed in matter of seconds. The window size in some instances is scrollable by simply dragging the window left or right. In the IGV window, each sample in some instances is interrogated to determine, for example, the specific change which is highlighted by a color change from the parental allele. The alternative allele is selected to determine the base change, while the parent allele can be detected to determine pathogenic risk score based on several public algorithms as well as the conservation of the allele across several vertebrate and mammalian species. This variant annotation further provides links to several databases to provide greater detail of the impact of the genomic alteration.

[0074] The systems and methods described herein may provide a visualization of genomic and multi-omic data having a large number of datasets. In some instances, the genomic data comprises at least 1, 2, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 750, 1000, or at least 1500 datasets. In some instances, the genomic data comprises 1-1000, 5-1000, 10-1000, 5-10,000, 100-10,000, 100-10,000, 100-1000, 10-500, 10-750, 50-750, or 50-500 datasets. In some instances each sample dataset corresponds to a single cell.

[0075] The systems and methods described herein may provide a visualization of genomic data comprises datasets with a large number of variants. In some instances, each dataset comprises at least 500, 1000, 2000, 5000, 10,000, 50,000, 100,000, 150,000, 250,000, 500,000 1 million, 2 million, 3 million, 4 million, 5 million, 6 million, 10 million or at least 15 million variants. In some instances, each dataset comprises about 500, 1000, 2000, 5000, 10,000, 50,000, 100,000, 150,000, 250,000, 500,000 1 million, 2 million, 3 million, 4 million, 5 million, 6 million, 10 million or about 15 million variants. In some instances, each dataset comprises 100-1 million, 100-100,000, 100,000-1 million, 100,000-5 million, 100-500,000, 500-5 million, 1 million-2 million, 2 million to 6 million, 3 million to 10 million, or 4 million to 7 million variants.

[0076] The systems and methods described herein may provide a visualization of genomic data within datasets represented in table format. In some instances, datasets comprise at least 1, 2, 5, 10, 20, 25, 50, 75, 80, 85, 90, 95, 100, 110, 120, 150, 200, or at least 250 million rows of data. In some instances, datasets comprise no more than 1, 2, 5, 10, 20, 25, 50, 75, 80, 85, 90, 95, 100, 110, 120, 150, 200, or no more than 250 million rows of data. In some instances,

datasets comprise 1-250 , 1-100, 1-50, 1-25, 5-25, 5-50, 10-100, 10-200, 50-200, 50-150, 100-400 or 100-300 million rows of data.

[0077] The systems and methods described herein may provide a visualization of genomic data in a short period of time. In some instances, a system for visualizing genomic data comprises one or more of a device comprising at least one processor and instructions executable by the at least one processor to provide a first application configured to perform operations comprising: i. accessing one or more datasets comprising genomic data; and ii. generating a visual representation of the one or more datasets. In some instances, the visualization comprises a circos plot. In some instances, the circos plot is generated in no more than 10, 5, 4, 3, 2, 1, 0.5, 0.2, 0.1, 0.05, or no more than 0.01 seconds. In some instances, the circos plot is generated in 0.01-10, 0.05-10, 0.1-50, 0.5-10, 1-10, 2-10, 5-10, 0.01-0.05, 0.01-0.1, 0.01-0.5, 0.1-0.5, 0.1-1, or 0.1-5 seconds. In some instances, the circos plot is generated in no more than 10, 5, 4, 3, 2, 1, 0.5, 0.2, 0.1, 0.05, or no more than 0.01 seconds for dataset having at least 5 cells. In some instances, the circos plot is generated in 0.01-10, 0.05-10, 0.1-50, 0.5-10, 1-10, 2-10, 5-10, 0.01-0.05, 0.01-0.1, 0.01-0.5, 0.1-0.5, 0.1-1, or 0.1-5 seconds for a dataset having at least 5 cells. In some instances, the circos plot is generated in no more than 10, 5, 4, 3, 2, 1, 0.5, 0.2, 0.1, 0.05, or no more than 0.01 seconds for dataset having at least 10 cells. In some instances, the circos plot is generated in 0.01-10, 0.05-10, 0.1-50, 0.5-10, 1-10, 2-10, 5-10, 0.01-0.05, 0.01-0.1, 0.01-0.5, 0.1-0.5, 0.1-1, or 0.1-5 seconds for a dataset having at least 10 cells. In some instances, the circos plot is generated in no more than 10, 5, 4, 3, 2, 1, 0.5, 0.2, 0.1, 0.05, or no more than 0.01 seconds for dataset having at least 20 cells. In some instances, the circos plot is generated in 0.01-10, 0.05-10, 0.1-50, 0.5-10, 1-10, 2-10, 5-10, 0.01-0.05, 0.01-0.1, 0.01-0.5, 0.1-0.5, 0.1-1, or 0.1-5 seconds for a dataset having at least 20 cells. In some instances, the circos plot is generated in no more than 10, 5, 4, 3, 2, 1, 0.5, 0.2, 0.1, 0.05, or no more than 0.01 seconds for dataset having at least 1 million variants per cell. In some instances, the circos plot is generated in 0.01-10, 0.05-10, 0.1-50, 0.5-10, 1-10, 2-10, 5-10, 0.01-0.05, 0.01-0.1, 0.01-0.5, 0.1-0.5, 0.1-1, or 0.1-5 seconds for a dataset having least 1 million variants per cell. In some instances, the circos plot is generated in no more than 10, 5, 4, 3, 2, 1, 0.5, 0.2, 0.1, 0.05, or no more than 0.01 seconds for dataset having at least 4 million variants per cell. In some instances, the circos plot is generated in 0.01-10, 0.05-10, 0.1-50, 0.5-10, 1-10, 2-10, 5-10, 0.01-0.05, 0.01-0.1, 0.01-0.5, 0.1-0.5, 0.1-1, or 0.1-5 seconds for a dataset having least 4 million variants per cell. In some instances, the circos plot is generated using no more than 1, 2, 3, 4, 5, 6, 7, or no more than 8 processors. In some instances, the circos plot is generated using at least 1, 2, 3, 4, 5, 6, 7, or at least 8 processors. In some instances, the circos plot is generated using about 1, 2, 3, 4, 5, 6, 7, or about 8 processors.

[0078] In some instances the visualization further comprises a phylogenic tree. In some instances the visualization further comprises sequencing quality metrics. In some instances the visualization further comprises annotated variations. In some instances the visualization further comprises number of variations. In some instances the visualization further comprises cell and cell population statistics.

[0079] Also described herein are platforms comprising: a database, in a computer memory, comprising biologic information for member of a population of individuals or samples, the biologic information comprising genome data, the biologic information obtained by analysis of one or more biologic samples from each sample and/or individual, each individual and/or sample having an ID; and a processor configured to provide a biologic information visual synthesis application comprising: a software module presenting an interface allowing a user to query the database one or more of: inputting a phenotype, inputting a gene name, inputting an individual ID, and inputting a sample ID; a software module generating a genome browser, the genome browser comprising: a whole genome display comprising an icon representing each chromosome, each icon indicating a density of gene variants; and a chromosome display comprising an iconic representation of each chromosome, the representation indicating a density of gene variants located at the relevant portion of the chromosome, wherein selection of a chromosome by a user generates a linear display of the chromosome demonstrating the spatial relationship on the chromosome of the genes and the variants; and a software module generating a lineage viewer, the lineage viewer comprising: a geographic display of autosomal ancestry, a geographic display of maternal line ancestry, and a geographic display of paternal line ancestry and multi-omic displays.

[0080] Information from a system or application may be exported. In some instances, information is exported automatically, or by action of a user. In some instances, visualization data is exported. In some instances, annotated datasets are exported. In some instances, information is exported as one more files within a project through a file explorer embedded directly on the platform. In some instances, configurations of visualization applications is exported. In some instances, configurations of visualization applications is exported in a format readable by the platform's visualization applications. In some instances, a system is configured for export of images of visual modules or applications which preserve manipulations performed by the user. In some instances, a system is configured for specific output directories. In some instances, a user has rights and access to place information in the output directory. In some instances, a system is configured for export of a project or pipeline to specified output locations.

[0081] Certain definitions

[0082] Unless otherwise defined, all technical terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. As used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural references unless the context clearly dictates otherwise. Any reference to "or" herein is intended to encompass "and/or" unless otherwise stated.

[0083] Unless defined otherwise, all terms of art, notations and other technical and scientific terms or terminology used herein are intended to have the same meaning as is commonly understood by one of ordinary skill in the art to which the claimed subject matter pertains. In some cases, terms with commonly understood meanings are defined herein for clarity and/or for ready reference, and the inclusion of such definitions herein should not necessarily be construed to represent a substantial difference over what is generally understood in the art.

[0084] Throughout this application, various embodiments may be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the disclosure. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[0085] As used in the specification and claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a sample" includes a plurality of samples, including mixtures thereof.

[0086] The terms "determining," "measuring," "evaluating," "assessing," "assaying," and "analyzing" are often used interchangeably herein and can refer to forms of measurement. The terms include determining if an element is present or not (for example, detection). These terms can include quantitative, qualitative or quantitative and qualitative determinations. Assessing can be relative or absolute. "Detecting the presence of" can include determining the amount of something present in addition to determining whether it is present or absent depending on the context.

[0087] As used herein, the term "gene" can refer to a linear sequence of nucleotides along a segment of DNA that provides the coded instructions for synthesis of RNA, which, when translated into protein, leads to the expression of hereditary character.

[0088] As used herein, the term "nucleic acid molecule" can mean DNA, RNA, single-stranded, double-stranded or triple stranded and any chemical modifications thereof. Virtually

any modification of the nucleic acid is contemplated. A "nucleic acid molecule" can be of almost any length, from 10, 20, 30, 40, 50, 60, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 6000, 7000, 8000, 9000, 10,000, 15,000, 20,000, 30,000, 40,000, 50,000, 75,000, 100,000, 150,000, 200,000, 500,000, 1,000,000, 1,500,000, 2,000,000, 5,000,000 or even more bases in length, including increments therein, up to a full-length chromosomal DNA molecule. For methods that analyze expression of a gene, the nucleic acid isolated from a sample is typically RNA.

[0089] A single-stranded nucleic acid molecule is "complementary" to another single-stranded nucleic acid molecule, in certain embodiments of the subject matter described herein, when it can base-pair (hybridize) with all or a portion of the other nucleic acid molecule to form a double helix (double-stranded nucleic acid molecule), based on the ability of guanine (G) to base pair with cytosine (C) and adenine (A) to base pair with thymine (T) or uridine (U). For example, the nucleotide sequence 5′-TATAC-3′ is complementary to the nucleotide sequence 5′-GTATA-3′.

[0090] As used herein, the term "mutation" can refer to a change in the genome with respect to the standard wild-type sequence. Mutations can be deletions, insertions, or rearrangements of nucleic acid sequences at a position in the genome, or they can be single base changes at a position in the genome, referred to as "point mutations." Mutations can be inherited, or they can occur in one or more cells during the lifespan of an individual. In some instances, mutation and variant are used synonymously.

[0091] As used herein, the term "kit" or "research kit" can refer to a collection of products that are used to perform a biological research reaction, procedure, or synthesis, such as, for example, a detection, assay, separation, purification, etc., which are typically shipped together, usually within a common packaging, to an end user.

[0092] Described herein is a cloud-based solution for the storage, query, and analysis of longitudinal data comprising a multiplicity of whole genomes, a large number of public and proprietary annotation sources as well as associated high quality phenotypic data, including microbiome metagenomes and metabolomics profiles. In various embodiments, the data analyzed by the platforms, systems, media, and methods described herein comprises more than 1,000, more than 5,000, more than 10,000, more than 20,000, more than 50,000, more than 100,000, more than 500,000, or more than 1,000,000 whole genomes.

[0093] The data analyzed by the platforms, systems, media, and methods described herein comprises genomic data. The genomic data is produced, by way of example, at a next generation sequencing (NGS) lab. In some cases, an AWS analysis pipeline based on Illumina's HiSeq X and the ISIS Analysis Software are utilized to produce the genomic data. Sequencing reads are

mapped to the hg38 human reference sequence and the Isaac Variant Caller is used to call single nucleotide variants (SNVs) and insertions and deletions (indels). The genomic data comprises a multiplicity of unique SNVs. By way of examples, the genomic data comprises over 1 million, over 10 million, over 50 million, over 100 million, over 500 million, or over 1 billion unique SNVs.

[0094] The data analyzed by the platforms, systems, media, and methods described herein comprises metadata. The whole genomes are associated with high quality phenotypic information. A proprietary phenotype ingestion process enables the cleaning and standardization of phenotype data across disparate data sources. In some embodiments, the ingestion process includes: data integrity checks; standardization of units; standardization of terms; ontology/vocabulary mapping; and maintenance of the proprietary data dictionary.

[0095] In various embodiments, the phenotype data comprises more than 1000, more than 5000, more than 10,000, more than 100,000, more than 1,000,000, or more than 10,000,000 phenotype data fields with, more than 1 million, more than 5 million, more than 10 million, more than 50 million, more than 100 million, more than 500 million, or more than 1 billion data points. Phenotypic data in some instances comprises cellular phenotype data. In some instances, cellular phenotypic data obtained from microscopy. In some instances, cell phenotypic data comprises one or more observable phenotypic traits such as cell shape or morphology, size, texture, internal structure, patterns of distribution of one or more specific proteins, glycosylated proteins, nucleic acid molecules, lipid molecules, glycosylated lipid molecules, carbohydrate molecules, metabolites, and ions. In some instances, phenotypic data describes populations of cells described herein. In some instances, phenotypic data describes phenotypic traits of an organism such as a human. In some instances, a phenotypic data comprises a clinical designation or category, for example, a clinical diagnosis, a clinical parameter name, a clinical parameter value, a laboratory test name or a laboratory test value. In some instances, a phenotype is associated with an observable disease characteristic.

[0096] The data analyzed by the platforms, systems, media, and methods described herein comprises annotation data. Annotation data is also cleaned and standardized through an automated end-to-end solution, which allows: idempotence, immutability, persistence; high quality data; consistency between data sources; and scalability and flexibility.

[0097] Samples (also referred to as biosamples) described herein may represent biologic information obtained from individuals or populations of individuals (e.g., genomic information). In some instances, samples comprise single cells. In some instances, samples comprise 1, 2, 5, 10, 20, 25, 50, 75, 100, 200, 500, or more than 1000 cells from the same or different individual. In some instances, samples comprise 1000, 2000, 5000, 10,000 20,000, 50,000, 75,000, or at

least 100,000 cells from the same or different individual. Samples may be obtained from any species, including but not limited to viruses, bacteria, plants, fungi, protozoa, archaea, or animals. In some instances, samples are obtained from vertebrates. In some instances, samples are obtained from mammals. In some instances, samples are obtained from humans. Samples in some instances are obtained from any bodily fluid or tissue (blood, semen, spinal fluid, serum, saliva, mucus, or other fluid). In some instances, samples are obtained from diseased tissue such as a tumor.

[0098] Population of individuals

[0099] The platforms, systems, media, and methods described herein include biologic data pertaining to a population of individuals, or use of the same. In various embodiments, the population of individuals comprises more than 1,000, more than 5,000, more than 10,000, more than 20,000, more than 50,000, or more than 100,000, more than 500,000, more than 1,000,000 more than 10,000,000, more than 50,000,000, or more than 100,000,000 individuals. In some cases, the individuals in the population participated in academic medical research studies using consents allowing for genetic testing of specimens. In such cases, biologic specimens and phenotype data are collected for individuals from pharmaceutical clinical trials, academic research, and health care settings. In some cases, biologic data pertaining to a population of individuals is collected from integrated health records for individuals representing a spectrum of diseases with unmet medical needs.

[0100] Biologic or Molecular information

[0101] The platforms, systems, media, and methods described herein include biologic (molecular) information, or use of the same. In some instances, biologic information comprises genetic information. In some embodiments, the biologic information comprises whole human genome sequencing information. In some embodiments, the biologic information comprises human transcriptome sequencing information. In some instances biologic information comprises genetic information from humans, non-human primates, animals, plants, fungi, protozoa, archaea, or bacteria. In some instances biologic information comprises genetic information from the microbiome.

[0102] The biologic information may comprise genomic information. As used herein, genomic information refers to genetic information found within a biological sample arising from the genome (or DNA – nuclear, mitochondrial or otherwise). In some instances, genomic information comprises nucleic acid sequence copy number, location, and sequence. The genomic information is not limited to protein-coding sequence, it may refer to intronic sequence and intergenic sequence, each known to harbor multiple functional elements whereby DNA changes at those elements may be consequential in normal development and disease. In some

instances, genomic information comprises post-transcriptional modifications such as methylation. In some instances, genomic information is found within a chromosome, plasmid, or other medium comprising nucleic acids.

[0103] The biologic information may comprise transcript information. As used herein, transcript information refers to information obtained from a transcriptome within a biological sample. In some instances, transcript information comprises expression levels of genes and sequence of corresponding nucleic acids expressed from genes.

[0104] The biologic information may comprise microbiome information. As used herein, "microbiome" refers to the bacteria and other microorganisms that live in and on the human body. In some embodiments, the microbiome information comprises metagenomic microbiome characterization. In various embodiments, the microbiome information comprises one or more of: microflora genus and/or species information, microflora relative abundance information, and microflora gene and/or gene variant information.

[0105] The biologic information may comprise proteome information. In some embodiments, the proteome information comprises information regarding abundance, localization, identity, post-transcriptional modifications, or other protein information.

[0106] The biologic information may comprise methylome information. In some embodiments, methylome information comprises post-transcriptional modifications such as the location of 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), CpG islands, or other post-transcriptional modification to nucleic acids.

[0107] The biologic information may comprise metabolome information. As used herein, "metabolome" refers to the small-molecule chemicals found within a biological sample. In some embodiments, metabolome information comprises the presence of one or more small-molecule chemicals. In further embodiments, the metabolome information comprises a qualitative measurement of one or more small-molecule chemicals. In still further embodiments, the metabolome information comprises a quantitative measurement of one or more small-molecule chemicals. In various embodiments, the microbiome information comprises measurements of at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, at least 1200, at least 1300, at least 1400, or at least 1500 substances (e.g., molecules).

[0108] Data security

[0109] Databases and visualizations described herein may comprise sensitive information pertaining to an individual's health. In some instances, a system provided herein comprises a platform for data security. Provided herein are platforms for data security comprising one or more of an access control for one or more users; a security framework; and biological data from

an individual. In some instances, one or more security measures are implemented via security frameworks to restrict access or protect an individual's health information. Security frameworks in some instances comprises standards. In some instances, security frameworks include HIPPA standards. In some instances, security frameworks comprise NIST cybersecurity framework. Access controls in some instances restrict access to certain individuals or groups of individuals. Access controls in some instances comprise passwords, biometrics, or other method of user authentication. In some instances, a platform for data security comprises one or more of an access control for one or more users; a security framework; and molecular data from an individual or biological species. In some instances, a back end module is configured to host a platform for data security. In some instances, a back end module is configured to perform user authentication and registration. In some instances, access to genomic data is limited to authorized users who are a part of authorized organizations. In some instances, the system is accessible by at least 2, 3, 4, 5, 10, 12, 15, 20, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 300, 400, 500, 600, 700, 800, 900 or at least 1000 users concurrently. In some instances, the system is accessible by 2-1000, 5-1000, 10-1000, 20-1000, 50-1000, 100-1000, 200-1000, 500-1000, or 500-10,000 users concurrently.

[0110] Use Cases

[0111] The system can be applied in a variety of fields. In some instances, the system provides useful data and analysis to pharmaceutical companies, including informaticians, bench scientists, medical director, the senior executive team, or commercial organizations. Such data and analysis in some instances includes analysis of clinical trial data for patient stratification and biomarker discovery, identification and in silico validation of novel genetic targets, discovery of novel disease and dose response biomarkers/signatures, compound repurposing and expand indications of marketed drugs, rescue of failed clinical trial assets, real time genetic analysis of adverse events, or targeted accelerated recruitment for clinical trials. For academic research groups, including physicians/principal investigators, informaticians, research scientists and geneticists, the system in some instances offers analysis of specific cohorts, analysis of individual patients, or large scale analysis of variation in populations. Clinics, hospitals and cancer centers, including physicians and genetic counsellors, in some instances will find the system useful in the analysis of individuals, analysis of cohorts, wellness focus, or oncology focus. The data and analysis in some instances also has value to insurance companies, actuarial teams, or health economists.

[0112] Specifically, for pharma and researchers, the system can serve as or enable a reference set of knowledge/evidence, a hypothesis generation engine, a platform for analysis of pharma's own data, a platform for combination of pharma data and data and analysis provided

by the system, a platform for combining data from multiple collaborators, a platform for sharing data within a company, etc. For physicians or genetic counsellors, the system can similarly be used as part of a care tool to identify the most relevant results for treatment and prevention, a reference set of knowledge/evidence, or a tool to identify other physicians with similar patients/ share knowledge. In addition, for insurance companies, the system can be useful as part of a tool for detect individual care pathway and incentivize healthy living or a tool to help quantify risk that they have in the insured population.

[0113] Kits

[0114] The systems described herein may accompany or be provided as a service with a kit. In some instances, the kit comprises reagents for acquiring biological information. In some instances, the kit is configured to obtain genomic or transcriptome data. In some instances, the kit is configured to obtain genomic, methylome, transcriptome or proteome data from single cells. In some instances, provided herein are kits comprising reagents for obtaining biological data from single cells, and instructions for using the kit. In some instances, the instructions comprise links to a web-based portal or mobile based software application to import, analyze, and/or compare biological data obtained from the kit.

[0115] Digital processing device

[0116] In some embodiments, the platforms, systems, media, and methods described herein include a digital processing device, or use of the same. In further embodiments, the digital processing device includes one or more hardware central processing units (CPUs) or general purpose graphics processing units (GPGPUs) that carry out the device's functions. In still further embodiments, the digital processing device further comprises an operating system configured to perform executable instructions. In some embodiments, one or more resources related to the systems described herein is stored locally. In some embodiments, the digital processing device is optionally connected a computer network. In further embodiments, the digital processing device is optionally connected to the Internet such that it accesses the World Wide Web. In still further embodiments, the digital processing device is optionally connected to a cloud computing infrastructure. In other embodiments, the digital processing device is optionally connected to an intranet. In other embodiments, the digital processing device is optionally connected to a data storage device.

[0117] In accordance with the description herein, suitable digital processing devices include, by way of non-limiting examples, server computers, desktop computers, laptop computers, and notebook computers.

[0118] In some embodiments, the digital processing device includes an operating system configured to perform executable instructions. The operating system is, for example, software,

including programs and data, which manages the device's hardware and provides services for execution of applications. Those of skill in the art will recognize that suitable server operating systems include, by way of non-limiting examples, FreeBSD, OpenBSD, NetBSD®, Linux, Apple® Mac OS X Server®, Oracle® Solaris®, Windows Server®, and Novell® NetWare®. Those of skill in the art will recognize that suitable personal computer operating systems include, by way of non-limiting examples, Microsoft® Windows®, Apple® Mac OS X®, UNIX®, and UNIX-like operating systems such as GNU/Linux®. In some embodiments, the operating system is provided by cloud computing.

[0119] In some embodiments, the device includes a storage and/or memory device. The storage and/or memory device is one or more physical apparatuses used to store data or programs on a temporary or permanent basis. In some embodiments, the device is volatile memory and requires power to maintain stored information. In some embodiments, the device is non-volatile memory and retains stored information when the digital processing device is not powered. In further embodiments, the non-volatile memory comprises flash memory. In some embodiments, the non-volatile memory comprises dynamic random-access memory (DRAM). In some embodiments, the non-volatile memory comprises ferroelectric random access memory (FRAM). In some embodiments, the non-volatile memory comprises phase-change random access memory (PRAM). In other embodiments, the device is a storage device including, by way of non-limiting examples, CD-ROMs, DVDs, flash memory devices, magnetic disk drives, magnetic tapes drives, optical disk drives, and cloud computing based storage. In further embodiments, the storage and/or memory device is a combination of devices such as those disclosed herein.

[0120] In some embodiments, the digital processing device includes a display to send visual information to a user. In some embodiments, the display is a cathode ray tube (CRT). In some embodiments, the display is a liquid crystal display (LCD). In further embodiments, the display is a thin film transistor liquid crystal display (TFT-LCD). In some embodiments, the display is an organic light emitting diode (OLED) display. In various further embodiments, on OLED display is a passive-matrix OLED (PMOLED) or active-matrix OLED (AMOLED) display. In some embodiments, the display is a plasma display. In other embodiments, the display is a video projector. In some embodiments, the display is a wearable display. In still further embodiments, the display is a combination of devices such as those disclosed herein.

[0121] In some embodiments, the digital processing device includes an input device to receive information from a user. In some embodiments, the input device is a keyboard. In some embodiments, the input device is a pointing device including, by way of non-limiting examples, a mouse, trackball, track pad, joystick, game controller, or stylus. In some embodiments, the

input device is a touch screen or a multi-touch screen. In other embodiments, the input device is a microphone to capture voice or other sound input. In other embodiments, the input device is a video camera or other sensor to capture motion or visual input. In further embodiments, the input device is a Kinect, Leap Motion, or the like. In still further embodiments, the input device is a combination of devices such as those disclosed herein.

[0122] Non-transitory computer readable storage medium

[0123] In some embodiments, the platforms, systems, media, and methods disclosed herein include one or more non-transitory computer readable storage media encoded with a program including instructions executable by the operating system of an optionally networked digital processing device. In further embodiments, a computer readable storage medium is a tangible component of a digital processing device. In still further embodiments, a computer readable storage medium is optionally removable from a digital processing device. In some embodiments, a computer readable storage medium includes, by way of non-limiting examples, CD-ROMs, DVDs, flash memory devices, solid state memory, magnetic disk drives, magnetic tape drives, optical disk drives, cloud computing systems and services, and the like. In some cases, the program and instructions are permanently, substantially permanently, semi-permanently, or non-transitorily encoded on the media.

[0124] Computing system

[0125] Referring to **FIG. 11**, a block diagram is shown depicting an exemplary machine that includes a computer system **1100** (e.g., a processing or computing system) within which a set of instructions can execute for causing a device to perform or execute any one or more of the aspects and/or methodologies for static code scheduling of the present disclosure. The components in **FIG. 11** are examples only and do not limit the scope of use or functionality of any hardware, software, embedded logic component, or a combination of two or more such components implementing particular embodiments.

[0126] Computer system **1100** may include one or more processors **1101**, a memory **1103**, and a storage **1108** that communicate with each other, and with other components, via a bus **1140**. The bus **1140** may also link a display **1132**, one or more input devices **1133** (which may, for example, include a keypad, a keyboard, a mouse, a stylus, etc.), one or more output devices **1134**, one or more storage devices **1135**, and various tangible storage media 1**136**. All of these elements may interface directly or via one or more interfaces or adaptors to the bus **1140**. For instance, the various tangible storage media **1136** can interface with the bus **1140** via storage medium interface **1126**. Computer system 1**100** may have any suitable physical form, including but not limited to one or more integrated circuits (ICs), printed circuit boards (PCBs), mobile

handheld devices (such as mobile telephones or PDAs), laptop or notebook computers, distributed computer systems, computing grids, or servers.

[0127] Computer system **1100** includes one or more processor(s) **1101** (e.g., central processing units (CPUs), general purpose graphics processing units (GPGPUs), or quantum processing units (QPUs)) that carry out functions. Processor(s) **1101** optionally contains a cache memory unit **1102** for temporary local storage of instructions, data, or computer addresses. Processor(s) **1101** are configured to assist in execution of computer readable instructions. Computer system **1100** may provide functionality for the components depicted in **FIG. 11** as a result of the processor(s) **1101** executing non-transitory, processor-executable instructions embodied in one or more tangible computer-readable storage media, such as memory **1103**, storage **1108**, storage devices **1135**, and/or storage medium **1136**. The computer-readable media may store software that implements particular embodiments, and processor(s) **1101** may execute the software. Memory **1103** may read the software from one or more other computer-readable media (such as mass storage device(s) **1135**, **1136**) or from one or more other sources through a suitable interface, such as network interface **1120**. The software may cause processor(s) **1101** to carry out one or more processes or one or more steps of one or more processes described or illustrated herein. Carrying out such processes or steps may include defining data structures stored in memory **1103** and modifying the data structures as directed by the software.

[0128] The memory **1103** may include various components (e.g., machine readable media) including, but not limited to, a random access memory component (e.g., RAM **1104**) (e.g., static RAM (SRAM), dynamic RAM (DRAM), ferroelectric random access memory (FRAM), phase-change random access memory (PRAM), etc.), a read-only memory component (e.g., ROM **1105**), and any combinations thereof. ROM **1105** may act to communicate data and instructions unidirectionally to processor(s) **1101**, and RAM **1104** may act to communicate data and instructions bidirectionally with processor(s) **1101**. ROM **1105** and RAM **1104** may include any suitable tangible computer-readable media described below. In one example, a basic input/output system **106** (BIOS), including basic routines that help to transfer information between elements within computer system **1100**, such as during start-up, may be stored in the memory **1103**.

[0129] Fixed storage **1108** is connected bidirectionally to processor(s) 1**101**, optionally through storage control unit **1107**. Fixed storage **1108** provides additional data storage capacity and may also include any suitable tangible computer-readable media described herein. Storage **108** may be used to store operating system **1109**, executable(s) **1110**, data **1111**, applications **1112** (application programs), and the like. Storage **1108** can also include an optical disk drive, a solid-state memory device (e.g., flash-based systems), or a combination of any of the above.

Information in storage **1108** may, in appropriate cases, be incorporated as virtual memory in memory **1103**.

[0130] In one example, storage device(s) **1135** may be removably interfaced with computer system **1100** (e.g., via an external port connector (not shown)) via a storage device interface **1125**. Particularly, storage device(s) **1135** and an associated machine-readable medium may provide non-volatile and/or volatile storage of machine-readable instructions, data structures, program modules, and/or other data for the computer system **1100**. In one example, software may reside, completely or partially, within a machine-readable medium on storage device(s) **1135**. In another example, software may reside, completely or partially, within processor(s) **1101**.

[0131] Bus **1140** connects a wide variety of subsystems. Herein, reference to a bus may encompass one or more digital signal lines serving a common function, where appropriate. Bus **140** may be any of several types of bus structures including, but not limited to, a memory bus, a memory controller, a peripheral bus, a local bus, and any combinations thereof, using any of a variety of bus architectures. As an example and not by way of limitation, such architectures include an Industry Standard Architecture (ISA) bus, an Enhanced ISA (EISA) bus, a Micro Channel Architecture (MCA) bus, a Video Electronics Standards Association local bus (VLB), a Peripheral Component Interconnect (PCI) bus, a PCI-Express (PCI-X) bus, an Accelerated Graphics Port (AGP) bus, HyperTransport (HTX) bus, serial advanced technology attachment (SATA) bus, and any combinations thereof.

[0132] Computer system **1100** may also include an input device **1133**. In one example, a user of computer system **1100** may enter commands and/or other information into computer system **1100** via input device(s) **1133**. Examples of an input device(s) **1133** include, but are not limited to, an alpha-numeric input device (e.g., a keyboard), a pointing device (e.g., a mouse or touchpad), a touchpad, a touch screen, a multi-touch screen, a joystick, a stylus, a gamepad, an audio input device (e.g., a microphone, a voice response system, etc.), an optical scanner, a video or still image capture device (e.g., a camera), and any combinations thereof. In some embodiments, the input device is a Kinect, Leap Motion, or the like. Input device(s) **1133** may be interfaced to bus **1140** via any of a variety of input interfaces **1123** (e.g., input interface **1123**) including, but not limited to, serial, parallel, game port, USB, FIREWIRE, THUNDERBOLT, or any combination of the above.

[0133] In particular embodiments, when computer system **1100** is connected to network **1130**, computer system **1100** may communicate with other devices, specifically mobile devices and enterprise systems, distributed computing systems, cloud storage systems, cloud computing systems, and the like, connected to network **1130**. Communications to and from computer

system **100** may be sent through network interface **1120**. For example, network interface **1120** may receive incoming communications (such as requests or responses from other devices) in the form of one or more packets (such as Internet Protocol (IP) packets) from network **1130**, and computer system **100** may store the incoming communications in memory **1103** for processing. Computer system **100** may similarly store outgoing communications (such as requests or responses to other devices) in the form of one or more packets in memory 1**103** and communicated to network **1130** from network interface **1120**. Processor(s) **1101** may access these communication packets stored in memory **1103** for processing.

[0134] Examples of the network interface **1120** include, but are not limited to, a network interface card, a modem, and any combination thereof. Examples of a network **1130** or network segment **1130** include, but are not limited to, a distributed computing system, a cloud computing system, a wide area network (WAN) (e.g., the Internet, an enterprise network), a local area network (LAN) (e.g., a network associated with an office, a building, a campus or other relatively small geographic space), a telephone network, a direct connection between two computing devices, a peer-to-peer network, and any combinations thereof. A network, such as network **1130**, may employ a wired and/or a wireless mode of communication. In general, any network topology may be used.

[0135] Information and data can be displayed through a display **1132**. Examples of a display **1132** include, but are not limited to, a cathode ray tube (CRT), a liquid crystal display (LCD), a thin film transistor liquid crystal display (TFT-LCD), an organic liquid crystal display (OLED) such as a passive-matrix OLED (PMOLED) or active-matrix OLED (AMOLED) display, a plasma display, and any combinations thereof. The display **1132** can interface to the processor(s) **1101**, memory **1103**, and fixed storage **1108**, as well as other devices, such as input device(s) **1133**, via the bus **1140**. The display **1132** is linked to the bus **1140** via a video interface **1122**, and transport of data between the display **1132** and the bus **1140** can be controlled via the graphics control **1121**. In some embodiments, the display is a video projector. In some embodiments, the display is a head-mounted display (HMD) such as a VR headset. In further embodiments, suitable VR headsets include, by way of non-limiting examples, HTC Vive, Oculus Rift, Samsung Gear VR, Microsoft HoloLens, Razer OSVR, FOVE VR, Zeiss VR One, Avegant Glyph, Freefly VR headset, and the like. In still further embodiments, the display is a combination of devices such as those disclosed herein.

[0136] In addition to a display **1132**, computer system **1100** may include one or more other peripheral output devices **1134** including, but not limited to, an audio speaker, a printer, a storage device, and any combinations thereof. Such peripheral output devices may be connected to the bus **1140** via an output interface **1124**. Examples of an output interface **1124** include, but

are not limited to, a serial port, a parallel connection, a USB port, a FIREWIRE port, a THUNDERBOLT port, and any combinations thereof.

[0137] In addition, or as an alternative, computer system **1100** may provide functionality as a result of logic hardwired or otherwise embodied in a circuit, which may operate in place of or together with software to execute one or more processes or one or more steps of one or more processes described or illustrated herein. Reference to software in this disclosure may encompass logic, and reference to logic may encompass software. Moreover, reference to a computer-readable medium may encompass a circuit (such as an IC) storing software for execution, a circuit embodying logic for execution, or both, where appropriate. The present disclosure encompasses any suitable combination of hardware, software, or both.

[0138] Those of skill in the art will appreciate that the various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the embodiments disclosed herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, circuits, and steps have been described above generally in terms of their functionality.

[0139] The various illustrative logical blocks, modules, and circuits described in connection with the embodiments disclosed herein may be implemented or performed with a general purpose processor, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described herein. A general purpose processor may be a microprocessor, but in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration.

[0140] The steps of a method or algorithm described in connection with the embodiments disclosed herein may be embodied directly in hardware, in a software module executed by one or more processor(s), or in a combination of the two. A software module may reside in RAM memory, flash memory, ROM memory, EPROM memory, EEPROM memory, registers, hard disk, a removable disk, a CD-ROM, or any other form of storage medium known in the art. An exemplary storage medium is coupled to the processor such the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. The processor and the storage medium may reside in an ASIC. The

ASIC may reside in a user terminal. In the alternative, the processor and the storage medium may reside as discrete components in a user terminal.

[0141] In accordance with the description herein, suitable computing devices include, by way of non-limiting examples, server computers, desktop computers, laptop computers, notebook computers, sub-notebook computers, netbook computers, netpad computers, set-top computers, media streaming devices, handheld computers, Internet appliances, mobile smartphones, tablet computers, and personal digital assistants. Suitable tablet computers, in various embodiments, include those with booklet, slate, and convertible configurations, known to those of skill in the art.

[0142] In some embodiments, the computing device includes an operating system configured to perform executable instructions. The operating system is, for example, software, including programs and data, which manages the device's hardware and provides services for execution of applications. Operating systems in some instances are stored locally, or accessed via a network. Those of skill in the art will recognize that suitable server operating systems include, by way of non-limiting examples, FreeBSD, OpenBSD, NetBSD®, Linux, Apple® Mac OS X Server®, Oracle® Solaris®, Windows Server®, and Novell® NetWare®. Those of skill in the art will recognize that suitable personal computer operating systems include, by way of non-limiting examples, Microsoft® Windows®, Apple® Mac OS X®, UNIX®, and UNIX-like operating systems such as GNU/Linux®. In some embodiments, the operating system is provided by cloud computing. Those of skill in the art will also recognize that suitable mobile smartphone operating systems include, by way of non-limiting examples, Nokia® Symbian® OS, Apple® iOS®, Research In Motion® BlackBerry OS®, Google® Android®, Microsoft® Windows Phone® OS, Microsoft® Windows Mobile® OS, Linux®, and Palm® WebOS®.

[0143] Computer systems described herein may be utilized as part of the systems and methods of the present invention. In some embodiments, a computer system may be utilized as a device configured for use by a researcher, patient, partner, caretaker, or healthcare provider.

[0144] Non-transitory computer readable storage medium

[0145] In some embodiments, the platforms, systems, media, and methods disclosed herein include one or more non-transitory computer readable storage media encoded with a program including instructions executable by the operating system of an optionally networked computing device. In further embodiments, a computer readable storage medium is a tangible component of a computing device. In still further embodiments, a computer readable storage medium is optionally removable from a computing device. In some embodiments, a computer readable storage medium includes, by way of non-limiting examples, CD-ROMs, DVDs, flash memory devices, solid state memory, magnetic disk drives, magnetic tape drives, optical disk drives,

distributed computing systems including cloud computing systems and services, and the like. In some cases, the program and instructions are permanently, substantially permanently, semi-permanently, or non-transitorily encoded on the media.

[0146] Computer program

[0147] In some embodiments, the platforms, systems, media, and methods disclosed herein include at least one computer program, or use of the same. A computer program includes a sequence of instructions, executable by one or more processor(s) of the computing device's CPU, written to perform a specified task. Computer readable instructions may be implemented as program modules, such as functions, objects, Application Programming Interfaces (APIs), computing data structures, and the like, that perform particular tasks or implement particular abstract data types. In light of the disclosure provided herein, those of skill in the art will recognize that a computer program may be written in various versions of various languages.

[0148] The functionality of the computer readable instructions may be combined or distributed as desired in various environments. In some embodiments, a computer program comprises one sequence of instructions. In some embodiments, a computer program comprises a plurality of sequences of instructions. In some embodiments, a computer program is provided from one location. In other embodiments, a computer program is provided from a plurality of locations. In various embodiments, a computer program includes one or more software modules or features. In various embodiments, a computer program includes, in part or in whole, one or more web applications, one or more mobile applications, one or more standalone applications, one or more web browser plug-ins, extensions, add-ins, or add-ons, or combinations thereof.

[0149] In some embodiments, one or more web applications, one or more mobile applications, one or more standalone applications, one or more web browser plug-ins, extensions, add-ins, or add-ons, or combinations thereof are utilized to perform the methods as described herein. In some embodiments, one or more web applications, one or more mobile applications, one or more standalone applications, one or more web browser plug-ins, extensions, add-ins, or add-ons, or combinations thereof are utilized as part of the systems as described herein. In some embodiments, one or more web applications, one or more mobile applications, one or more standalone applications, one or more web browser plug-ins, extensions, add-ins, or add-ons, or combinations thereof are utilized to fully or partially automate the systems and methods as described herein. In some embodiments, automation allows methods to be carried out which are beyond the limits of what can be processed by a human.

[0150] Web application

[0151] In some embodiments, a computer program includes a web application. In light of the disclosure provided herein, those of skill in the art will recognize that a web application, in various embodiments, utilizes one or more software frameworks and one or more database systems. In some embodiments, a web application is created upon a software framework such as Microsoft® .NET or Ruby on Rails (RoR). In some embodiments, a web application utilizes one or more database systems including, by way of non-limiting examples, relational, non-relational, object oriented, associative, XML, and document oriented database systems. In further embodiments, suitable relational database systems include, by way of non-limiting examples, Microsoft® SQL Server, mySQL™, and Oracle®. Those of skill in the art will also recognize that a web application, in various embodiments, is written in one or more versions of one or more languages. A web application may be written in one or more markup languages, presentation definition languages, client-side scripting languages, server-side coding languages, database query languages, or combinations thereof. In some embodiments, a web application is written to some extent in a markup language such as Hypertext Markup Language (HTML), Extensible Hypertext Markup Language (XHTML), or eXtensible Markup Language (XML). In some embodiments, a web application is written to some extent in a presentation definition language such as Cascading Style Sheets (CSS). In some embodiments, a web application is written to some extent in a client-side scripting language such as Asynchronous JavaScript and XML (AJAX), Flash® ActionScript, JavaScript, or Silverlight®. In some embodiments, a web application is written to some extent in a server-side coding language such as Active Server Pages (ASP), ColdFusion®, Perl, Java™, JavaServer Pages (JSP), Hypertext Preprocessor (PHP), Python™, Ruby, Tcl, Smalltalk, WebDNA®, or Groovy. In some embodiments, a web application is written to some extent in a database query language such as Structured Query Language (SQL). In some embodiments, a web application integrates enterprise server products such as IBM® Lotus Domino®. In some embodiments, a web application includes a media player element. In various further embodiments, a media player element utilizes one or more of many suitable multimedia technologies including, by way of non-limiting examples, Adobe® Flash®, HTML 5, Apple® QuickTime®, Microsoft® Silverlight®, Java™, and Unity®.

[0152] Referring to **FIG. 12**, in a particular embodiment, an application provision system comprises one or more databases **1200** accessed by a relational database management system (RDBMS) **1210**. Suitable RDBMSs include Firebird, MySQL, PostgreSQL, SQLite, Oracle Database, Microsoft SQL Server, IBM DB2, IBM Informix, SAP Sybase, Teradata, and the like. In this embodiment, the application provision system further comprises one or more application severs **1220** (such as Java servers, .NET servers, PHP servers, and the like) and one or more web servers **1230** (such as Apache, IIS, GWS and the like). The web server(s) optionally expose one

or more web services via app application programming interfaces (APIs) **1240**. Via a network, such as the Internet, the system provides browser-based and/or mobile native user interfaces.

[0153] Referring to **FIG. 13**, in a particular embodiment, an application provision system alternatively has a distributed, cloud-based architecture **1300** and comprises elastically load balanced, auto-scaling web server resources **1310** and application server resources **1320** as well synchronously replicated databases **1330**.

[0154] The web applications may be utilized as part of the systems as described herein. The web applications may be utilized to perform the systems as described herein. In some application, web applications are utilized to provide features or modules of the systems described herein. In some application, web applications are utilized to fully or partially automate systems and methods described herein. In some embodiments, automation allows methods to be carried out which are beyond the limits of what can be processed by a human.

[0155] Mobile application

[0156] In some embodiments, a computer program includes a mobile application provided to a mobile computing device. In some embodiments, the mobile application is provided to a mobile computing device at the time it is manufactured. In other embodiments, the mobile application is provided to a mobile computing device via the computer network described herein.

[0157] In view of the disclosure provided herein, a mobile application is created by techniques known to those of skill in the art using hardware, languages, and development environments known to the art. Those of skill in the art will recognize that mobile applications are written in several languages. Suitable programming languages include, by way of non-limiting examples, C, C++, C#, Objective-C, Java™, JavaScript, Pascal, Object Pascal, Python™, Ruby, VB.NET, WML, and XHTML/HTML with or without CSS, or combinations thereof.

[0158] Suitable mobile application development environments are available from several sources. Commercially available development environments include, by way of non-limiting examples, AirplaySDK, alcheMo, Appcelerator®, Celsius, Bedrock, Flash Lite, .NET Compact Framework, Rhomobile, and WorkLight Mobile Platform. Other development environments are available without cost including, by way of non-limiting examples, Lazarus, MobiFlex, MoSync, and PhoneGap. Also, mobile device manufacturers distribute software developer kits including, by way of non-limiting examples, iPhone and iPad (iOS) SDK, Android™ SDK, BlackBerry® SDK, BREW SDK, Palm® OS SDK, Symbian SDK, webOS SDK, and Windows® Mobile SDK.

[0159] Those of skill in the art will recognize that several commercial forums are available for distribution of mobile applications including, by way of non-limiting examples, Apple® App Store, Google® Play, Chrome WebStore, BlackBerry® App World, App Store for Palm devices, App Catalog for webOS, Windows® Marketplace for Mobile, Ovi Store for Nokia® devices, and Samsung® Apps.

[0160] The mobile applications may be utilized as part of the systems as described herein. The mobile applications may be utilized to perform the systems as described herein. In some application, mobile applications are utilized to provide features or modules of the systems described herein. In some application, mobile applications are utilized to fully or partially automate systems and methods described herein. In some embodiments, automation allows methods to be carried out which are beyond the limits of what can be processed by a human.

[0161] Web browser plug-in

[0162] In some embodiments, the computer program includes a web browser plug-in (e.g., extension, etc.). In computing, a plug-in is one or more software components that add specific functionality to a larger software application. Makers of software applications support plug-ins to enable third-party developers to create abilities which extend an application, to support easily adding new features, and to reduce the size of an application. When supported, plug-ins enable customizing the functionality of a software application. For example, plug-ins are commonly used in web browsers to play video, generate interactivity, scan for viruses, and display particular file types. Those of skill in the art will be familiar with several web browser plug-ins including, Adobe® Flash® Player, Microsoft® Silverlight®, and Apple® QuickTime®. In some embodiments, the toolbar comprises one or more web browser extensions, add-ins, or add-ons. In some embodiments, the toolbar comprises one or more explorer bars, tool bands, or desk bands.

[0163] In view of the disclosure provided herein, those of skill in the art will recognize that several plug-in frameworks are available that enable development of plug-ins in various programming languages, including, by way of non-limiting examples, C++, Delphi, Java™, PHP, Python™, and VB .NET, or combinations thereof.

[0164] Web browsers (also called Internet browsers) are software applications, designed for use with network-connected computing devices, for retrieving, presenting, and traversing information resources on the World Wide Web. Suitable web browsers include, by way of non-limiting examples, Microsoft® Edge®, Mozilla® Firefox®, Google® Chrome, Apple® Safari®, Opera Software® Opera®, and KDE Konqueror. In some embodiments, the web browser is a mobile web browser. Mobile web browsers (also called microbrowsers, mini-browsers, and wireless browsers) are designed for use on mobile computing devices including, by way of non-

limiting examples, handheld computers, tablet computers, netbook computers, subnotebook computers, smartphones, and personal digital assistants (PDAs). Suitable mobile web browsers include, by way of non-limiting examples, Google® Android® browser, RIM BlackBerry® Browser, Apple® Safari®, Palm® Blazer, Palm® WebOS® Browser, Mozilla® Firefox® for mobile, Microsoft® Internet Explorer® Mobile, Amazon® Kindle® Basic Web, Nokia® Browser, Opera Software® Opera® Mobile, and Sony® PSP™ browser.

[0165] Software modules

[0166] In some embodiments, the platforms, systems, media, and methods disclosed herein include software, server, and/or database modules, or use of the same. In view of the disclosure provided herein, software modules are created by techniques known to those of skill in the art using machines, software, and languages known to the art. The software modules disclosed herein are implemented in a multitude of ways. In various embodiments, a software module comprises a file, a section of code, a programming object, a programming structure, or combinations thereof. In further various embodiments, a software module comprises a plurality of files, a plurality of sections of code, a plurality of programming objects, a plurality of programming structures, or combinations thereof. In various embodiments, the one or more software modules comprise, by way of non-limiting examples, a web application, a mobile application, and a standalone application. In some embodiments, software modules are in one computer program or application. In other embodiments, software modules are in more than one computer program or application. In some embodiments, software modules are hosted on one machine. In other embodiments, software modules are hosted on more than one machine. In further embodiments, software modules are hosted on a distributed computing platform such as a cloud computing platform. In some embodiments, software modules are hosted on one or more machines in one location. In other embodiments, software modules are hosted on one or more machines in more than one location.

[0167] Databases

[0168] In some instances, users query one or more databases to identify information about biological data in his or her dataset. For example, user may use an interface to display specific information about a variant, such as the variants role in cancer or other diseases. In some embodiments, the platforms, systems, media, and methods disclosed herein include one or more databases, or use of the same. In view of the disclosure provided herein, those of skill in the art will recognize that many databases are suitable for storage and retrieval of, for example, patient, photo, video, skin condition, visit, physician, and insurance information. In various embodiments, suitable databases include, by way of non-limiting examples, relational databases, non-relational databases, object oriented databases, object databases, entity-relationship model

databases, associative databases, XML databases, document oriented databases, and graph databases. Further non-limiting examples include SQL, PostgreSQL, MySQL, Oracle, DB2, Sybase, and MongoDB. In some embodiments, a database is Internet-based. In further embodiments, a database is web-based. In still further embodiments, a database is cloud computing-based. In a particular embodiment, a database is a distributed database. In other embodiments, a database is based on one or more local computer storage devices.

[0169] Databases may comprise information (e.g., annotations) regarding genetic variants. In some instances, databases provide information on somatic, germline, or somatic and germline variants. In some instances, a database comprises one or more of ClinVar, COSMIC, NCBI database of Genotypes and Phenotypes (dbGaP), gnomAD, 69 genomes from CGI, Personalized Genome Project, NCI Genomic Data Commons (GDC), cBioPortal, Intogen, and the Pediatric Cancer Genome Project. In some instances, databases provide information on variants related to cancer or other disease.

EXAMPLES

[0170] The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

[0171] **EXAMPLE 1**

[0172] User Account creation is provided as first option for new users onto the account showing the image in **FIG. 14A**. Fields (**1400**) provided within the form must be filled out entirely and agreements to terms and conditions made prior to email verification. Prior to gaining access to the platform, the user must review and approve the terms and conditions (**1410**), where access and review capability is shown in **FIG. 14B**. After completion of user access fields, a verification email is then sent providing a number token to be provided back to the platform to confirm email details. Tabs at the login screen allow for users to toggle between new account creation and login.

[0173] **EXAMPLE 2**

[0174] An administration console is provided to platform administrators to handle toggling access and assigning information to organizations and users (**1500**). This console is not available to individual accounts unless they have been grated Administrative access. The console provides information about the organizations with created accounts on the platform as shown in **FIG. 15A**.

[0175] Organization details are added by the BaseJumper Administrator as necessary (**1510**). Information about organizations contains pertinent company details about tracking that customer in third party platforms such as Zoho and SalesForce. Edits can be made to an organization by

selecting the Elipsis on the console (**1530**). An organization is allowed one or more workspaces, which are specific research groups established by the organization and typically conform to separate billable entities.

[0176] Selecting a specific workspace (**1540**) opens up a panel (**1550**) where all of an organization's workspaces are viewable as in **FIG. 15B**. Selecting the edit icon (**1560**) allows a user to specify data ingress locations such as BaseSpace Sequencing Hub (platform provided by Illumina®, **1570**) and a path available to a user on Amazon Web Services Simple Storage Serivce (s3) (**1580**).

[0177] Selecting the User tab on the console (**1500**) opens up a user dashboard where a platform admin can perform several functions. Queries can be made based on components of a user's name, their email or organization (**1590**). Selecting the Elipsis icon allows the Administrator to assign specific rights at the organization (**1592**) or Workspace (**1593**) level as in **FIG 15C.** This can also allow authorized users to have access to workspaces or organizations outside of their own, should collaborations or agreements between organizations authorize such an action.

[0178] **EXAMPLE 3**

[0179] Users are able to build projects (collections of biological samples) on the platform if they are authorized administrators of an organization or workspace. The former allows a user to create projects across workspaces of an organization and the latter allows project creation only within a workspace.

[0180] Project creation is done by selecting the New Project button (**1610**) at the top of a workspace's project listing in **FIG. 16A**.

[0181] The new project form in **FIG 16B** allows users to specify a name (**1620**) for a project (platform checks and reports if the name is redundant). Additional details about BioSkryb product lots (**1630**), associated genome (**1640**) and sequencing library information (**1650**) can be provided.

[0182] The user then determines which ingress method they want to use (**1660**).

[0183] A new window opens where a user can specify a specific project (**1670**) if BSSH (BaseSpace Sequence Hub®) in **FIG 16C** or specific characters of a sample name (**1680**) which can exist in BSSH or from a user's s3, as in **FIG 16D**. From there a box can be used to select all samples or individual sample boxes to select specific samples to be used for a project (**1690**).

[0184] **EXAMPLE 4**

[0185] The platform offers up analysis pipelines to process user's molecular data in a series of transformations automated by the platform.

[0186] Immediately after project creation, the platform alerts the user if they want to queue a pipeline immediately after downloading as in **FIG 17A**.

[0187] When a user decides to have a pipeline performed, the launch window in **FIG 17B** allows users to select what pipeline and what version of a pipeline to be run on the samples

[0188] Selecting a secondary analysis pipeline allows users to customize parameters and metadata used to provide additional instructions to steps in the pipeline as seen in **FIG 17C**. Specific tools/modules can be selected as well as details about chemistry.

[0189] When a user has customized their pipeline a final check is provided to the screen that has two tabs (**1700**) summarizes parameter details as well as what tab samples it will be used with (**FIG 17D**).

[0190] **EXAMPLE 5**

[0191] Pipelines are separated into secondary and tertiary analyses, determined by what inputs are provided into the pipeline.

[0192] Secondary pipelines start from sequencing data (typically FASTQs) and tertiary pipelines start from outputs of secondary analyses (typically aligned files, VCFs or counting information).

[0193] When selecting a project, the platform displays details (**1800**) about the biosamples of the project such as file size, total number of reads, read length BioSkryb LotID and upload date as shown in **FIG 18A**.

[0194] In that same window a user can select samples (**1810**) that will be considered in pipelines and after selection a user can select if secondary or tertiary pipelines are run via buttons (**1820**).

[0195] Tertiary analysis pipelines can only be run after secondary analysis pipelines and if a user selected a tertiary analysis (**1820**), it takes them to a window like in **FIG 18B**, where a set of secondary analyses that can be leveraged is provided. The user would then select a pipeline and advance to the next stage.

[0196] A tertiary workflow progress bar (**1830**) is provided at the top of the window to show the user what stage of execution the window portrays.

[0197] A window which shows which tertiary analysis pipeline can be selected (one or many) is displayed in **FIG 18C.**

[0198] **EXAMPLE 6**

[0199] After variants have been called by a secondary or tertiary analysis pipelines a user can select a variant table visualization as shown in **FIG 19A.**

[0200] Variant filtering visualization application has modules containing at least: a locus specific table showing variants called across all samples selected for pipeline output (**1900**) a

table for criteria to use for filtering (**1910**), an ability to select genotype details (**1920**) and a table export feature (**1930**).

[0201] Multiple databases to filter against are displayed in **FIG 19B.**

[0202] Users can provide specific lists of genes, amino acid (**1940**) changes or genomic region (**1950**) to filter the variant table against.

[0203] Specific cancer curated sets such as ClinVar, COSMIC or known cancer hallmarks can be leveraged to filter variant table specifically for known biomarker sets (**1960**).

[0204] Biomarkers characterized for specific protein effects or gene bodies can be leveraged in **1970.**

[0205] Polymorphic databases (such as but not limited to: dbSNP/gnomAD) where genotypes associated to populations can be leveraged and genotypes above or below certain minor allele frequencies can be selected (**1980**).

[0206] Individual sample genotypes, genotype frequencies and prevalence of alternate alleles can be shown by **FIG 19C** on the table if selections are made in **1920**.

[0207] A log of each filtering step is provided at the bottom of the application, highlighting one or more steps as shown in **FIG 19D.** A user can go back to previous steps or copy and paste the syntax to go directly to a previously used filtering combination.

[0208] **EXAMPLE 7**

[0209] For users that select to use copy number tertiary analyses an application (**FIG. 20**) can be leveraged to evaluate chromosomal gains and losses for samples in their project

[0210] Users can select some or all of the samples in their project by point and click or by searching for matching strings of sample sets they want to use (**2000**).

[0211] Read coverage is displayed for each selected samples in a scale-able frame and gains (red) and losses (blue) are highlighted as significantly different from expected copy (black) in **2010**.

[0212] A user can choose to select specific genomic regions of a sample (**2020**) and automatically initialize a table listing the genes that occupy that region (**2030**) for all selected samples in the application.

[0213] The user can use any of the tables filtering bars to select for specific features they are interested in such as (but not limited to): chromosomal start and stop positions, genes, samples or number of copies (**2040**).

[0214] At any point, a user can download the feature table (**2030**) with any or no filters applied (**2050**).

[0215] **EXAMPLE 8**

[0216] To handle expression-based molecular information, such as those that come from transcriptomic or proteomic sequencing data an application is created that leverages several modules showing different facets of the data

[0217] As with all applications on the platform, manipulations of each of the modules creates a custom view into the data of a project and if the user should so choose, they can export the manipulation settings or import a previously used manipulation schema as shown in **FIG. 21A.**

[0218] Initial details about the samples of a project are outlined in **FIG. 21B.**

[0219] Cellular phenotype data, generated by analytical pipelines on the platform are shown (**2100**), which highlights predicted identification feature such as (but not limited to): cell cycle phase, progenitor type, tissue and tumor and matched normal of cancer references like The Cancer Genome Atlas (TCGA).

[0220] This table (**2100**) can be exported but also a user is allowed to add and import additional classification columns (**2110**), such as those identifying biomarker states from investigations of other molecular types on the platform.

[0221] The application can also be configured to display relative expression values at the isoform-level (specific coding combinations of a gene) in **2120.**

[0222] A user can select which samples to remove from representation in the modules of the application by selecting or searching for sample names in **2130**.

[0223] To account for prevalence of expression at the gene or sample level, a user can select observational minimums to reduce sparsity of the results displayed (**2140**).

[0224] A read count figure is initially shown to identify specific samples that may have abnormally high (or low) sequencing information that may skew graphical representations (**2150**). Samples removed from **2130** are shown in contrast to approved samples

[0225] Heatmaps, showing correlation of expression at the gene or sample level is displayed in **FIG. 21C**

[0226] Users are able to select a variety of gene and sample-level normalizations, along with one or more labels from sample phenotypic information provided in **2100.**

[0227] A user can chose gene sets they wish to display on the heatmap from a variety of sources, such as: top variable genes (auto-calculated by the application), differentially expressed genes, custom gene lists or known biological pathways (**2170**).

[0228] One of the platform's tertiary analysis capabilities (differential expression) is performed as a module within the application as shown in **FIG. 21D.**

[0229] Specific groups can be selected to compare expression levels against as well as multiple statistical methods to perform the analysis are customizable (**2180**).

[0230] Results of differential expression performed in **FIG. 21D** are subsequently shown in a separate module with specific significance statistics for each identified gene is shown (**2190**) and able to be downloaded.

[0231] **EXAMPLE 9**

[0232] Other modules from the expression application include projections of expression data into two dimensional space.

[0233] The first allows the user to view a principal components analysis (PCA) of their data that leverages prior sample exclusion (as shown in **2130**) and is shown in **FIG. 22A**.

[0234] Samples can be colored by the phenotypic labels (**2200**) provided by the client in **2100.**

[0235] Gene or sample-level normalizations can be toggled on or off (**2210**), depending on specific numerical characteristics of the project's samples.

[0236] The user is able to select specific components to display (**2220**), highlighting different views of potential sample grouping along specific contributions of variability. Percent contribution of sample variability is shown for each component selected

[0237] For more complex and higher sample volume projects a module is provided that shows the Uniform Manifold Approximation and Projection (UMAP) representation of the data as shown in **FIG. 22B.** Color can be provided randomly to sample IDs or use the common phenotypic information used throughout the application.

[0238] Statistical details used to govern the UMAP are provided to the user to customize representation based on data such as number of neighbors and minimal distances (**2230**), calculated by the selectable distance function in **2240**.

[0239] To get additional context of cluster or group separation of cells in the UMAP, the user can display results in a rendered 3 dimensional view for the main graphic shown in **FIG 22B.**

[0240] **EXAMPLE 10**

[0241] A circos plot application is provided to users that can overlay multiple molecular data types in genomic context as shown in **FIG. 23A.**

[0242] Users are able to select one or multiple samples (**2300**), which can generate an aggregated plot or multiple plots per samples in the application.

[0243] Specific pipeline results, from secondary or tertiary pipelines run in a project can be selected to be represented on the plot (**2130**).

[0244] Specific chromosomes are also able to be selected (**2320**) as well to determine what region(s) are displayed on the plot.

[0245] The circos plots on the application (**2330**) display selected data types and chromosomes determined by the user.

[0246] As the user selects specific features displayed in the plot, a box highlighting details of that feature, along with genomic coordinates is displayed (**2340**). These coordinates can be used for modifying other visualization applications on the platform.

[0247] As with all graphics and tables on the platform, a user can save images of the resulting image to be used in their research (**2350**).

[0248] **EXAMPLE 11**

[0249] A web-based user interface was created which allowed rapid visualization of genomic data, having the architecture of **FIG. 1B**. A user logged into the interface, which displayed previous projects analyzed using the interface **FIG. 1C**. The user then loaded sample data comprising genomic sequences for 10 single cell experiments, which were available for viewing in a genome browser (**FIGS. 2D-2E**). Variants were annotated based on querying databases (e.g., COSMIC) **FIG. 2C**. Lineage trees for variants were generated and presented to the user (**FIGS. 3B** and **4A**). Information for statistics of individual cells and populations and whole-genome sequencing quality metrics were accessed through the web-based interface. A circos plot representing clinically relevant mutations found in the samples was generated (**FIG. 5A-5B**) in less than 100 ms (**FIG. 6**). Additional circos plots were generated with controls for which chromosomes to display and different types of mutations (controlled by filters) at **FIG. 7**.

[0250] **EXAMPLE 12**

[0251] Following the general methods of Example 1, a clinician obtains cells from a tumor biopsy and subjects the cells to single-cell sequencing. After generating sequencing files for the cells, the clinician imports the files using the web-based user interface, identifies clinically relevant mutations using interface, and prescribes specific therapies based on the presence of absence of these mutations. In some instances, driver mutations identified in a minority of cells are considered such that a combination treatment is prescribed.

[0252] **EXAMPLE 13**

[0253] Applications that processes potentially sensitive genomic data, have been designed with secure best practices. NIST Cybersecurity Framework and HIPAA as the basis for its data security program has been utilized. The NIST Cybersecurity Framework is a set of guidelines for mitigating organizational cybersecurity risks published by the US National Institute of Standards and Technology. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is a federal law that requires the creation of national standards to protect sensitive patient health information from being disclosed without the patient's consent or knowledge. Following best practices, an application described herein aims to ensure the

confidentiality, integrity, and availability of information within, using the following mechanisms to provide for the confidentiality, integrity, and availability of information:

[0254] All interactions with the data and files are logged using AWS CloudTrail.

[0255] All users selected a password conforming to the following guidelines:

Minimum of eight characters;

Contains at least 1 number;

Contains at least 1 special character (^ $ * . [ ] { } ( ) ? - " ! @ # % & / \ , > < ' : ; | _ ~ ` + =);

Contains at least 1 uppercase letter; and

Contains at least 1 lowercase letter.

[0256] Access permissions were managed, incorporating the principles of least privilege and separation of duties. Being able to access data does not automatically give a user authorization to view that data.

[0257] Network shares/data stores/repositories were reviewed to ensure that sensitive, confidential, or regulated data is not mistakenly saved in locations accessible by unauthorized users. For situations where access cannot be limited, tools must be utilized to log activity. User activity was reviewed to identify any access beyond the minimum required. Any unauthorized activity results result in discipline.

[0258] Network integrity was protected, incorporating network segregation where appropriate. If it is determined that segmentation is necessary, specific methods will be approved and adopted. Periodic internal vulnerability tests must be conducted and documented to validate controls are effective.

[0259] Log-in procedures limited the number of unsuccessful log-in attempts, after which a user must contact the information system administrator to have his or her password reset.

[0260] Network communication was encrypted by SSL certificates.

[0261] Access control was governed by AWS IAM, which provides role-based access control to the databases and files.

[0262] Files were structured so that organizations cannot access files outside of their organization. Database records were similarly structured so that only authorized users within an organization or workspace can access them.

[0263] Data was encrypted at rest. Data was encrypted when it is stored on a network that might be accessible by unauthorized individuals. Information that can be used to alter or defeat security measures was encrypted.

[0264] Data-in-Transit is protected. Secure transmission protocols include encryption of data and e-mails; VPN's, TLS links, SSL links, and similar technologies.

[0265] User authentication was provided by AWS Cognito, which is a managed authentication service similar to Okta or Auth0. Automation was used to automatically create groups for workspaces and Password organizations. Users were then allocated to these groups which provide access to the files and data.

[0266] Security alerts and findings were routed to a centralized security information and event system for proactive monitoring and remediation.

[0267] AWS GuardDuty, a tool that uses AI to detect anomalous behavior, was used proactively to monitor the accounts for suspicious activity.

[0268] Automatic log-off was implemented for protection against data leaks. Automatic controls to log users off after 15 minutes of inactivity was enabled. After being automatically logged off, a user must re-enter his or her username and password to resume the interrupted activity. Users may not disable this automatic log-off feature.

[0269] To protect against data leakage: 1) the workforce members and third-party stakeholders are forbidden from posting any organization information to social media sites without express permission from an organization executive; 2) workforce members and third-party stakeholders are forbidden from saving data to personally owned devices, personally maintained cloud services, and backup services.

[0270] Workforce members and third-party stakeholders were forbidden from sending organization data to personal mail accounts, or to anyone without authorized business purposes.

[0271] The development and testing environment(s) were separate from the production environment for data integrity and ongoing developmental tasks such as testing patches and updates, diagnosing devices that may contain malware, and staging equipment for users.

[0272] Configuration change control processes were in place. For example: 1) determining the types of changes to the information system that are configuration controlled; 2) approving configuration-controlled changes to the system with consideration for security; 3) documenting approved configuration-controlled changes to the system; 4) retaining and reviewing records of configuration-controlled changes to the system; 5) auditing activities associated with configuration-controlled changes to the system; and 6) coordinating and providing oversight for configuration change control activities through change request forms that must be approved by a Security Officer.

[0273] Backups of information were conducted, maintained, and tested periodically.

[0274] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now

occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

1. A system for visualizing biological data comprising:

a device comprising at least one processor and instructions executable by the at least one processor to provide a first application configured to perform operations comprising:

i. accessing one or more datasets comprising biological data;

ii. generating a visual representation of the one or more datasets, wherein the data is generated rapidly enough for live user interpretation.

2. The system of claim 1, wherein the visual representation is generated in no more than a minute.

3. The system of claim 1, wherein the first application comprises one or more modules.

4. The system of claim 3, wherein the one or more modules comprise one or more of plot, tree, cells, counting, or sequencing metrics modules.

5. The system of claim 4, wherein the sequencing metrics module comprises quality metrics.

6. The system of claim 5, wherein the biological data comprises one or more of molecular data.

7. The system of claim 6, wherein the molecular data comprises one or more of genomic data, proteomics data, transcriptomics data, and methylomics data.

8. The system of claim 6, wherein the molecular data comprises two or more of genomic data, proteomics data, transcriptomics data, and methylomics data.

9. The system of claim 7, wherein the genomic data comprises one or more genomic variants.

10. The system of claim 9, wherein the genomic variants comprise one or more of somatic, germline, SNP, or indels.

11. The system of claim 1, wherein one or more datasets corresponds to a project.

12. The system of claim 11, wherein a project comprises a virtual collection of samples

13. The system of claim 1, wherein the system comprises at least 10 datasets.

14. The system of claim 1, wherein the system comprises at least 1000 datasets.

15. The system of claim 1, wherein the system comprises at least 100,000 datasets.

16. The system of claim 1, wherein the one or more datasets are each representative of a single cell.

17. The system of claim 16, wherein the one or more datasets represent at least 30 cells.

18. The system of claim 16, wherein the one or more datasets represent at least 100 cells.

19. The system of claim 16, wherein the one or more datasets represent at least 1000 cells

20. The system of claim 16, wherein the one or more datasets represent at least a 100,000 cells

21. The system of claim 1, wherein the one or more datasets comprise at least 1 biomarker.

22. The system of claim 21, wherein the one or more datasets comprise at least 100 biomarkers.

23. The system of claim 1, wherein the system comprises no more than four computer processors.

24. The system of claim 1, wherein the system comprises at least four processors.

25. The system of claim 24, wherein the system comprises at least 16 processors.

26. The system of claim 1, wherein the system comprises a platform for data security comprising an access control for one or more users; a security framework; and molecular data from an individual or biological species

27. The system of claim 26, wherein the back end module is configured to perform user authentication and registration.

28. The system of claim 26, wherein access to genomic data is limited to authorized users who are a part of authorized organizations.

29. The system of claim 1, where the system is accessible by at least 2 users concurrently.

30. The system of claim 29, where the system can be accessible by at least 20 users concurrently.

31. The system of claim 29, where the system can be accessible by at least 200 users concurrently.

32. The system of claim 26, wherein the security framework comprises an NIST cybersecurity framework.

33. The system of claim 26, wherein the platform complies with HIPAA standards relating to the individual.

34. The system of claim 1, wherein the system is configured for a user to define one or more analytical workflows.

35. The system of claim 34, wherein the one or more analytical workflows operate on molecular data defined in a project.

36. The system of claim 34, wherein the one or more analytical workflows are executed by one or more pipeline modules.

37. The system of claim 36, wherein the pipeline module is configured to perform complex genomic analytical transformations.

38. The system of claims 37, wherein the pipeline module is configured to run genomics analysis tools to extract biomarkers from one or more sequencing data.

39. The system of claim 36, wherein the system is configured for a user to queue and/or order pipelines to be run on a dataset.

40. The system of claim 39, wherein the system is configured for a user to queue and/or order pipelines to be run on a dataset that does not require manual intervention.

41. The system of claim 36, wherein the system comprises a pipeline module configured to identify biomarkers present in the genome.

42. The system of claim 36, wherein biomarkers comprise one or more of single nucleotide variants, insertions and deletions.

43. The system of claim 42, wherein variants across every sample of a project are aggregated to a single table.

44. The system of claim 43, wherein the aggregation comprises joint genotyping.

45. The system of claim 34, wherein the pipeline module is configured to extract annotations and variants from an input file into a database.

46. The system of claim 45, wherein the database comprises COSMIC, ClinVar, Ensembl, or other genomic annotation databases.

47. The system of claim 34, wherein the pipeline module is configured to identify structural biomarkers present in a genome.

48. The system of claim 47, wherein structural biomarkers comprise variations in chromosomal copy number.

49. The system of claim 47, wherein structural biomarkers comprises one or more of transversions, inversions, tandem repeats, translocation, and element insertion.

50. The system of claim 34, wherein the pipeline module is configured to identify genomic arrangement information specific to classes of genes.

51. The system of claim 50, wherein the genomic arrangement information comprises one or more of major histocompatibility complexes, microsatellite instability, tumor mutational burden and immune repertoire.

52. The system of claim 34, wherein the pipeline module is configured to one or more run transcriptome analysis tools.

53. The system of claim 52, wherein transcriptome analysis tools are configured to extinct counts of isoforms from one or more sequencing files.

54. The system of claim 52, wherein transcriptome analysis tools are configured to identify fusion events from one or more sequencing files

55. The system of claim 52, wherein transcriptome analysis tools are configured to identify nucleotide variation in expressed genes from one or more sequencing files.

56. The system of claim 34, wherein the pipeline module is configured to run methylation analysis tools from one or more sequencing files.

57. The system of claim 56, wherein the methylation analysis tools are configured to extract methylation status from one or more sequencing files.

58. The system of claim 57, wherein methylation status comprises one or more of regions of methylated cytosine, such as mC, hmC and hhmC from one or more sequencing files.

59. The system of claim 34, wherein the pipeline module is configured to run proteomics analysis tools from one or more sequencing files.

60. The system of claim 59, wherein the proteomics analysis tools provide proteomic data.

61. The system of claim 60, wherein the proteomic data comprises sequence counts of identified native conformationally altered proteins.

62. The system of claim 1, comprising a platform for visualizing molecular data comprising:

a device comprising at least one processor and instructions executable by the at least one processor, wherein the instructions comprise:

a front end module comprising a graphical user interface; a security layer configured to reduce unauthorized access; a back end module; and one or more pipeline modules configured to generate a visual representation.

63. The system of claim 62, wherein the visual representation comprises at least one plot and genomics scale.

64. The system of claim 62, wherein the visual representation displays one or more biomarkers.

65. The system of claim 62, wherein the one or more biomarkers comprises specific molecular states associated with biological states of a sample.

66. The system of claim 62, wherein the system is configured for a user to select one or more applications.

67. The system of claim 62, wherein the one or more applications are configured for visually representing specific types of molecular data.

68. The system of claim 62, wherein the molecular data comprises one or more of genome, transcriptome, methylome and proteome.

69. The system of claim 62, where the graphical user interface is configured for a user to select one or more samples processed through a pipeline on the platform to be represented in a visualization application.

70. The system of claim 62, wherein the back end module is configured to encode data for one or more visualizations.

71. The system of claim 62, wherein visualization modules comprises a query engine.

72. The system of claim 71, wherein the query engine comprises one or more of Vaex and AG Grid.

73. The system of claim 62, wherein the back end module is configured to perform one or more calculations or filtering commands.

74. The system of claim 73, wherein one or more calculations or filtering commands are performed on biomarkers.

75. The system of claim 73, wherein one or more calculations or filtering commands are performed on a project or visualization application.

76. The system of claim 69, wherein a visualization application receives (as input) an output generated from one or more pipeline modules performed on the system.

77. The system of claim 74, wherein a visualization application is configured for a user to view biomarkers.

78. The system of claim 77, wherein the biomarkers are identified through the platform pipelines against one or more of an annotation and genomic reference tables.

79. The system of claim 51, where a visualization application comprises of one or more modules which represent one or more characteristics of a biological data type.

80. The system of claim 51, where a user is able to directly manipulate visual representations in a visualization application.

81. The system of claim 80, wherein the visual representation comprises one or more of feature selection, feature filtering, importing gene lists, and external biomarker results.

82. The system of claim 51, where configurations or manipulations of a visual application are exported or imported.

83. The system of claim 82, wherein configurations or manipulations control a layout of other visualization applications within the platform.

84. The system of claim 51, wherein chromosomal coordinates are leveraged to highlight various information across multiple visualization applications.

85. The system of claim 51, wherein biological groupings, label or group samples within a visualization module within an application.

86. The system of claim 85, wherein biological groupings comprises cellular phenotypes.

87. The system of claim 51, wherein a user is able to view biomarker information within genomic context, leveraging one or more genomic loci, along with associated annotation.

88. The system of claim 51, wherein the system depicts one or more modules within a visualization application for transcription.

89. The system of claim 88, wherein a visualization application depicts one or more of heatmaps, principal components analysis, differential expression, and alternative transcriptional mapping.

90. The system of claim 51, wherein a visualization application is configured for a user to view changes in copy number or structural changes within genome scale.

91. The system of claim 51, wherein a visualization application is configured for a user to view annotated information of changes in copy number or structural changes within the genome scale.

92. The system of claim 51, wherein the system is configured for syncopation of molecular data.

93. The system of claim 92, wherein syncopation of molecular data comprises managing and querying for one or more analysis tools.

94. The system of claim 92, wherein molecular data comprises data objects derived from the same cell.

95. The system of claim 92, wherein syncopated molecular information is visualized together on the same application within the platform.

96. The system of claim 95, wherein syncopated molecular information is visualized together using one or more of a Circos plots, genome viewers, and graph networks.

97. The system of claim 51, wherein the system is configured to associate molecular information with biological groupings.

98. The system of claim 97, wherein the system is configured to associate molecular information with biological groupings using one or more statistical methods.

99. The system of claim 98, wherein one or more statistical comprises differential expression (transcriptomic/proteomic), genome-wide association studies (genomic) and quantitative trait loci calculations using molecular information.

100. The system of claim 1, wherein the system is configured to export information in the system contained in one or more objects.

101. The system of claim 100, wherein the one or more objects comprise files, images, pipeline and visualization configurations, biomarker lists or projects.

102. The system of claim 100, wherein information is exported as one more files within a project through a file explorer embedded directly on the platform.

103. The system of claim 100, wherein configurations of visualization applications is exported.

104. The system of claim 100, wherein configurations of visualization applications is exported in a format readable by the platform's visualization applications.

105. The system of claim 100, wherein the system is configured for export of images of visual modules or applications which preserve manipulations performed by the user.

106. The system of claim 100, wherein the system is configured for specific output directories.

107. The system of claim 100, wherein a user has rights and access to place information in the output directory.

108. The system of claim 100, wherein the system is configured for export of a project or pipeline to specified output locations.

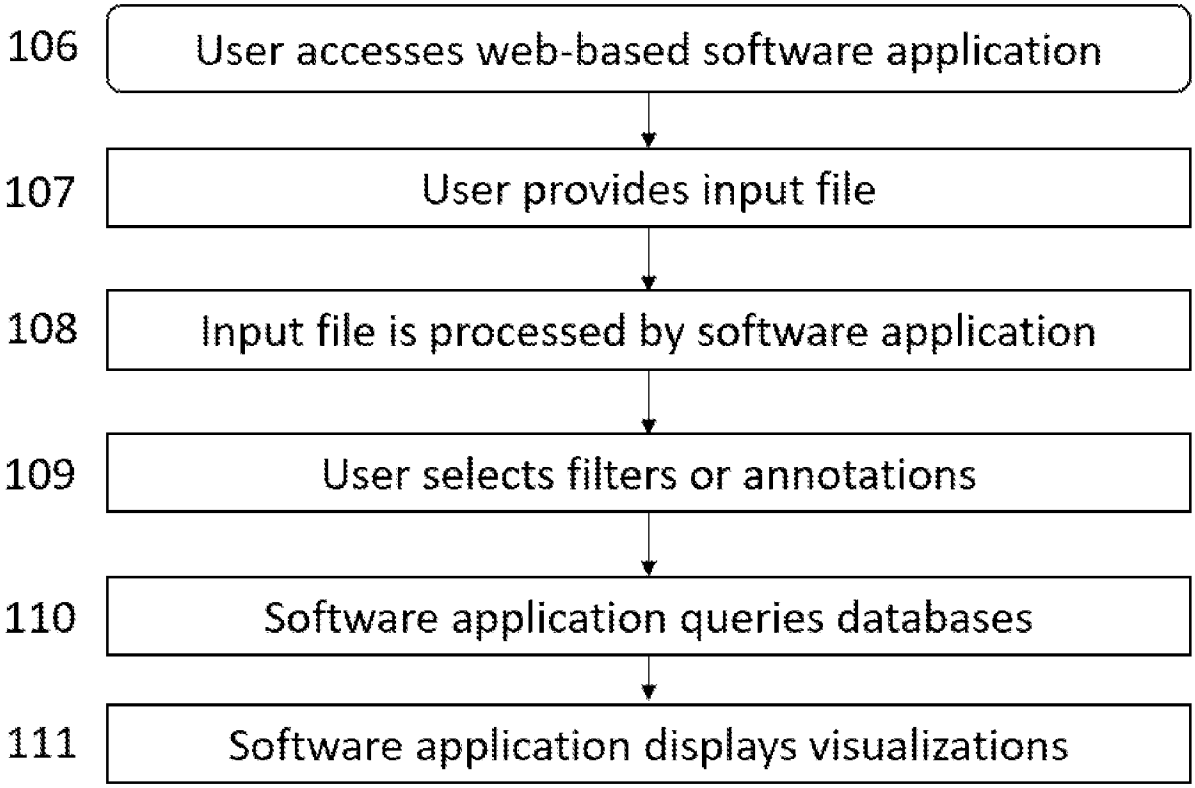

106
User accesses web-based software application
107
User provides input file
108
Input file is processed by software application
109
User selects filters or annotations
110
Software application queries databases
111
Software application displays visualizations

FIG. 1A

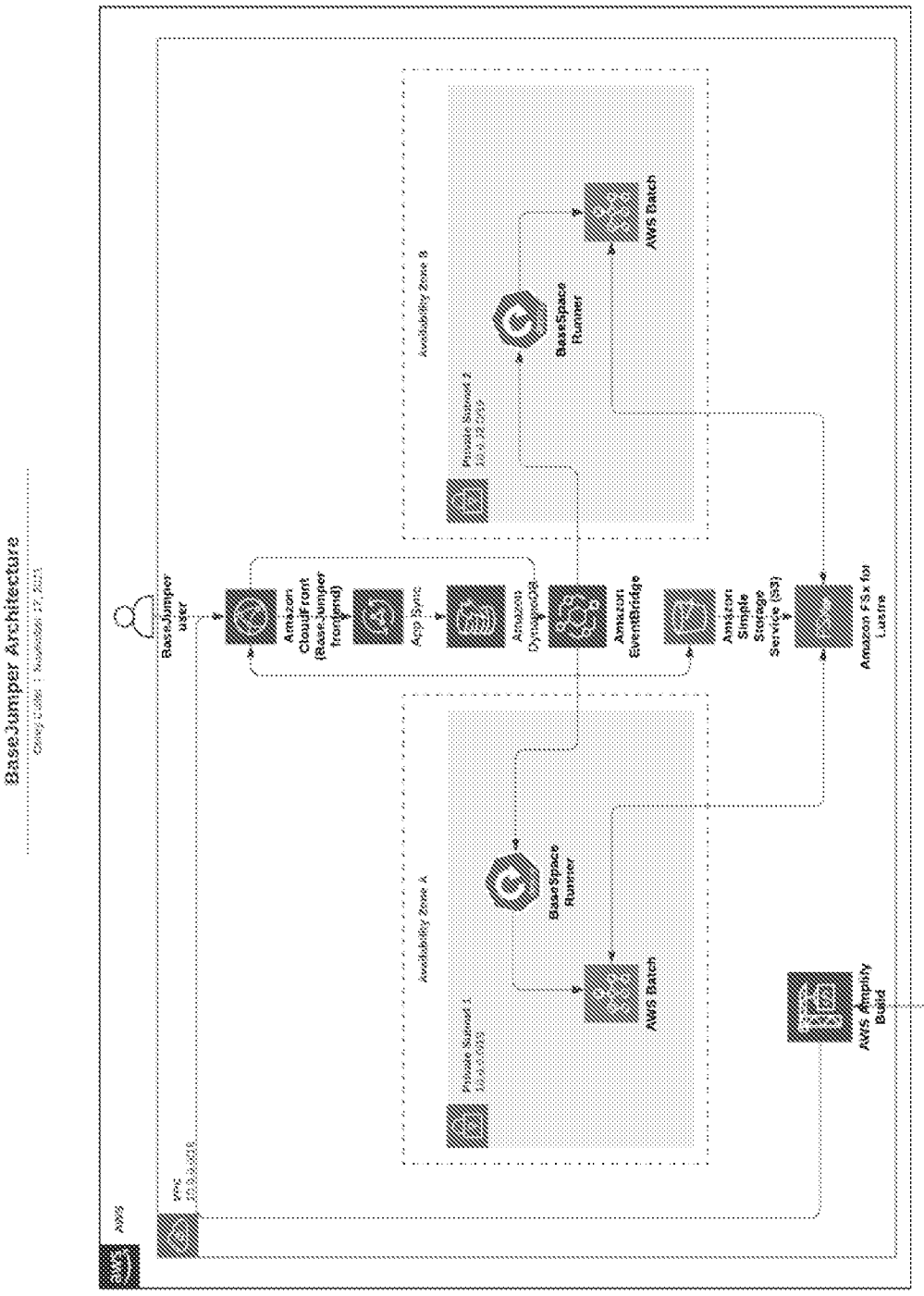

BaseJumper Architecture
BaseJumper user
Amazon CloudFront (BaseJumper frontend)
AWS AppSync
Amazon DynamoDB
Amazon EventBridge
Amazon Simple Storage Service (S3)
Amazon FSx for Lustre
Availability Zone A
Availability Zone B
BaseSpace Runner
AWS Batch
AWS Amplify Build
GitHub
Development Workflow
Developers

FIG. 1B

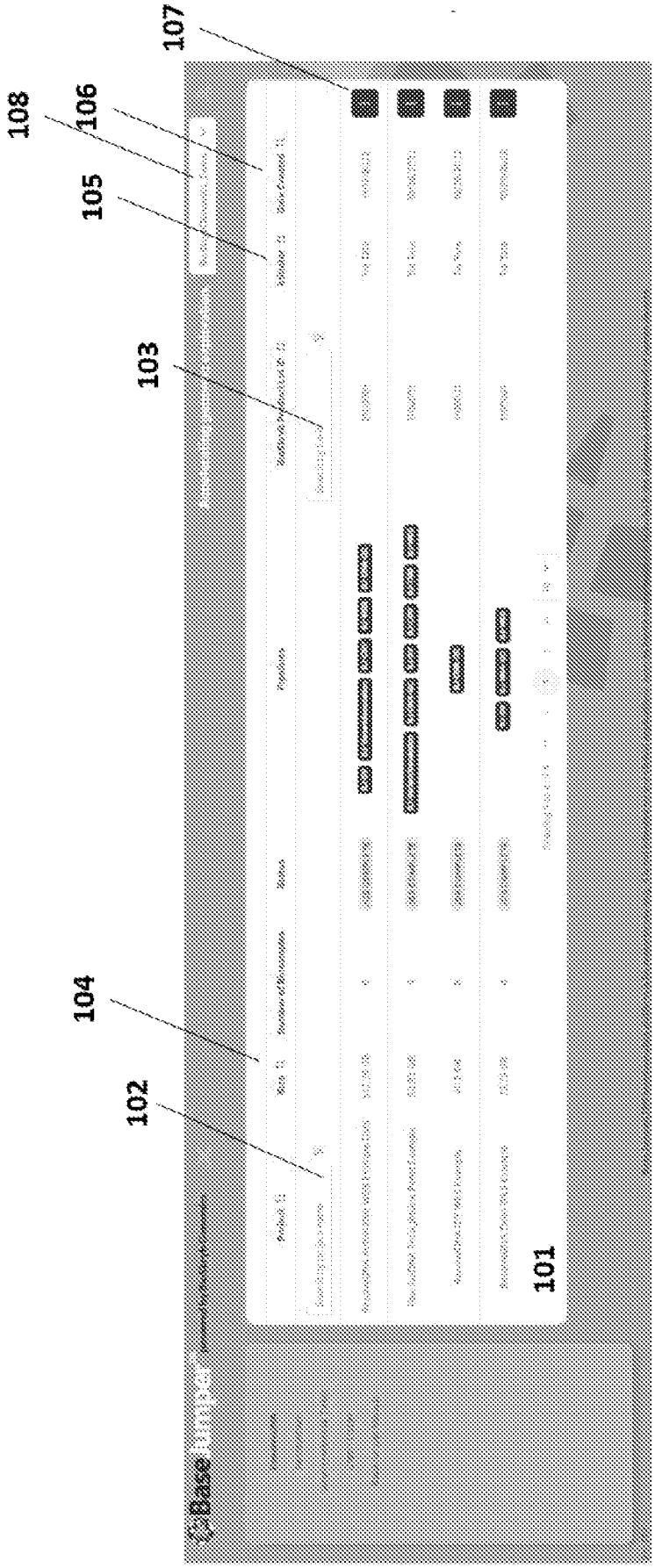


FIG. 1C

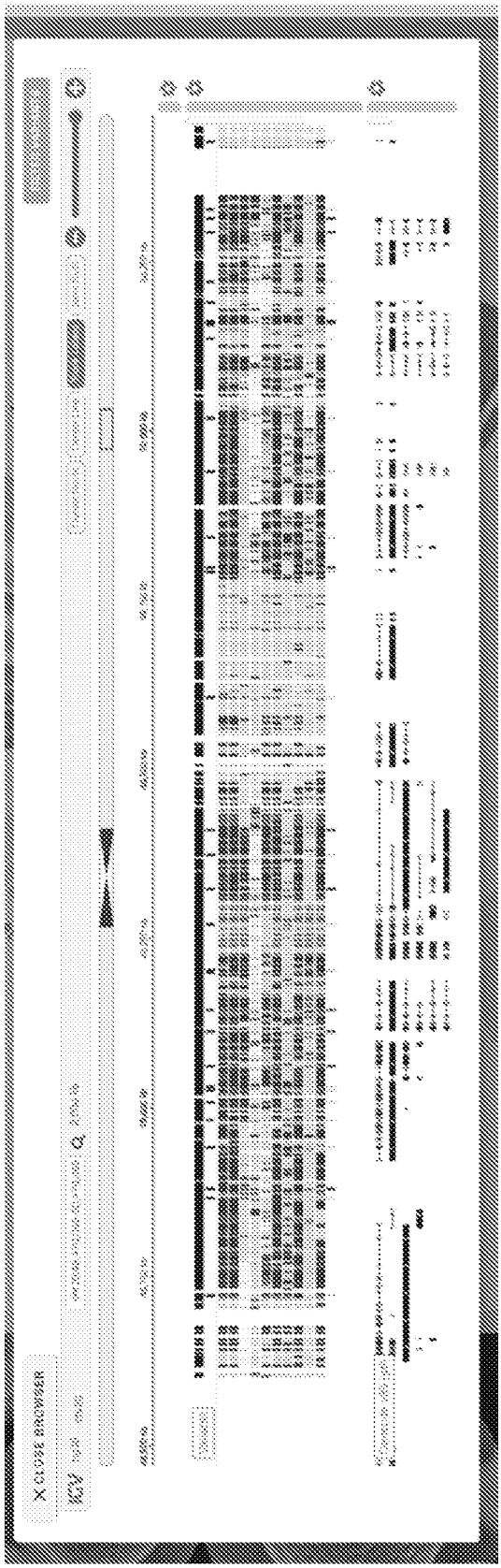


FIG. 2A

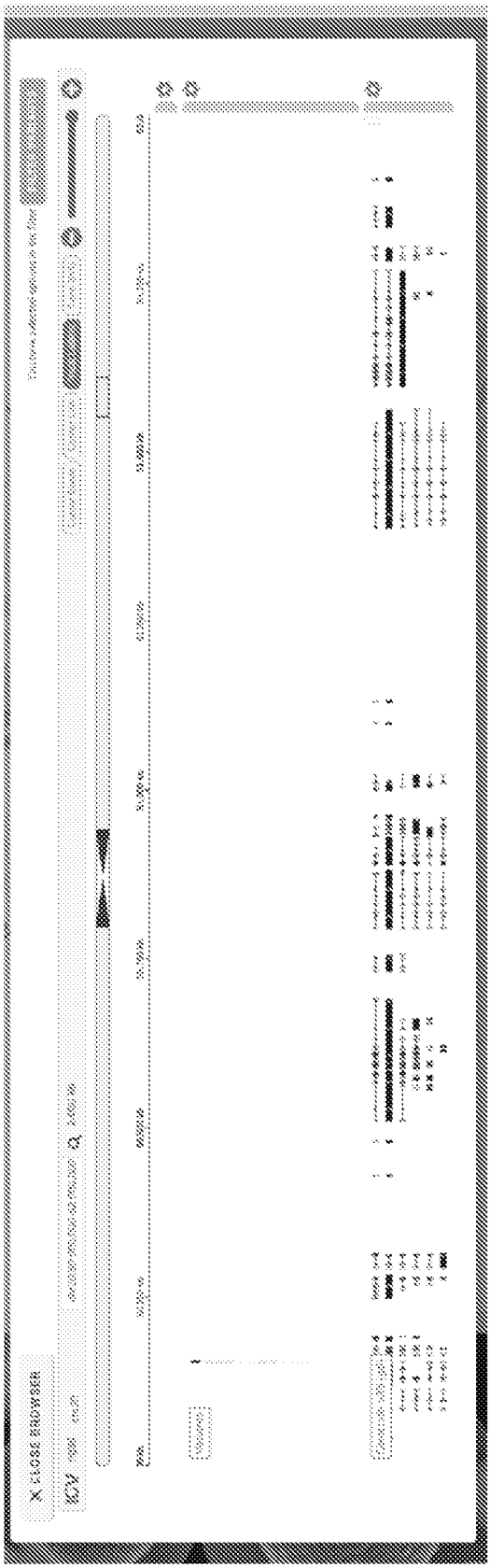


FIG. 2B

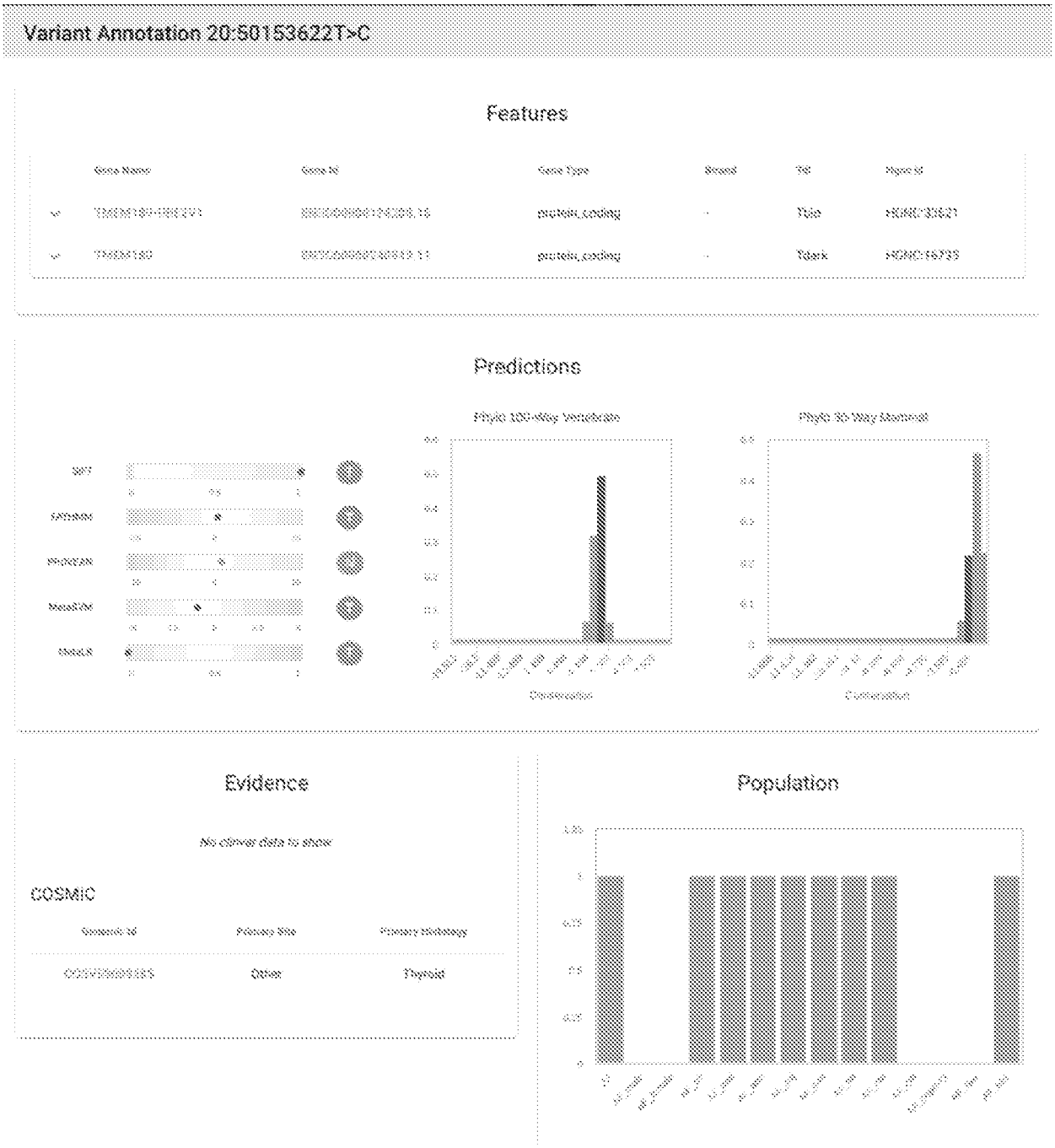

Variant Annotation 20:50153622T>C
Features
Predictions
Evidence
No clinvar data to show
COSMIC
Primary Site
Thyroid
Population

FIG. 2C

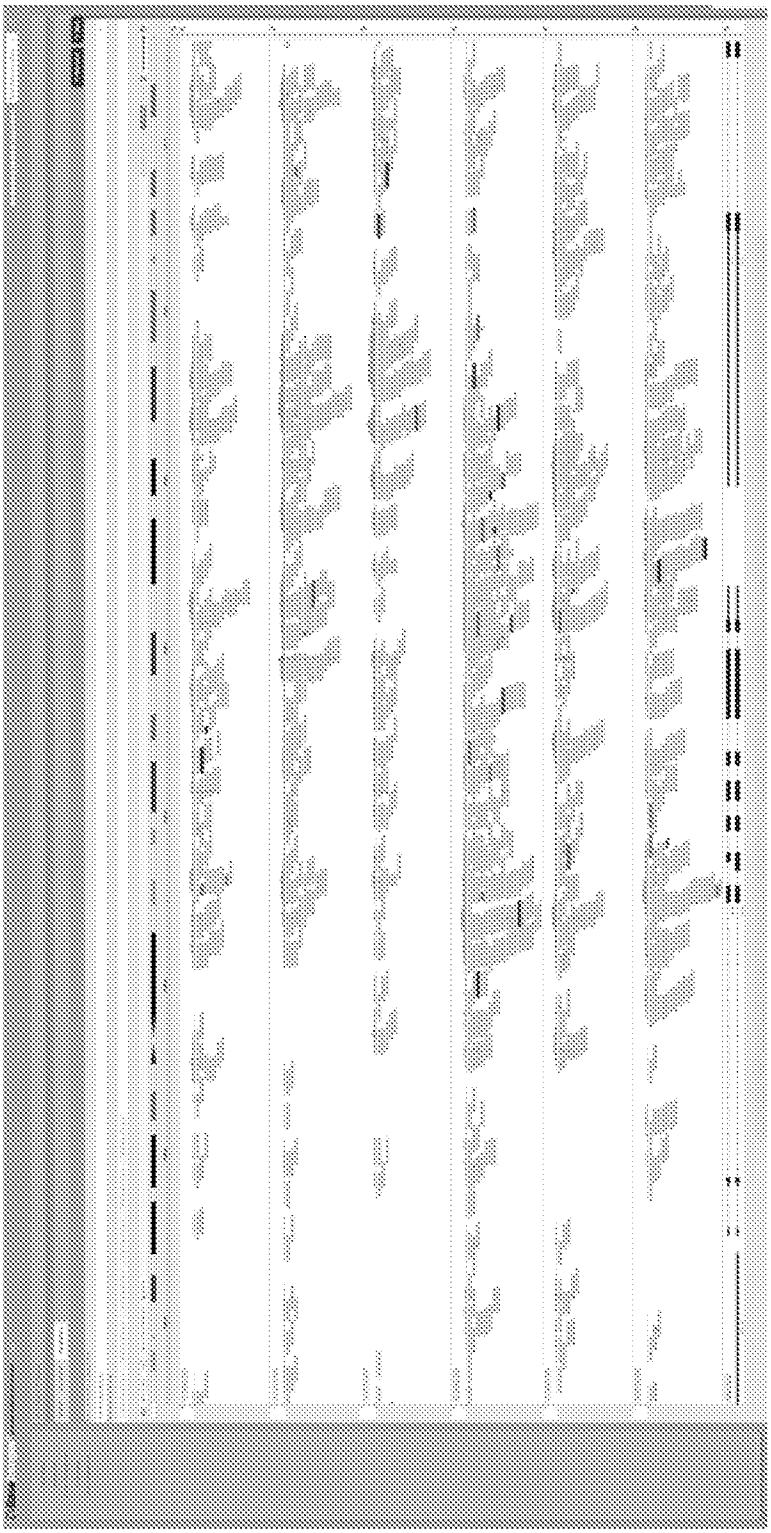


FIG. 2D

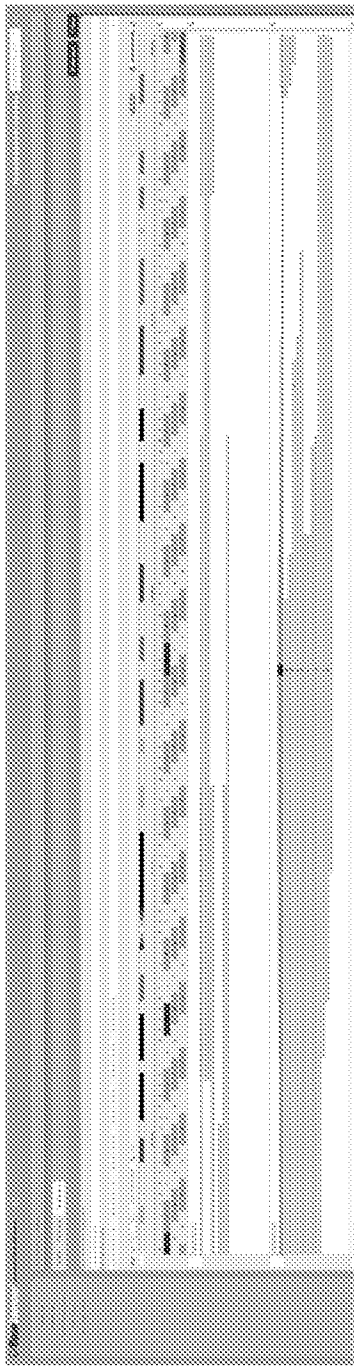


FIG. 2E

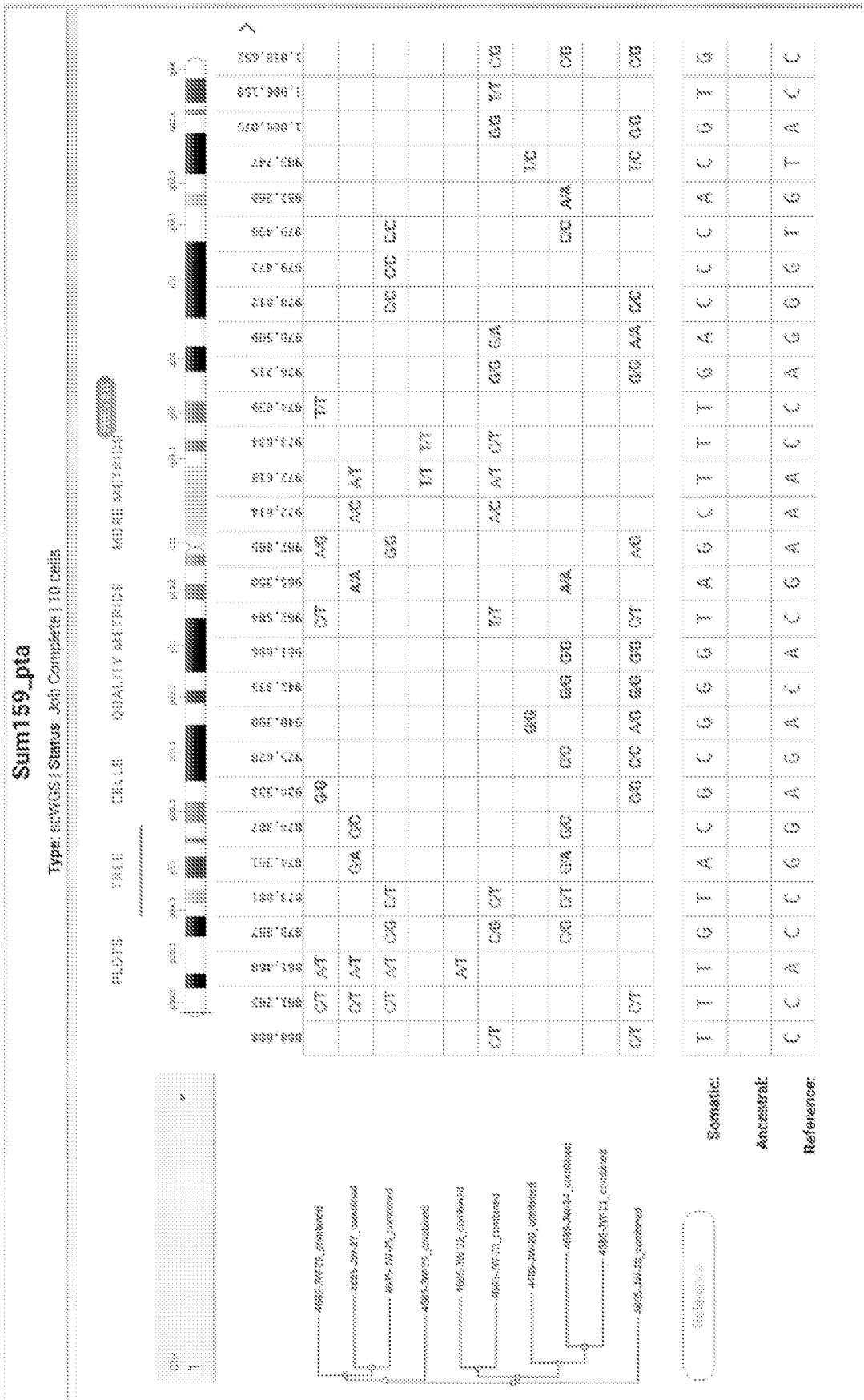

Sum159_pta
Type: scWGS | Status: Job Complete | 10 cells
Somatic:
Ancestral:
Reference:

FIG. 3A

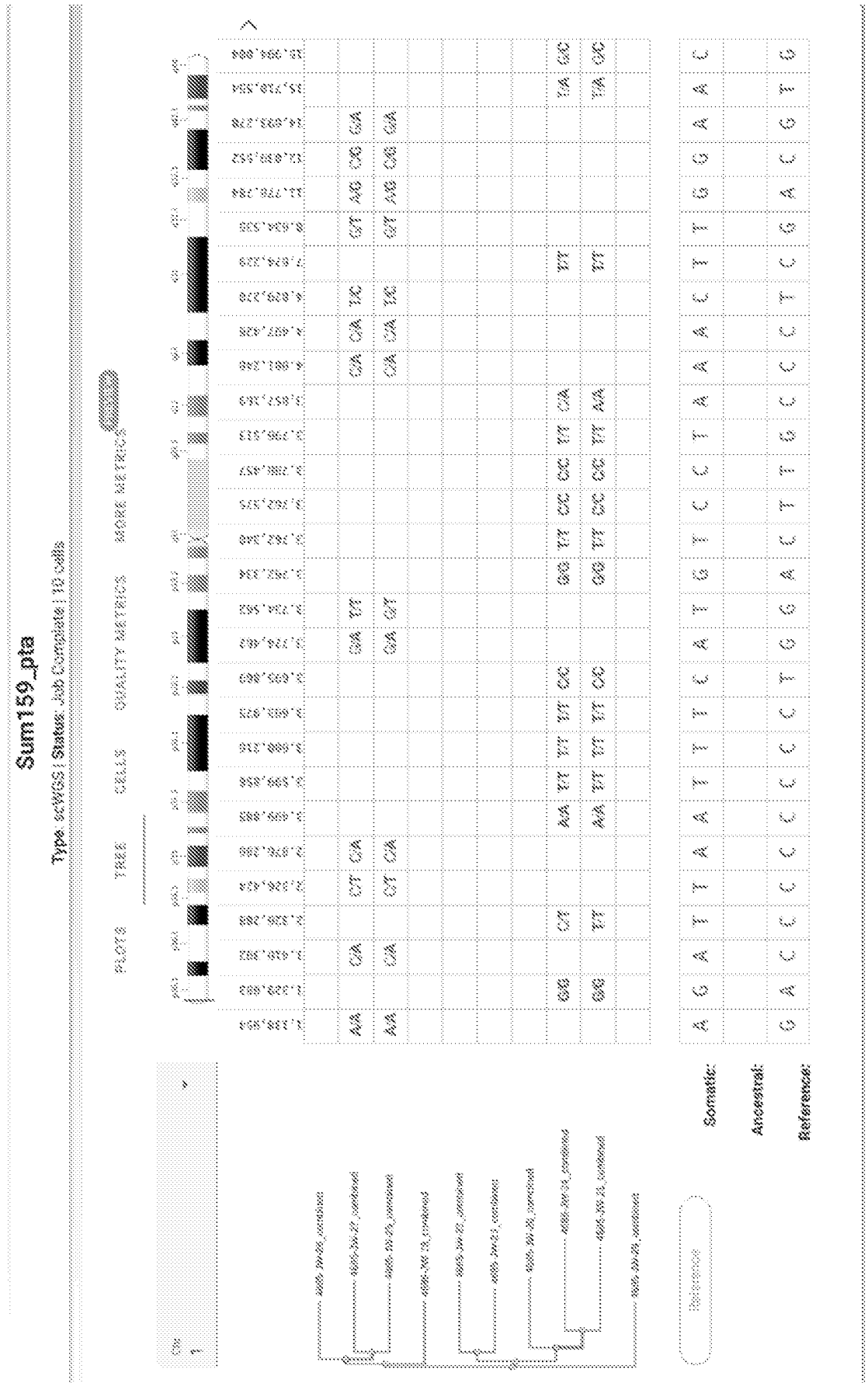

Sum159_pta
Type: scWGS | Status: Job Complete | 10 cells
PLOTS
TREE
CELLS
QUALITY METRICS
MORE METRICS
Somatic:
Ancestral:
Reference:

FIG. 3B

Sum159_pta
Type: scWGS ; Status: Job Complete | 10 cells
Inter-Cell Distance
Dimension 2
Dimension 1
Cell Statistics
Search
Cell
Genome
Genome
Sum
Index
Min
Median
Low
4685_JW-27_combined
52,062
2,819,289
2,871,351
0
1,388,887
1,388,826
782,388
4685_JW-26_combined
53,253
2,844,791
2,898,044
0
2,074,968
1,074,976
426,861
4685_JW-28_combined
63,249
2,253,642
2,289,891
0
990,529
1,241,005
954,636
4685_JW-21_combined
37,615
1,985,648
2,023,263
0
880,923
884,559
1,064,693
4685_JW-20_combined
49,222
2,592,096
2,643,318
0
1,230,623
1,264,652
953,680
4685_JW-23_combined
41,304
2,601,867
2,643,171
0
1,332,946
1,239,444
969,502
4685_JW-24_combined
51,175
2,274,975
2,326,151
0
1,015,318
977,118
988,435
4685_JW-22_combined
50,188
2,816,156
2,866,355
0
1,461,063
1,237,149
821,947
4685_JW-29_combined
60,208
1,835,426
1,885,634
0
812,283
1,054,496
872,168
4685_JW-25_combined
33,032
2,551,305
2,554,528
0
1,575,448
1,190,929
661,836
Rows per page:
1-10 of 10

FIG. 4A

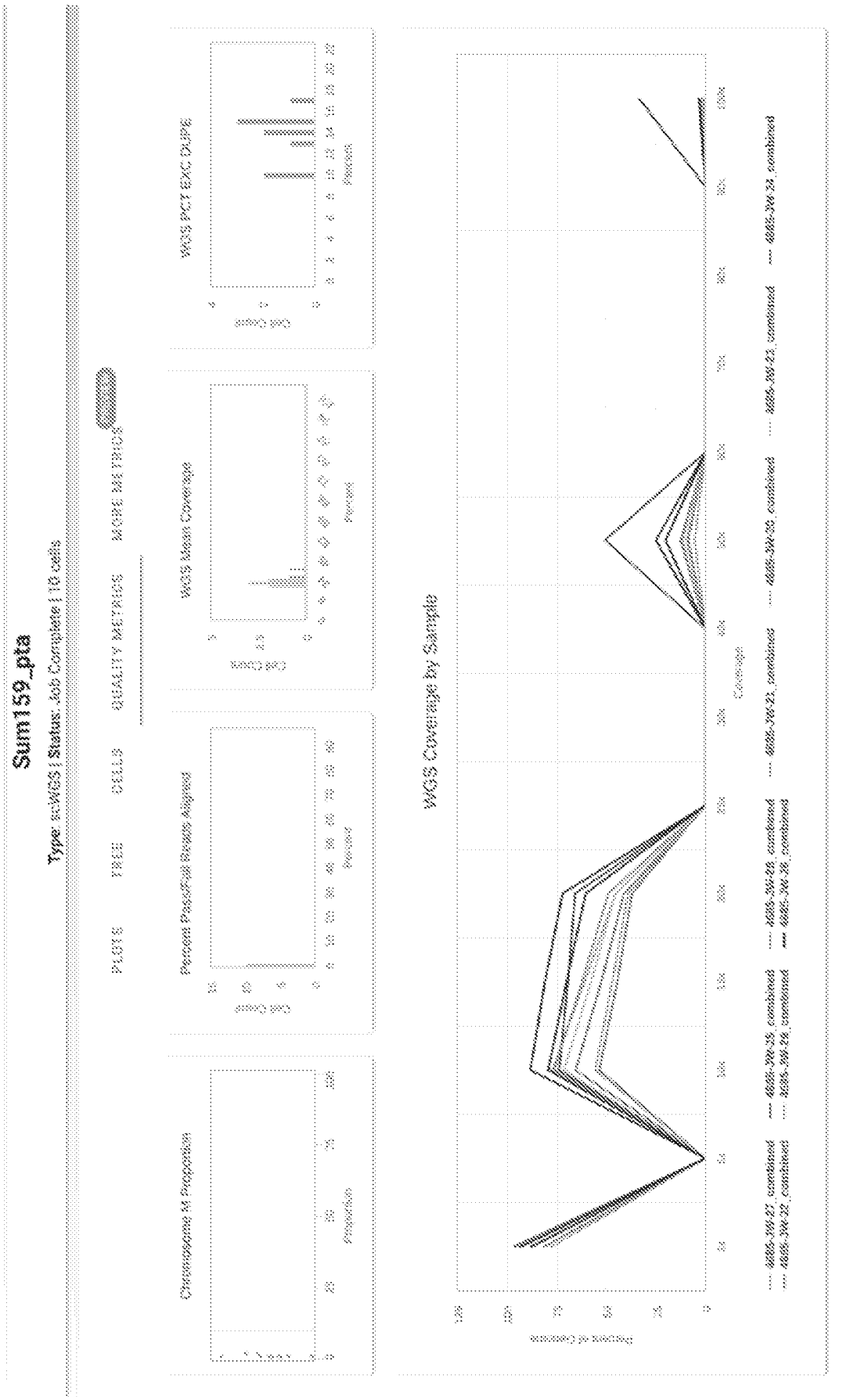

Sum159_pta
Type: scWGS | Status: Job Complete | 10 cells
PLOTS
FREE
CELLS
QUALITY METRICS
MORE METRICS
Chromosome M Proportion
Percent Properly Reads Aligned
WGS Mean Coverage
WGS PCT EXC DUPE
WGS Coverage by Sample
Coverage
Percent of Genome
4685-3W-27_combined
4685-3W-32_combined
4685-3W-35_combined
4685-3W-26_combined
4685-3W-28_combined
4685-3W-21_combined
4685-3W-33_combined
4685-3W-24_combined

FIG. 4B

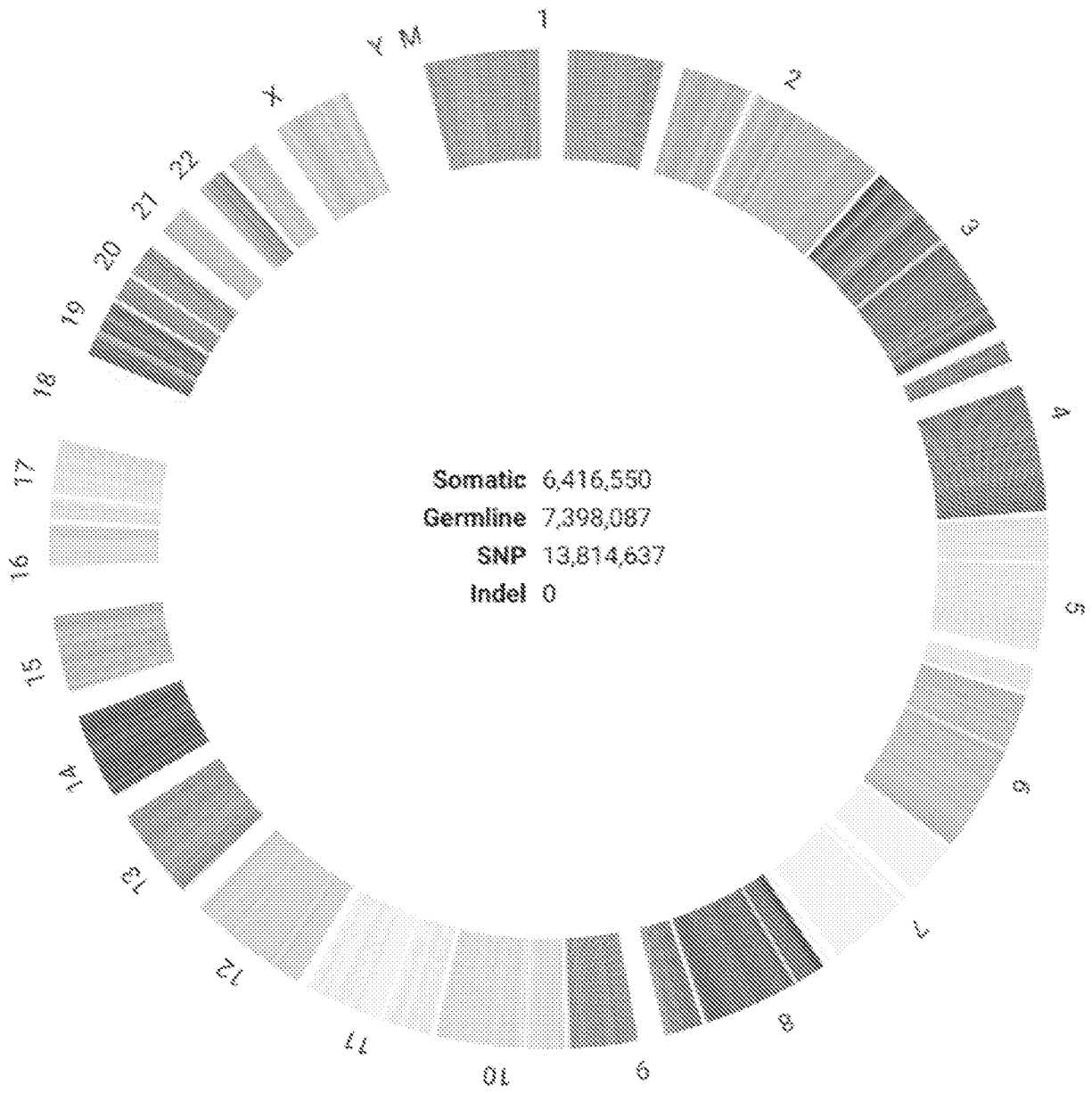

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
M
Somatic 6,416,550
Germline 7,398,087
SNP 13,814,637
Indel 0

FIG. 5A

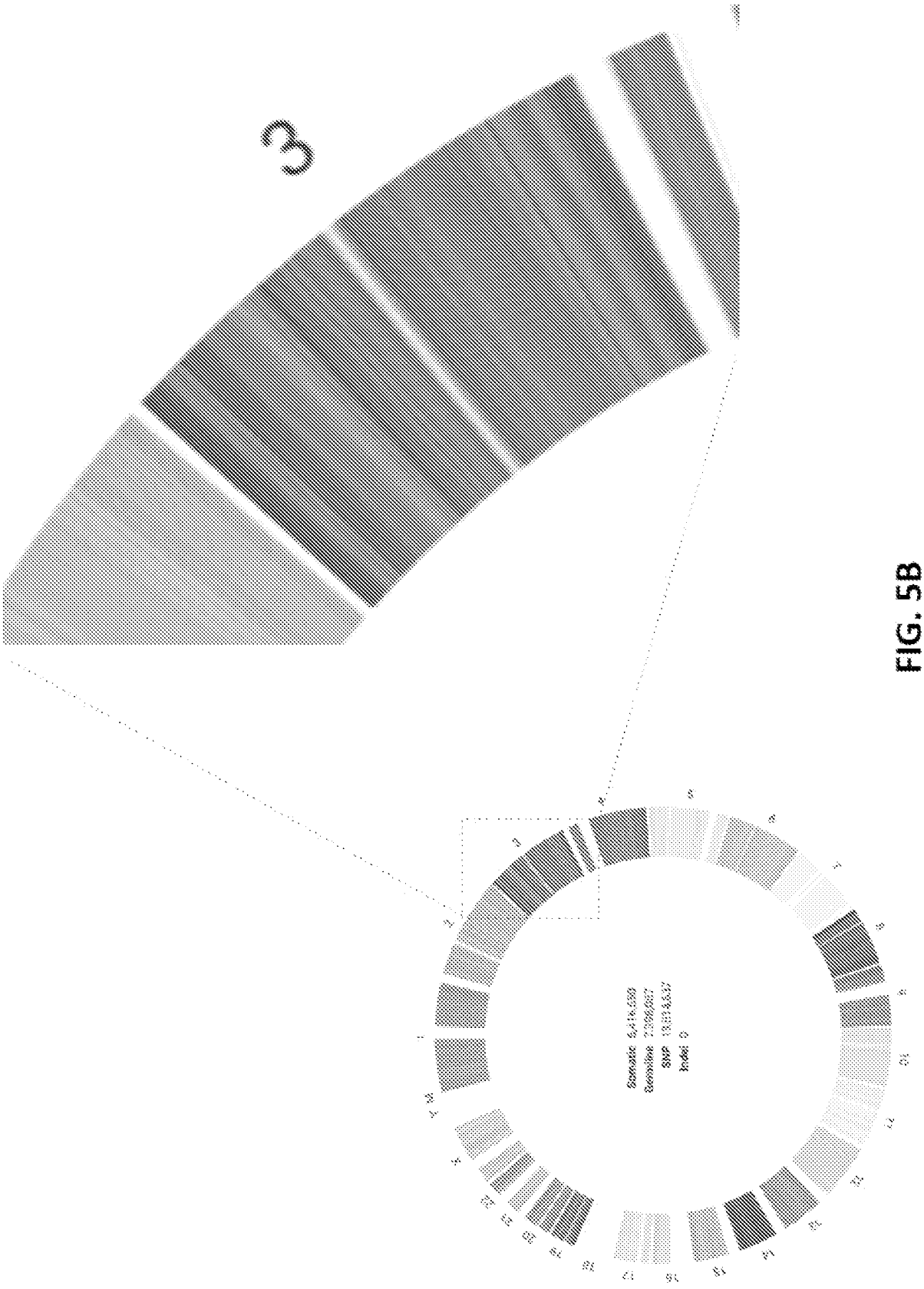


FIG. 5B

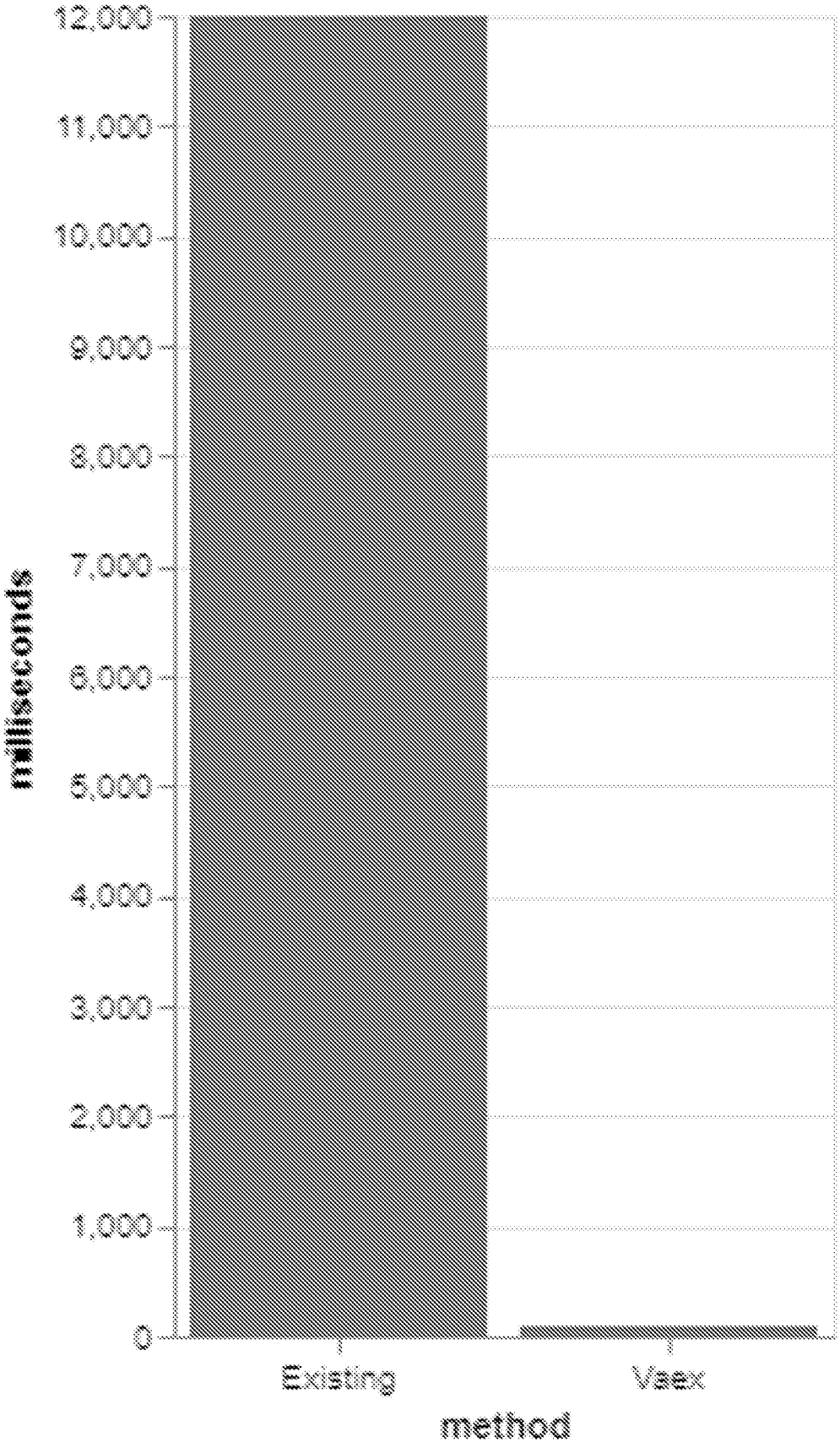

12,000
11,000
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
milliseconds
Existing
Vaex
method

FIG. 6

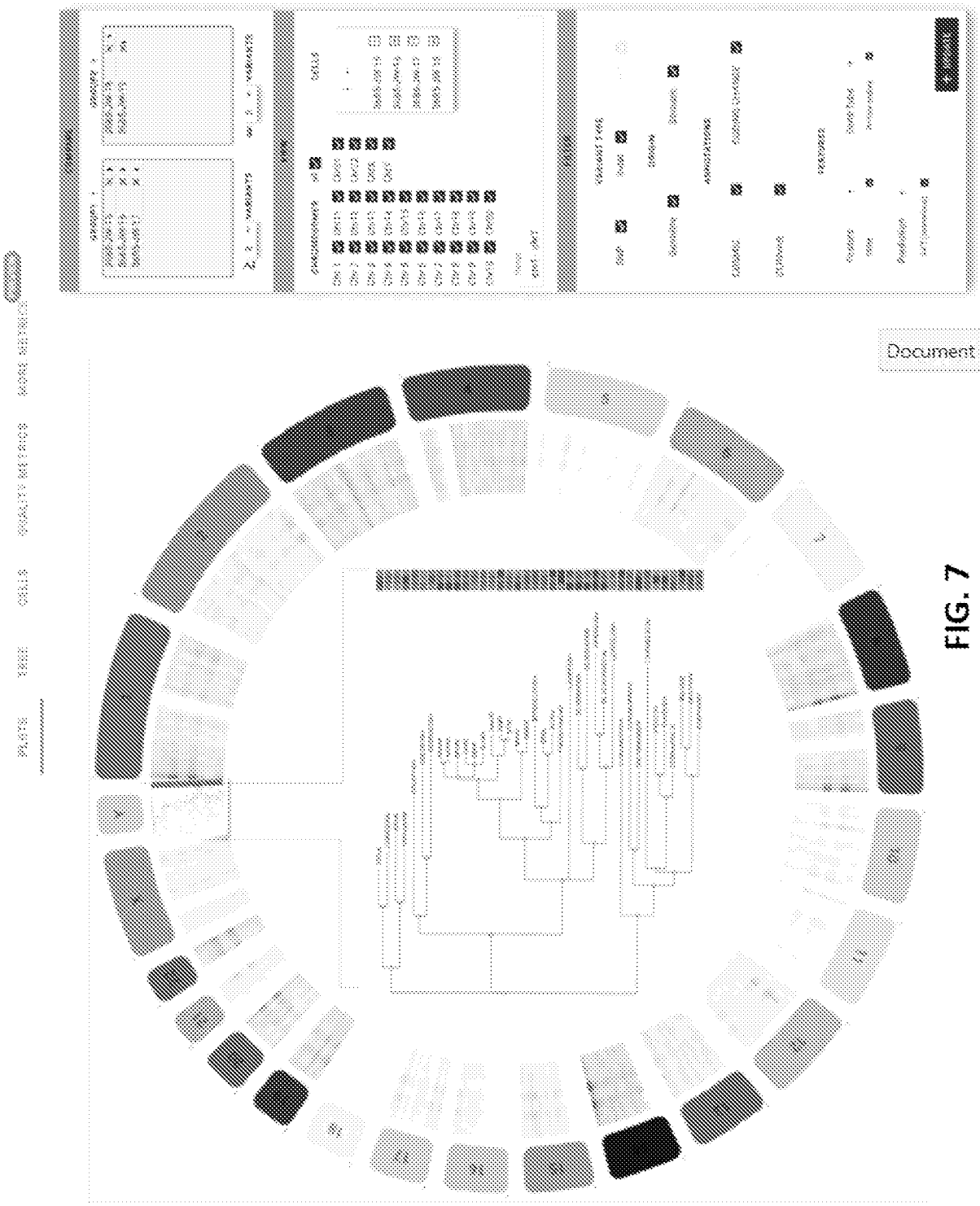


FIG. 7

Document was

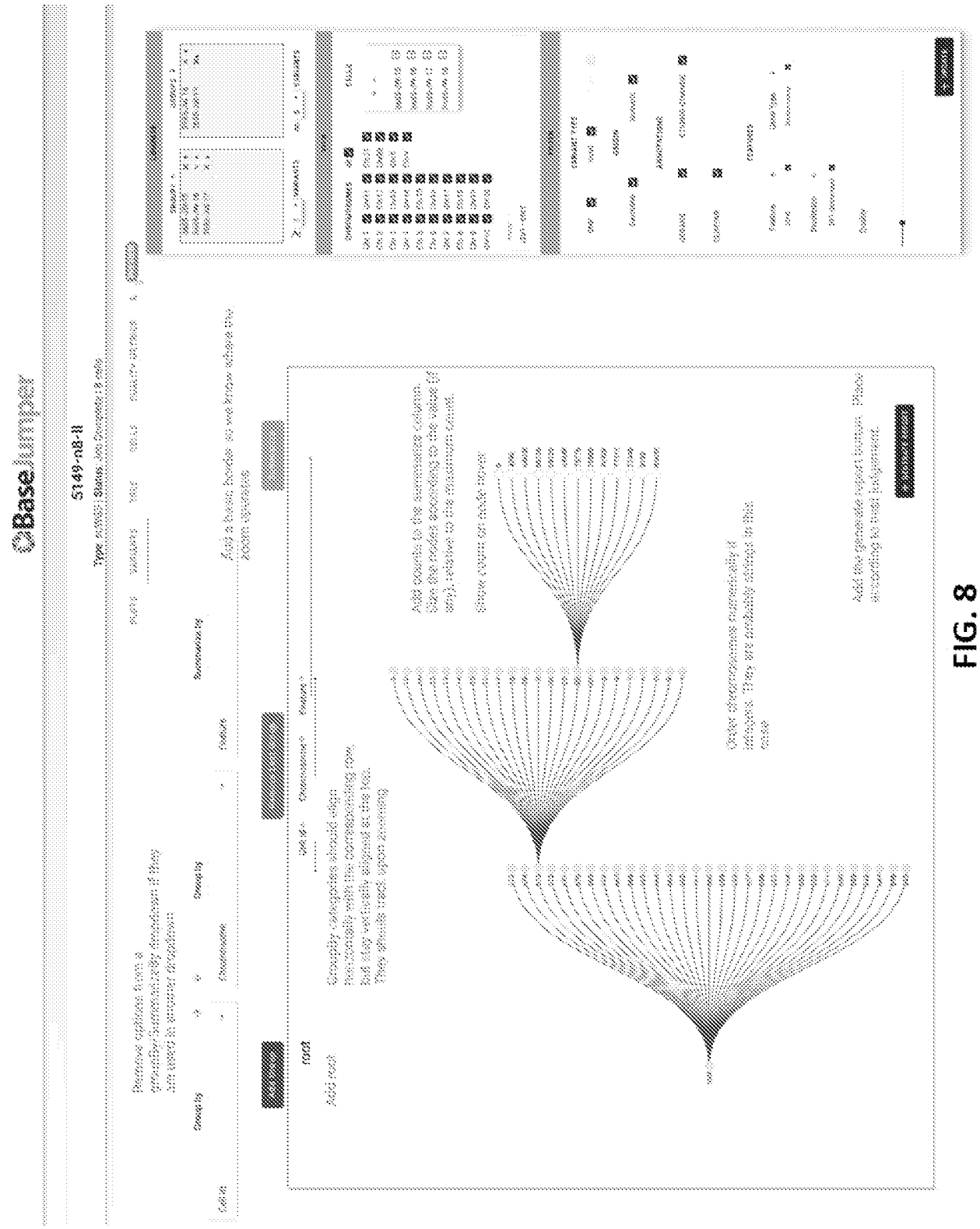


FIG. 8

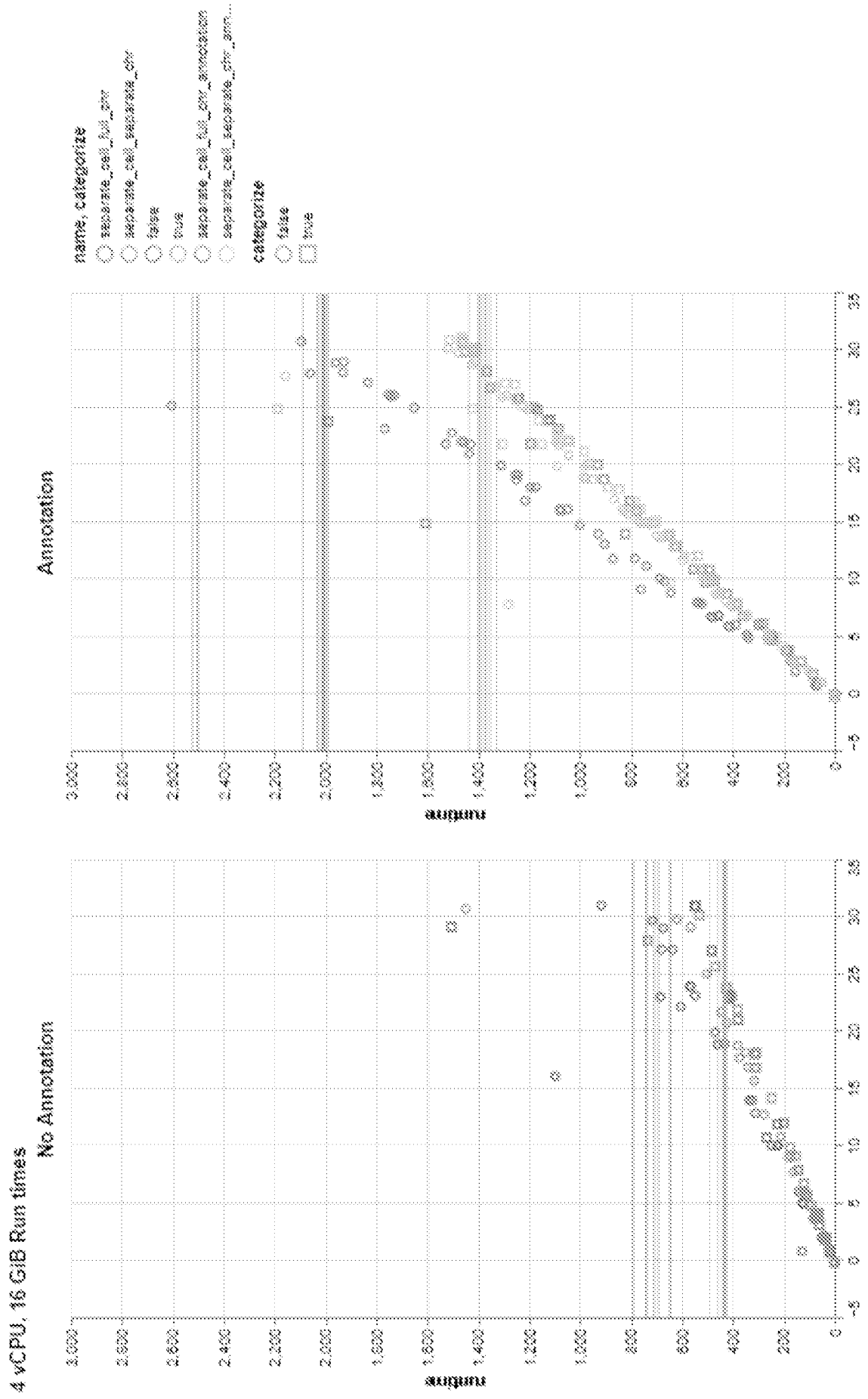

4 vCPU, 16 GiB Run times
No Annotation
Annotation
runtime
ncells_jitter
name, categorize
separate_cell_full_drv
separate_cell_separate_drv
false
true
separate_cell_full_drv_annotation
separate_cell_separate_drv_ann...
categorize
false
true

FIG. 9

FIG. 10

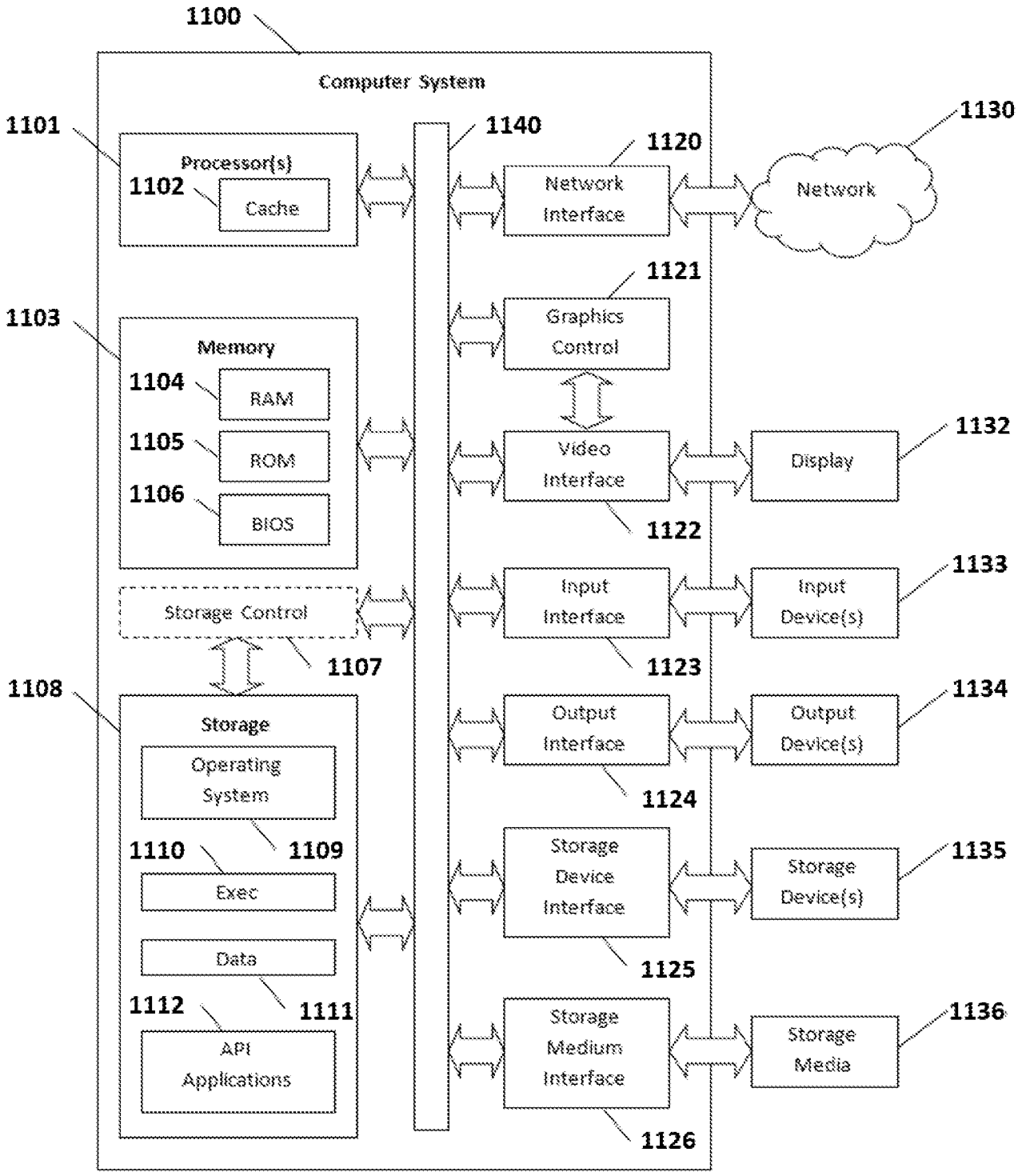

1100
Computer System
1101
Processor(s)
1102
Cache
1103
Memory
1104
RAM
1105
ROM
1106
BIOS
Storage Control
1107
1108
Storage
Operating System
1109
1110
Exec
Data
1111
1112
API Applications
1140
1120
Network Interface
1130
Network
1121
Graphics Control
Video Interface
1122
Display
1132
Input Interface
1123
Input Device(s)
1133
Output Interface
1124
Output Device(s)
1134
Storage Device Interface
1125
Storage Device(s)
1135
Storage Medium Interface
1126
Storage Media
1136

FIG. 11

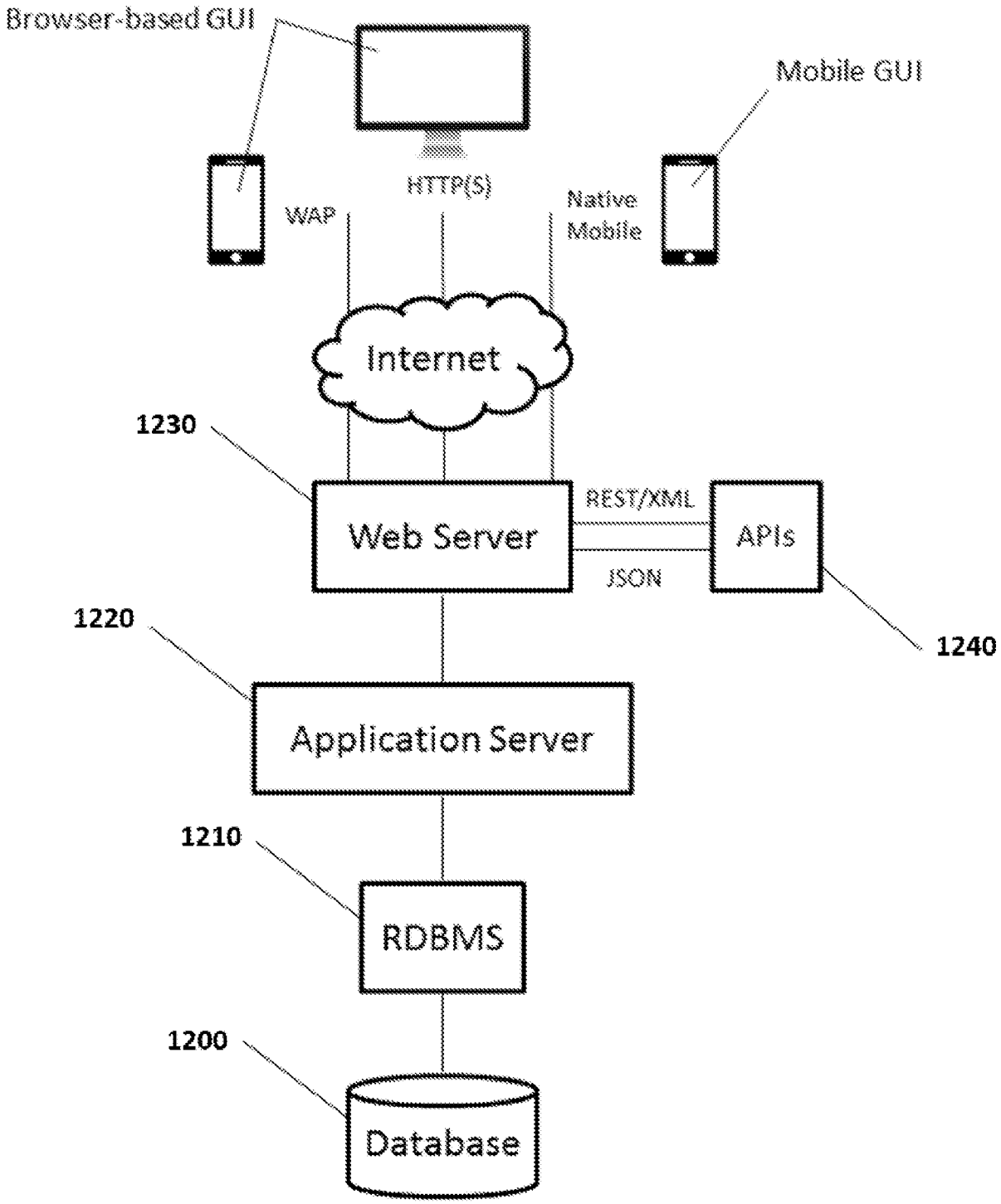

Browser-based GUI
Mobile GUI
HTTP(S)
WAP
Native Mobile
Internet
1230
Web Server
REST/XML
JSON
APIs
1240
1220
Application Server
1210
RDBMS
1200
Database

FIG. 12

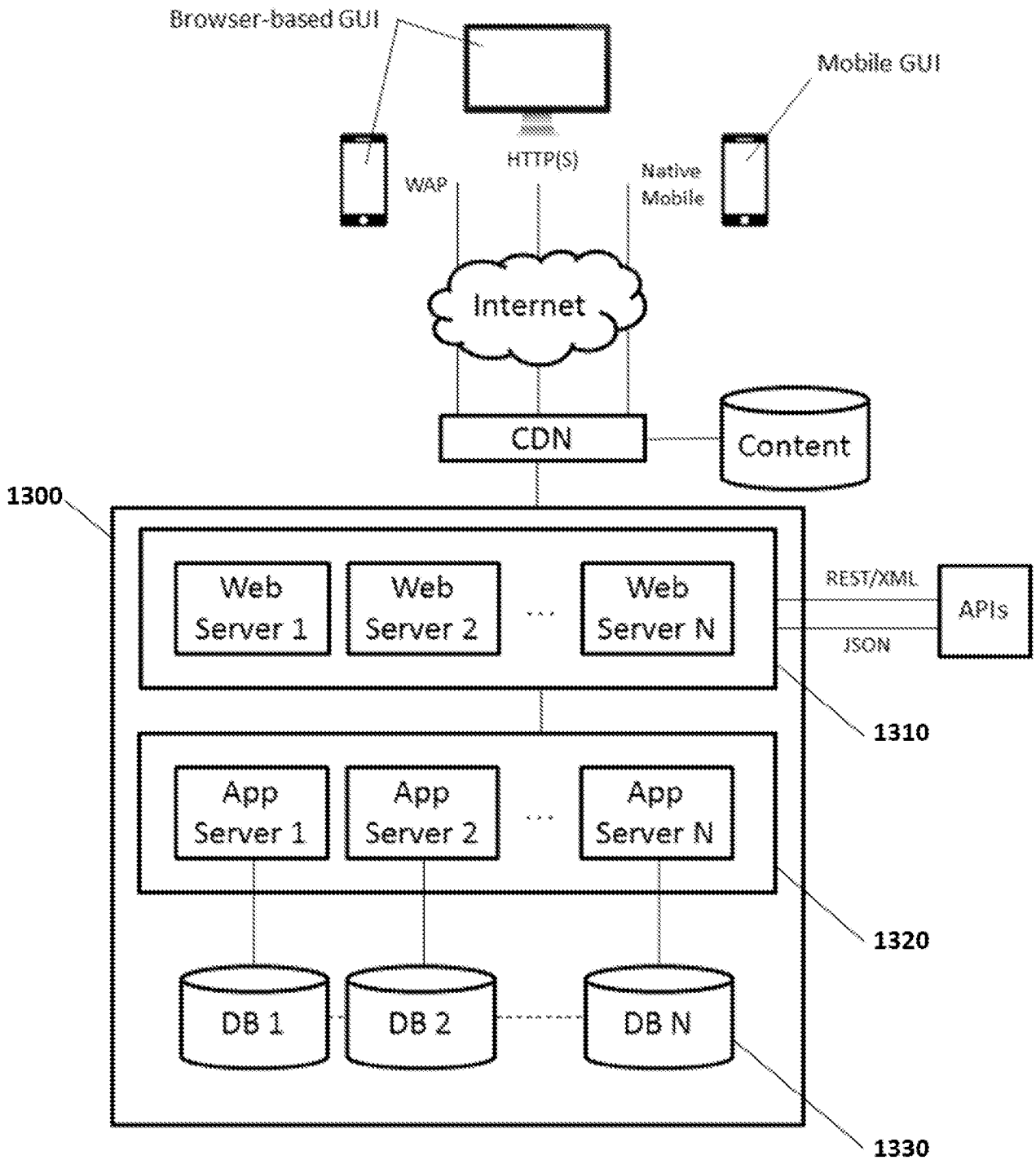

Browser-based GUI
Mobile GUI
HTTP(S)
WAP
Native Mobile
Internet
CDN
Content
1300
Web Server 1
Web Server 2
Web Server N
REST/XML
JSON
APIs
1310
App Server 1
App Server 2
App Server N
1320
DB 1
DB 2
DB N
1330

FIG. 13

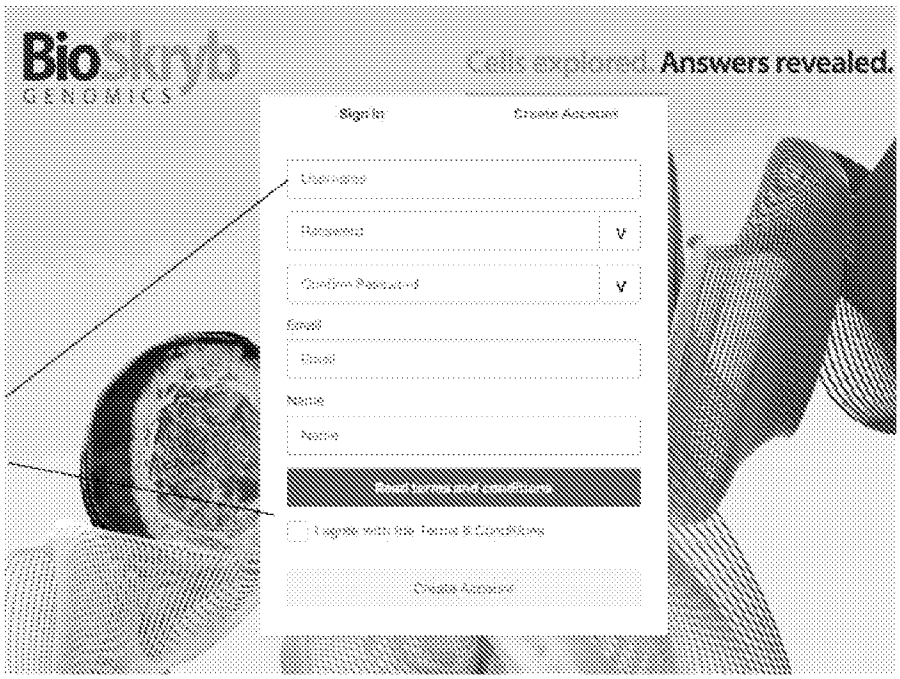


FIG. 14A

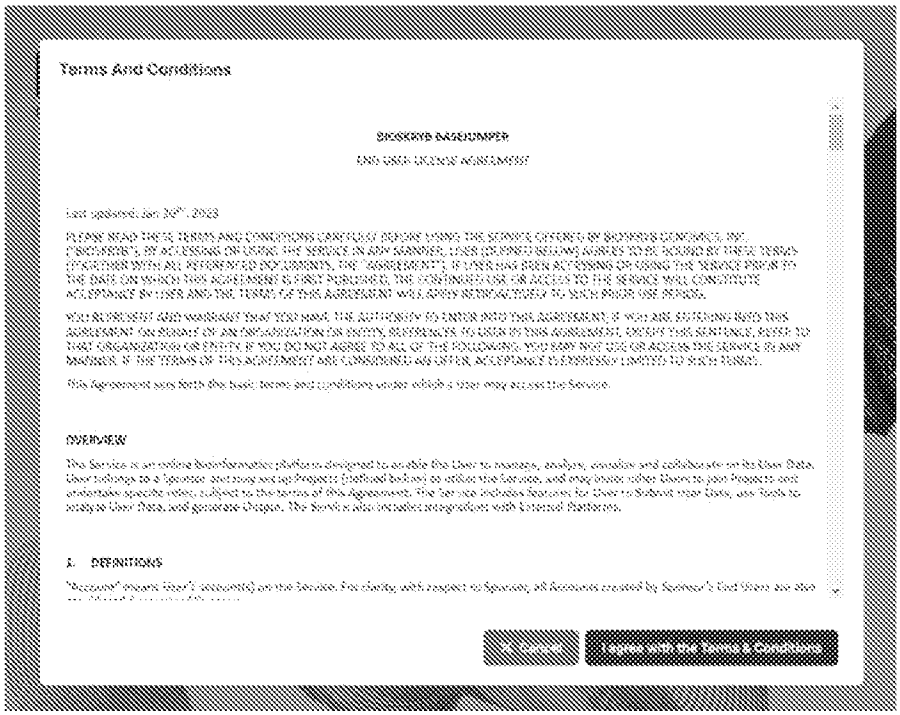


FIG. 14B

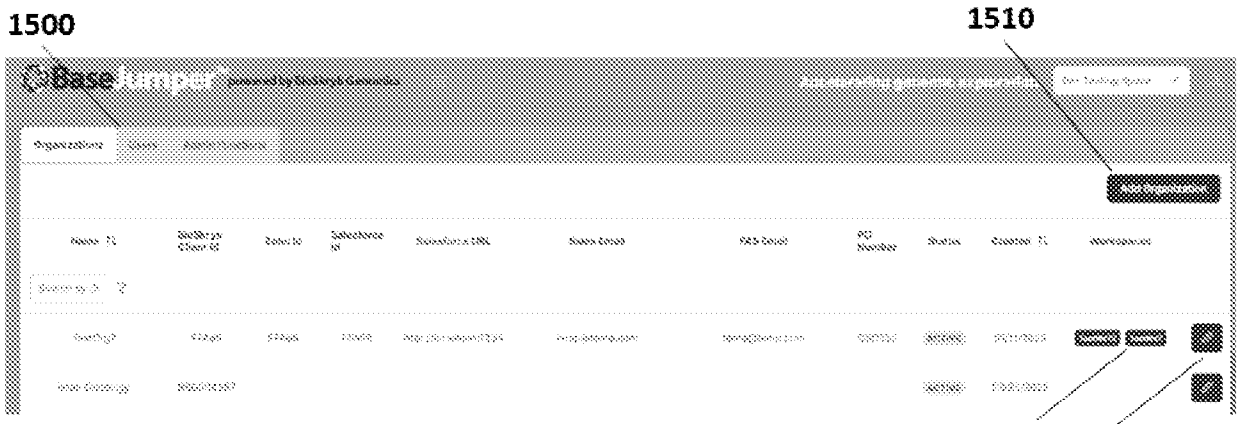


FIG. 15A

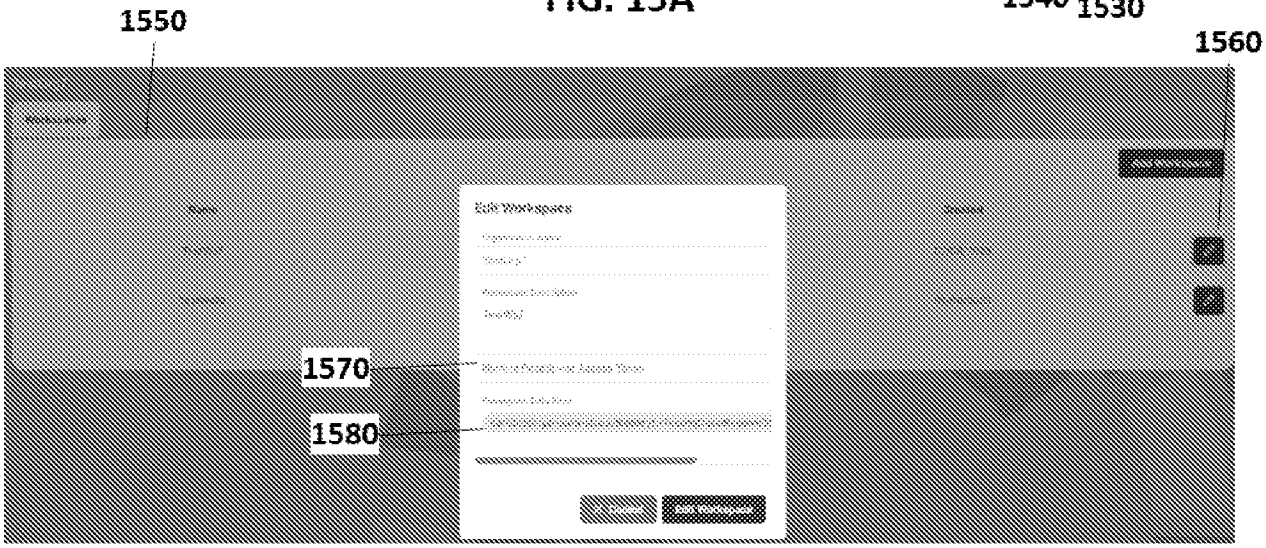


FIG. 15B

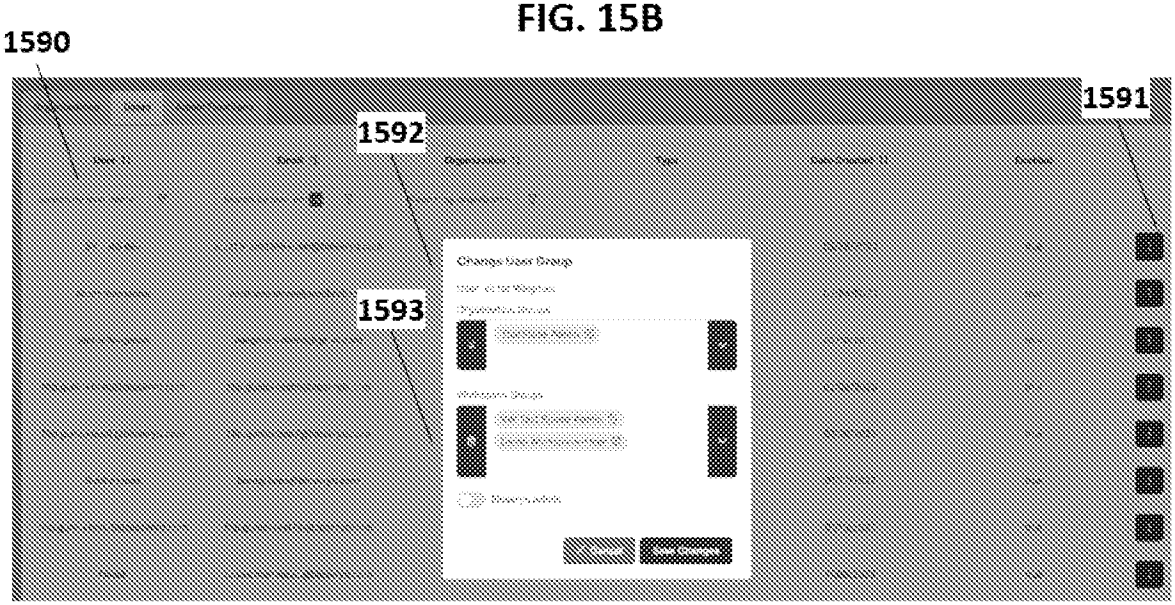


FIG. 15C

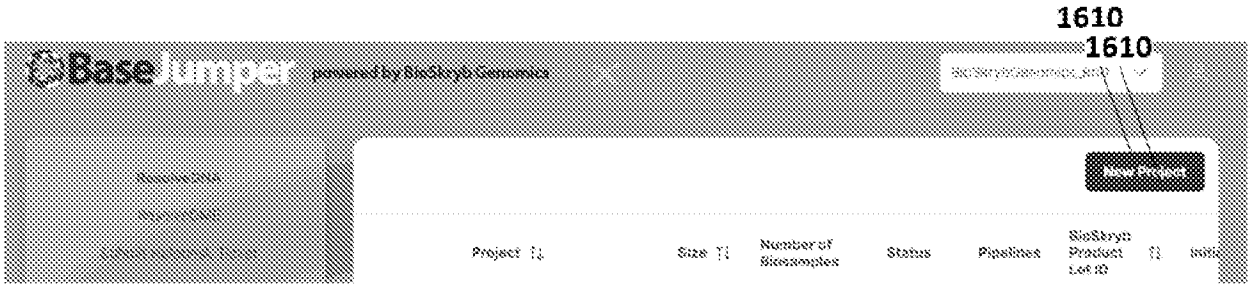


FIG. 16A

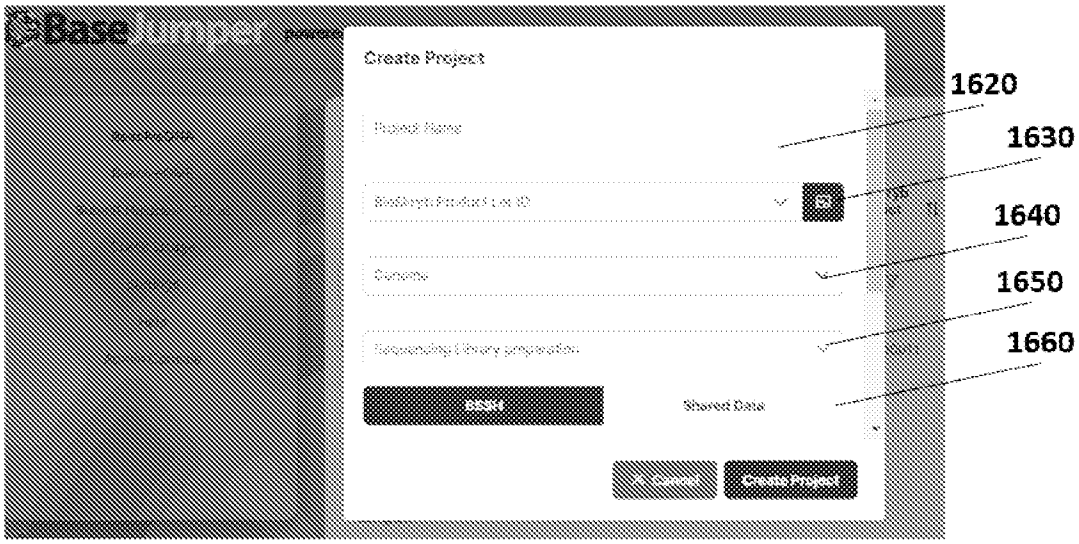


FIG. 16B

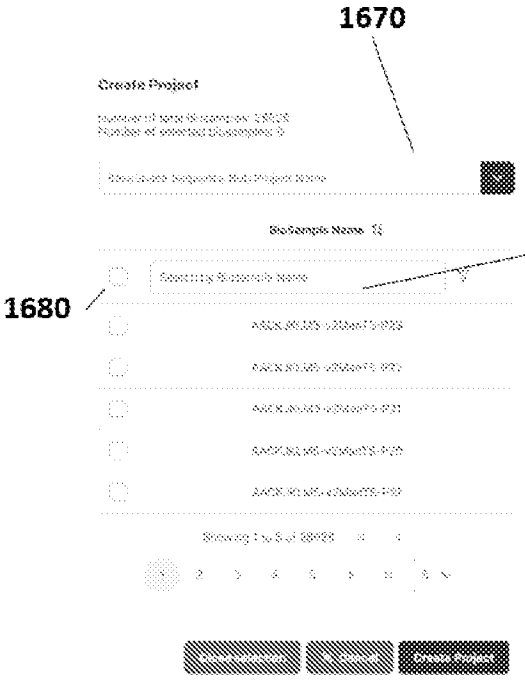


FIG. 16C

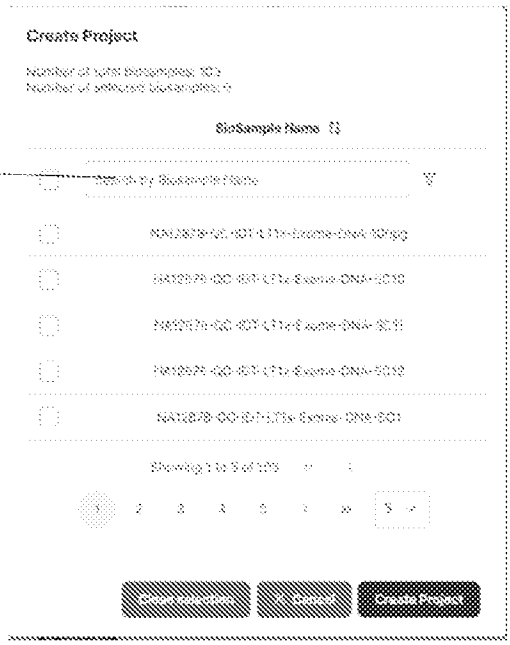


FIG. 16D

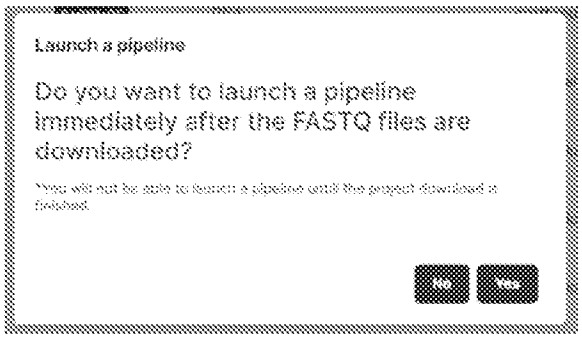


FIG. 17A

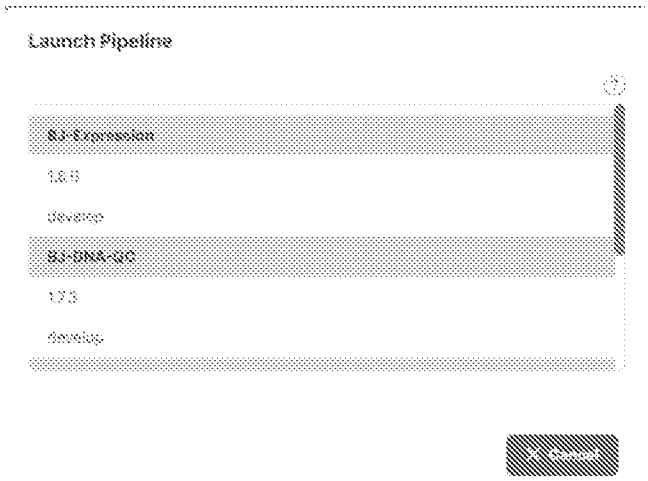


FIG. 17B

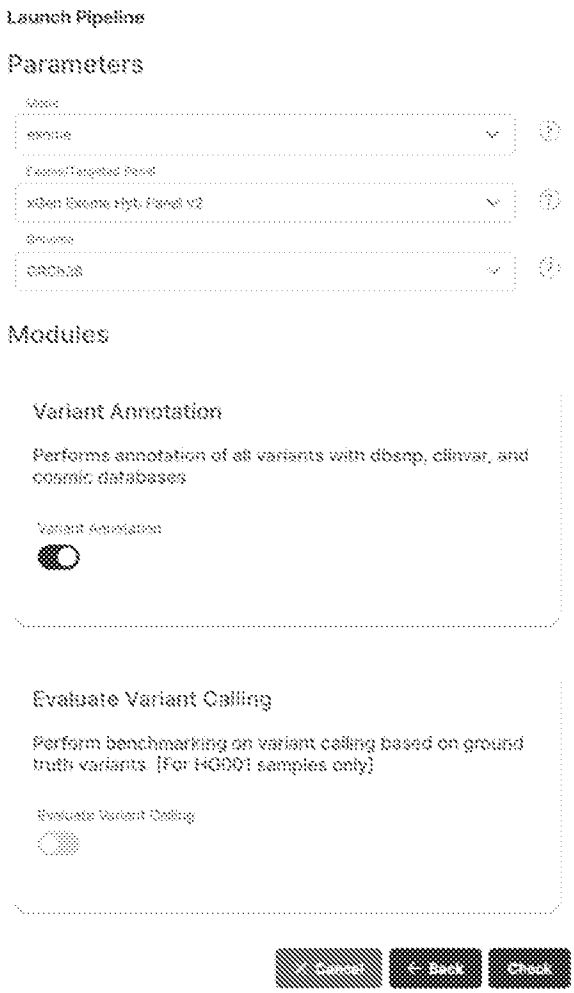


FIG. 17C

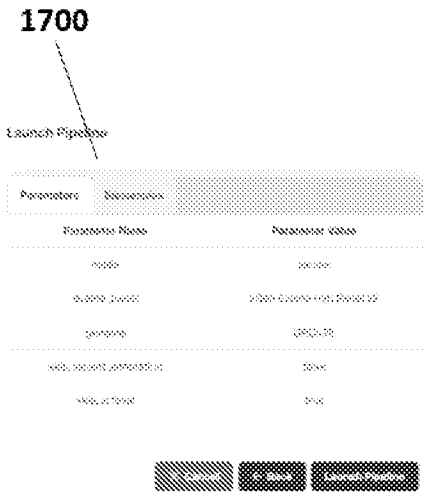


FIG. 17D

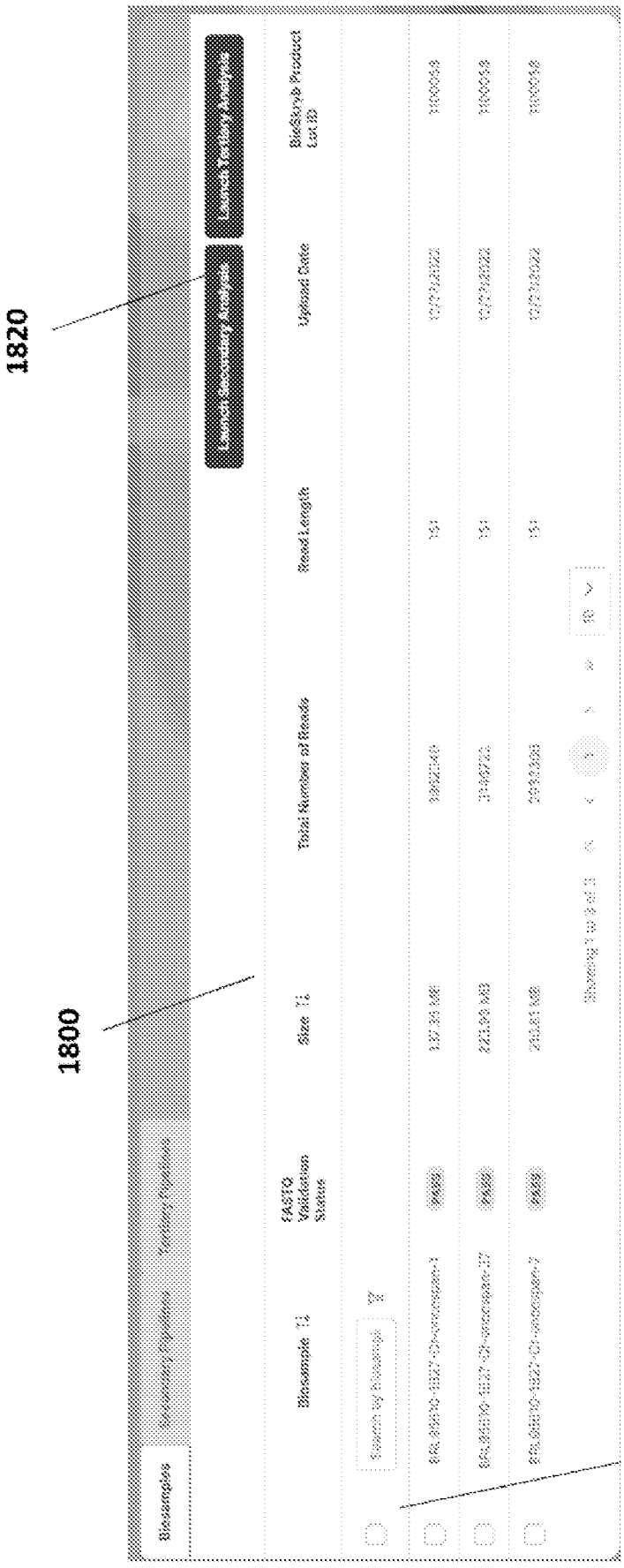

1820
1800
1810
Biosamples
Sequencing Runs
Variant Pipelines
Biosample
FASTQ Validation Status
Size
Total Number of Reads
Read Length
Upload Date
BioStorys Product Lot ID

FIG. 18A

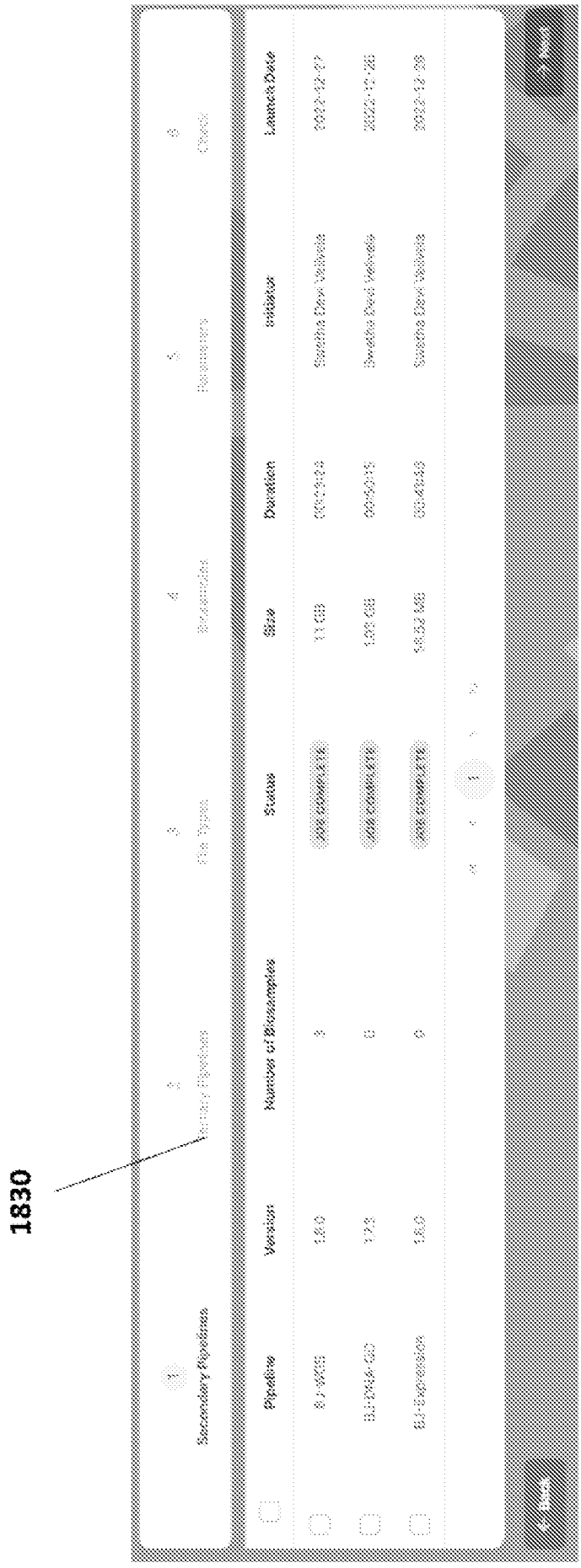

1830
Secondary Pipelines
Primary Pipelines
Pipeline
Version
Number of Biosamples
Status
Size
Duration
Initiator
Launch Date

FIG. 18B

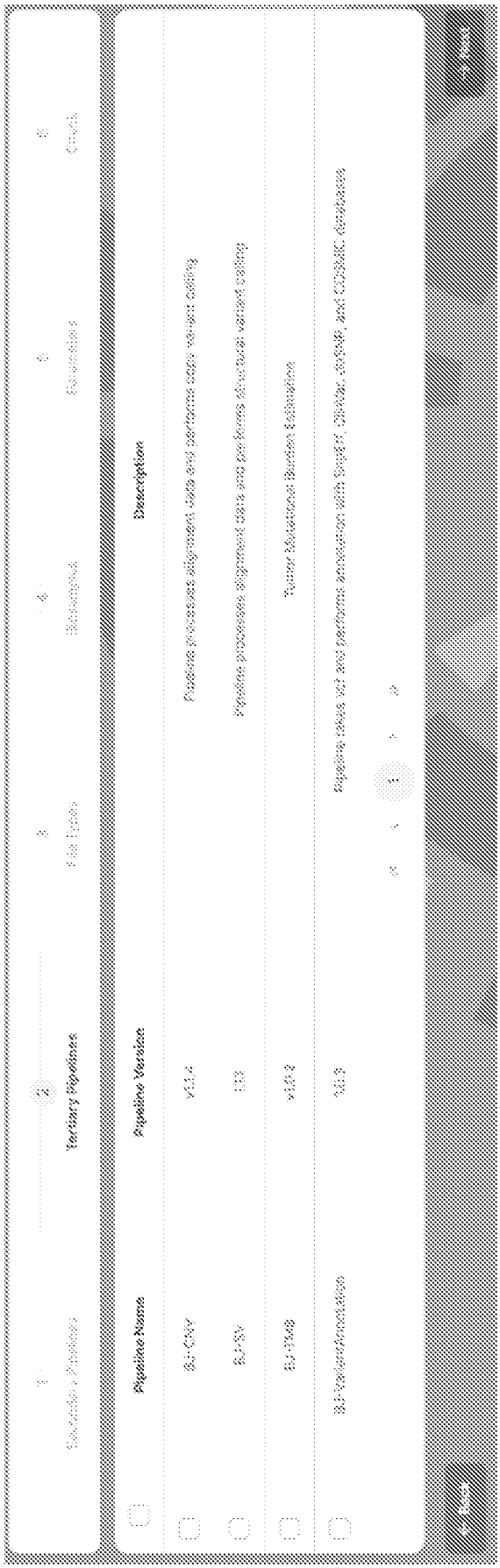

Tertiary Pipelines
Pipeline Name
Pipeline Version
Description
Pipeline processes alignment data and performs copy number calling
Pipeline processes alignment data and performs structural variant calling
Tumor Mutational Burden Estimation
Pipeline takes vcf and performs annotation with SnpEff, ClinVar, dbNSFP, and COSMIC databases

FIG. 18C

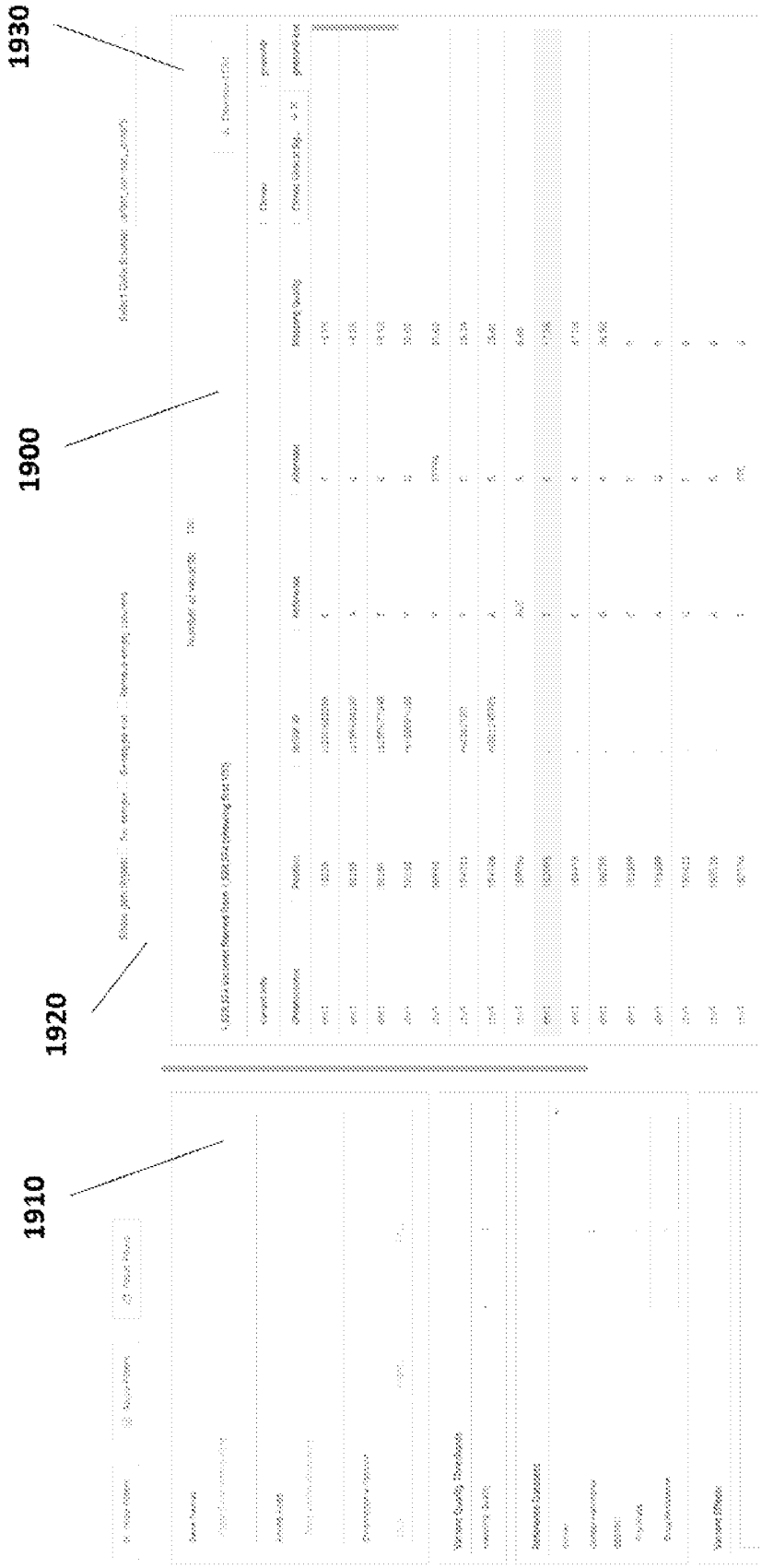


FIG. 19A

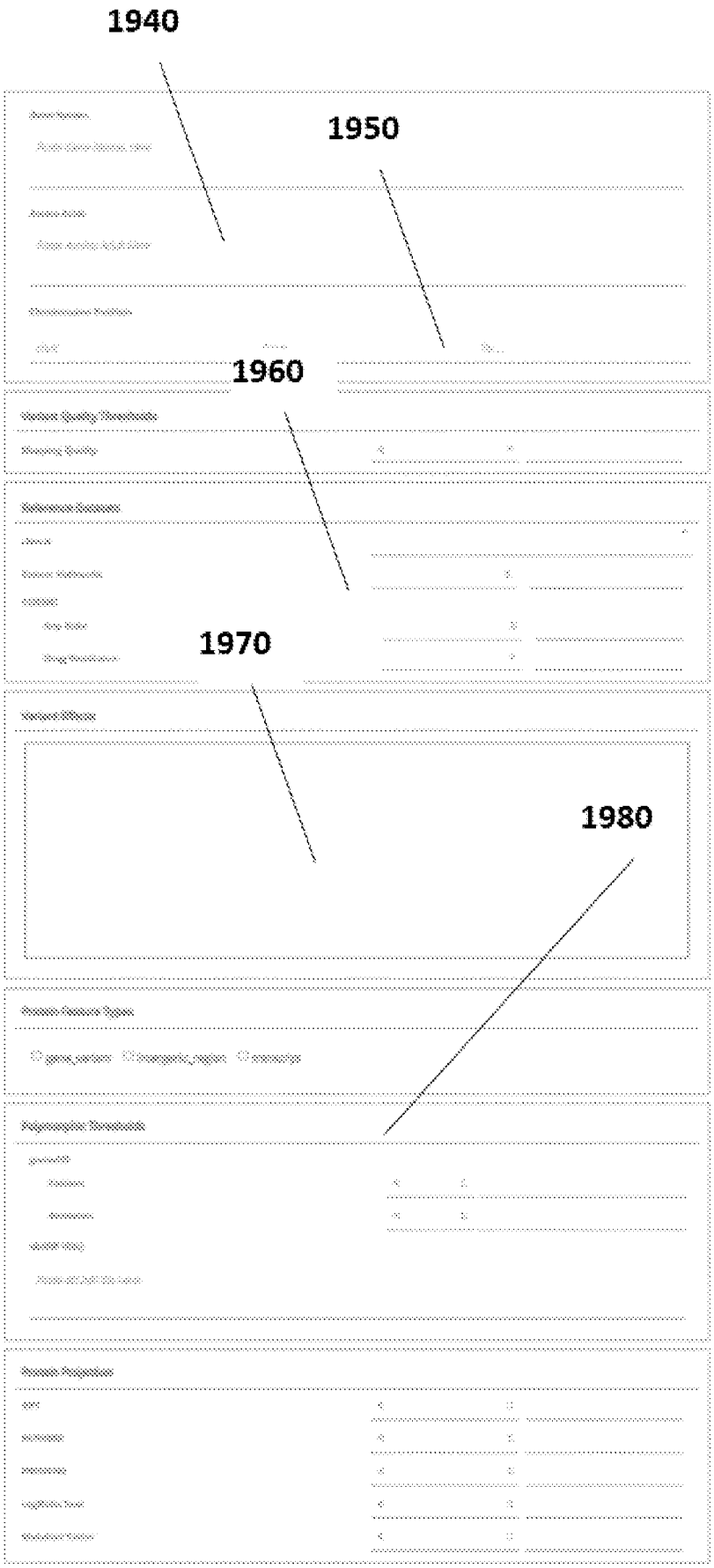


FIG. 19B

FIG. 19C

Current filters:
{'gnomad-genomes': 0.1, 'gnomad-genomes-exc': '', 'clinvar': ['association'], 'feature-type': ['gene_variant']}
Filter History:

FIG. 19D

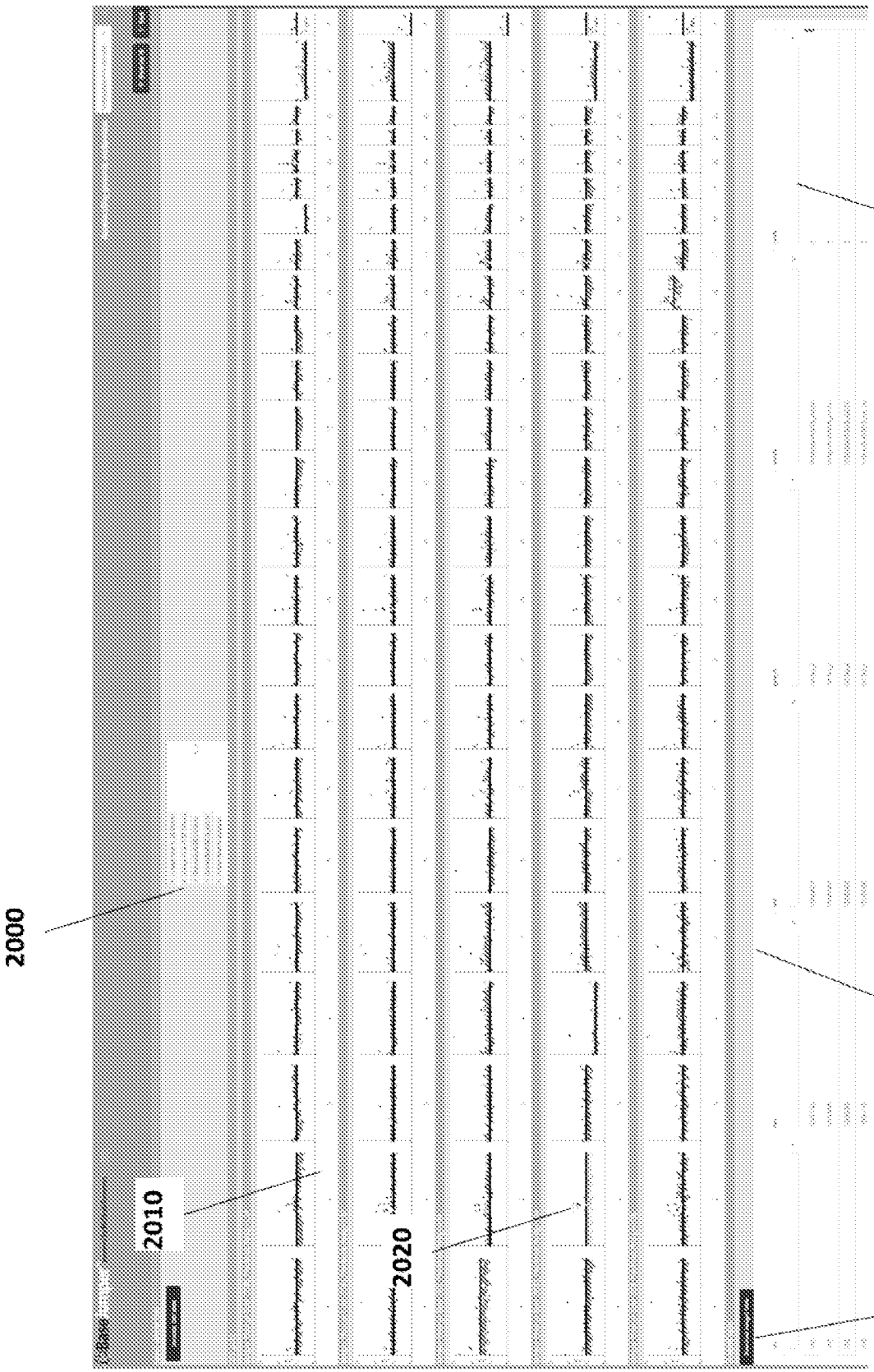


FIG. 20

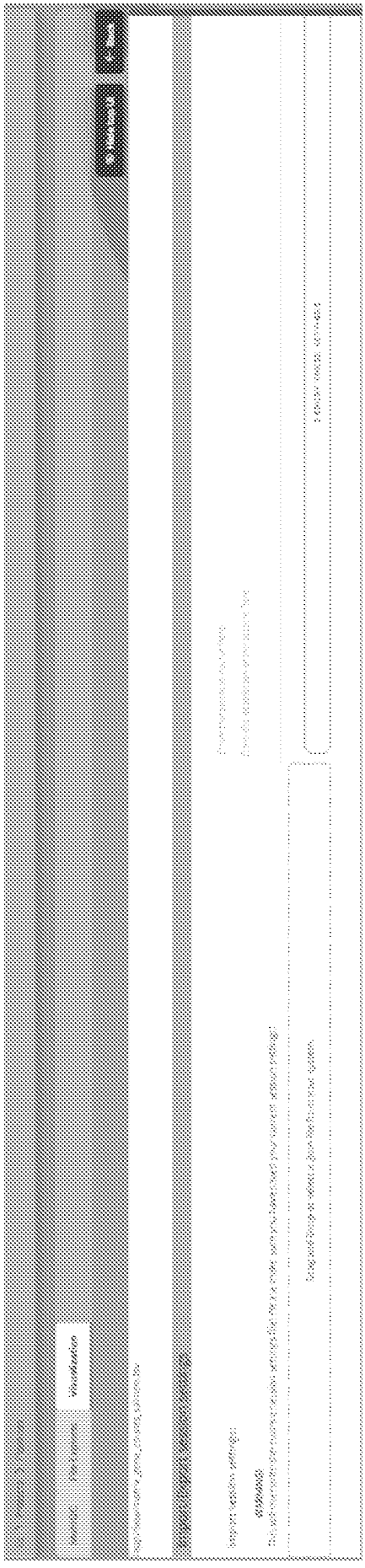


FIG. 21A

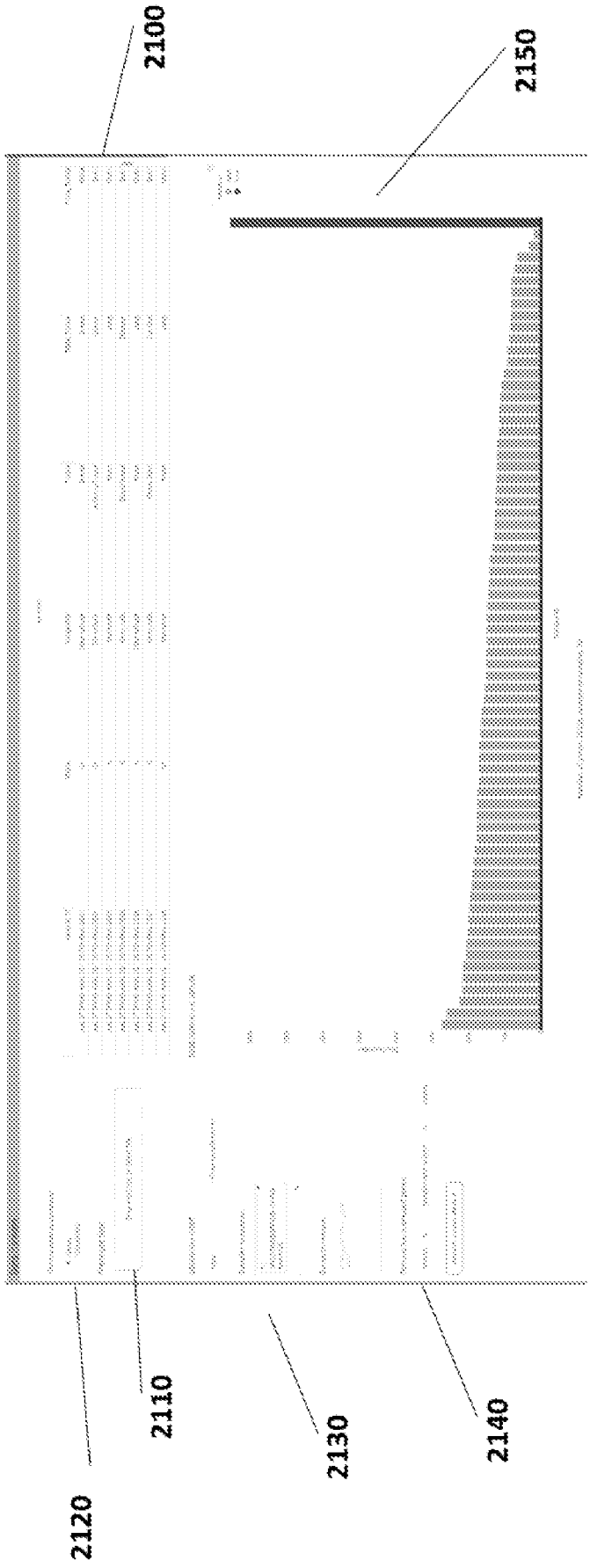


FIG. 21B

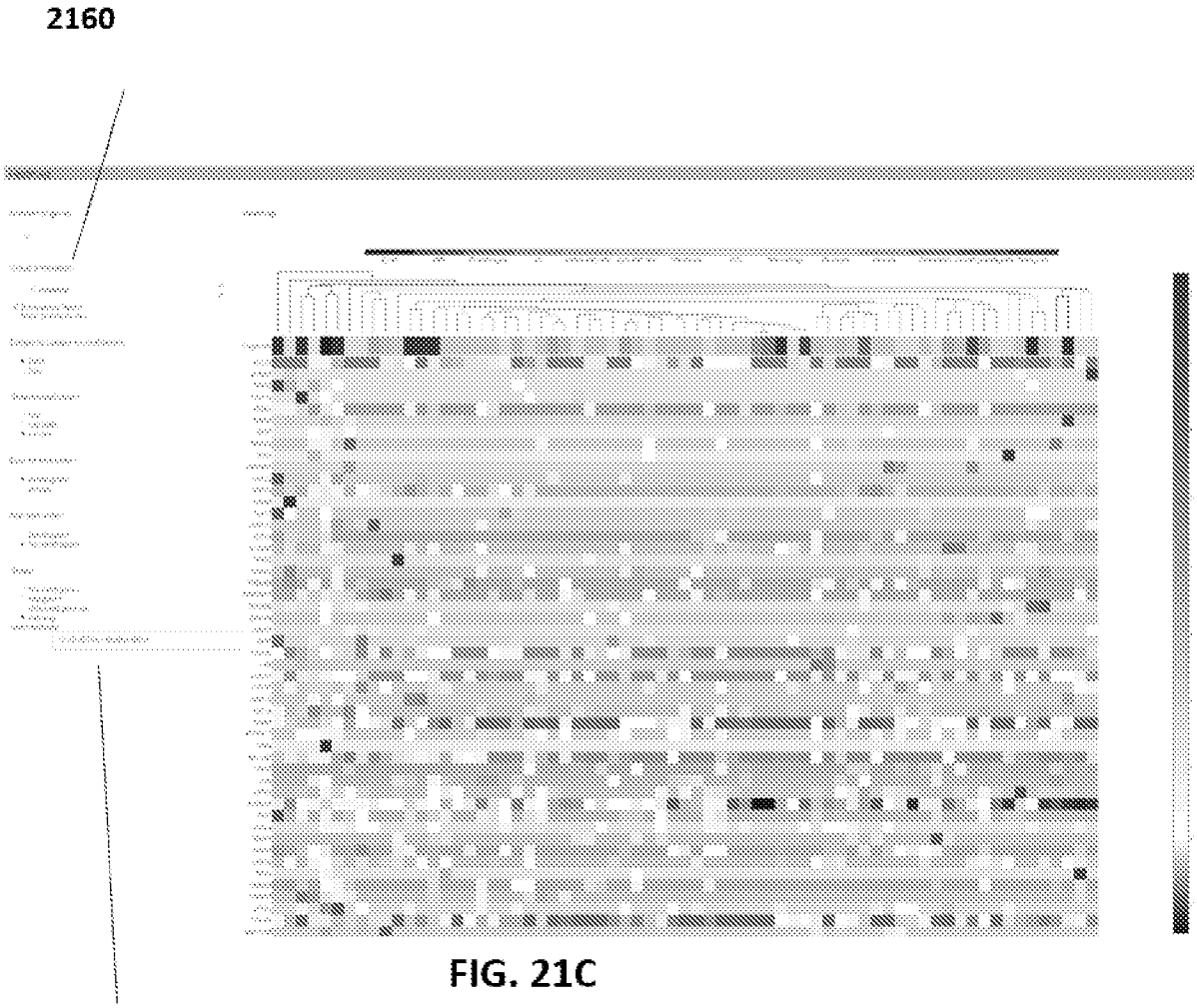


FIG. 21C

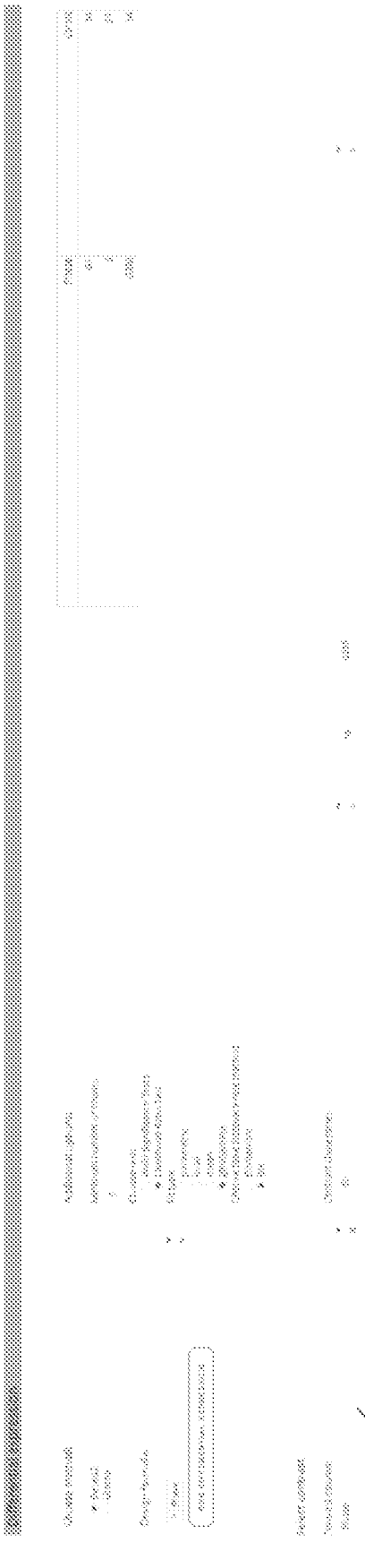


FIG. 21D

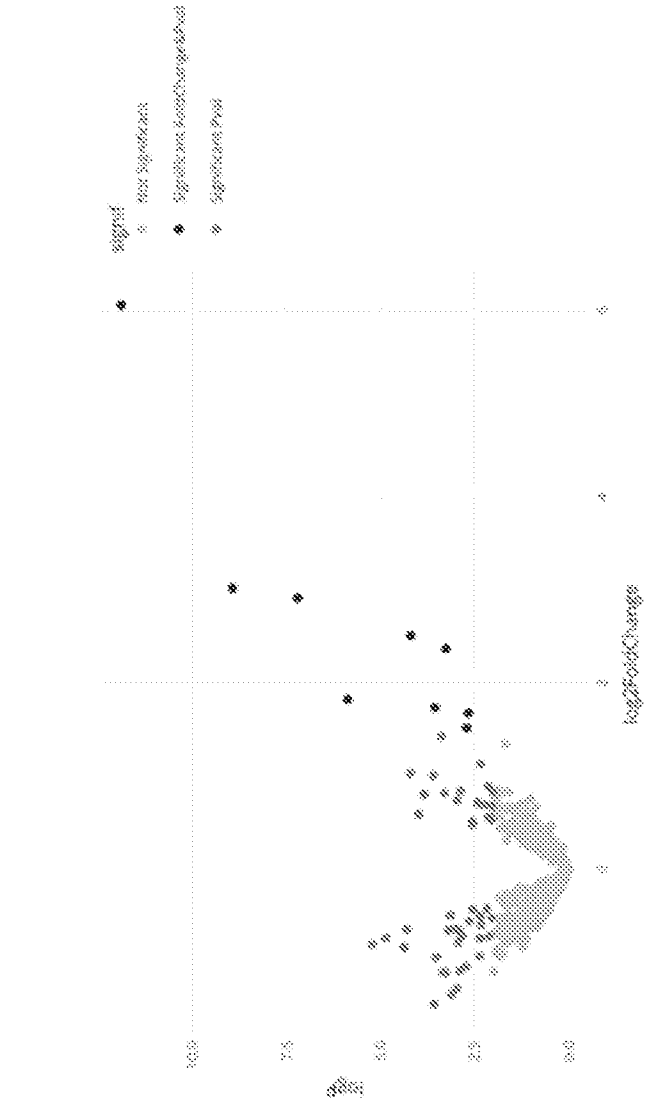

Volcano plot
log2 fc value threshold:
2
log2 fold change:
1.5
gene_name pvalue padj log2fc
MKLN1 1.3e-12 1.4e-8 6.1
MIR1302 1.3e-8 4.6e-7 4
MIR4661 3.3e-8 0.000014 2.5
XIST 0.0000015 0.00028 1.4
ITM... 0.000001 0.0014 -0.81
NAB... 0.000014 0.0026 -0.72
log2FoldChange
Legend
Not Significant
Significant Fold Change
Significant Both

FIG. 21E

2190

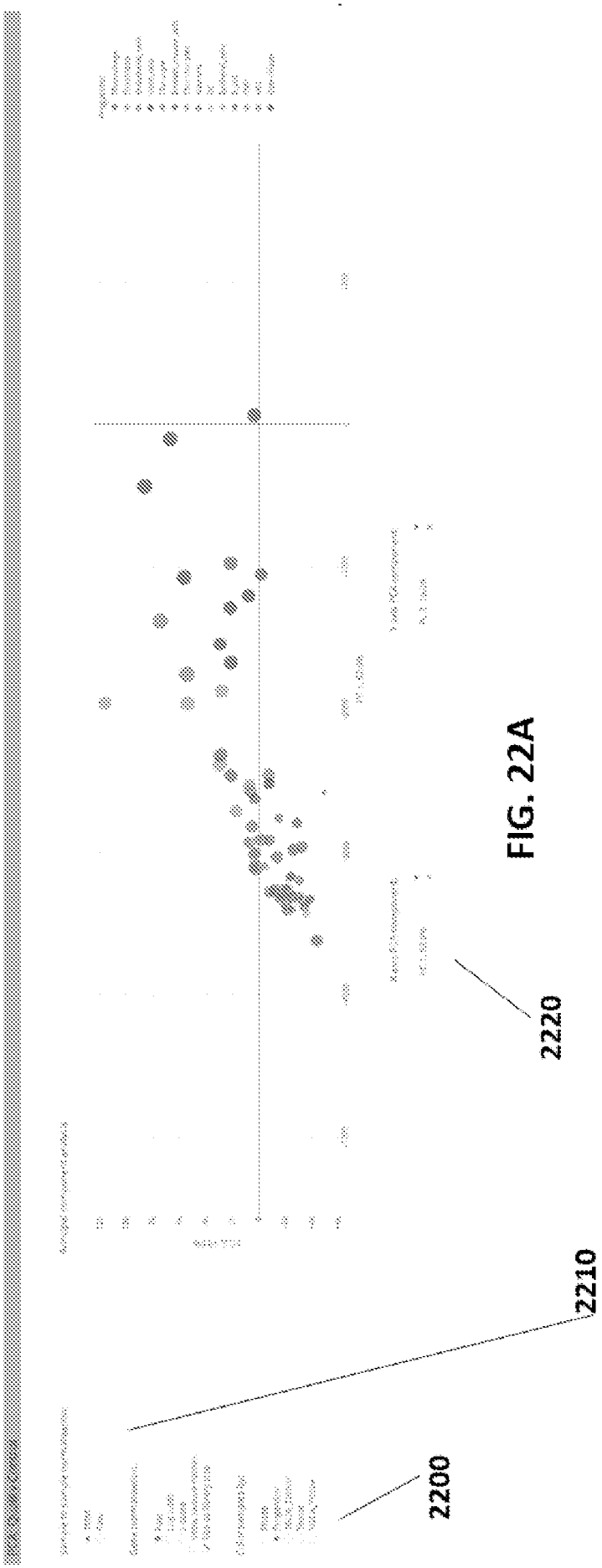


FIG. 22A

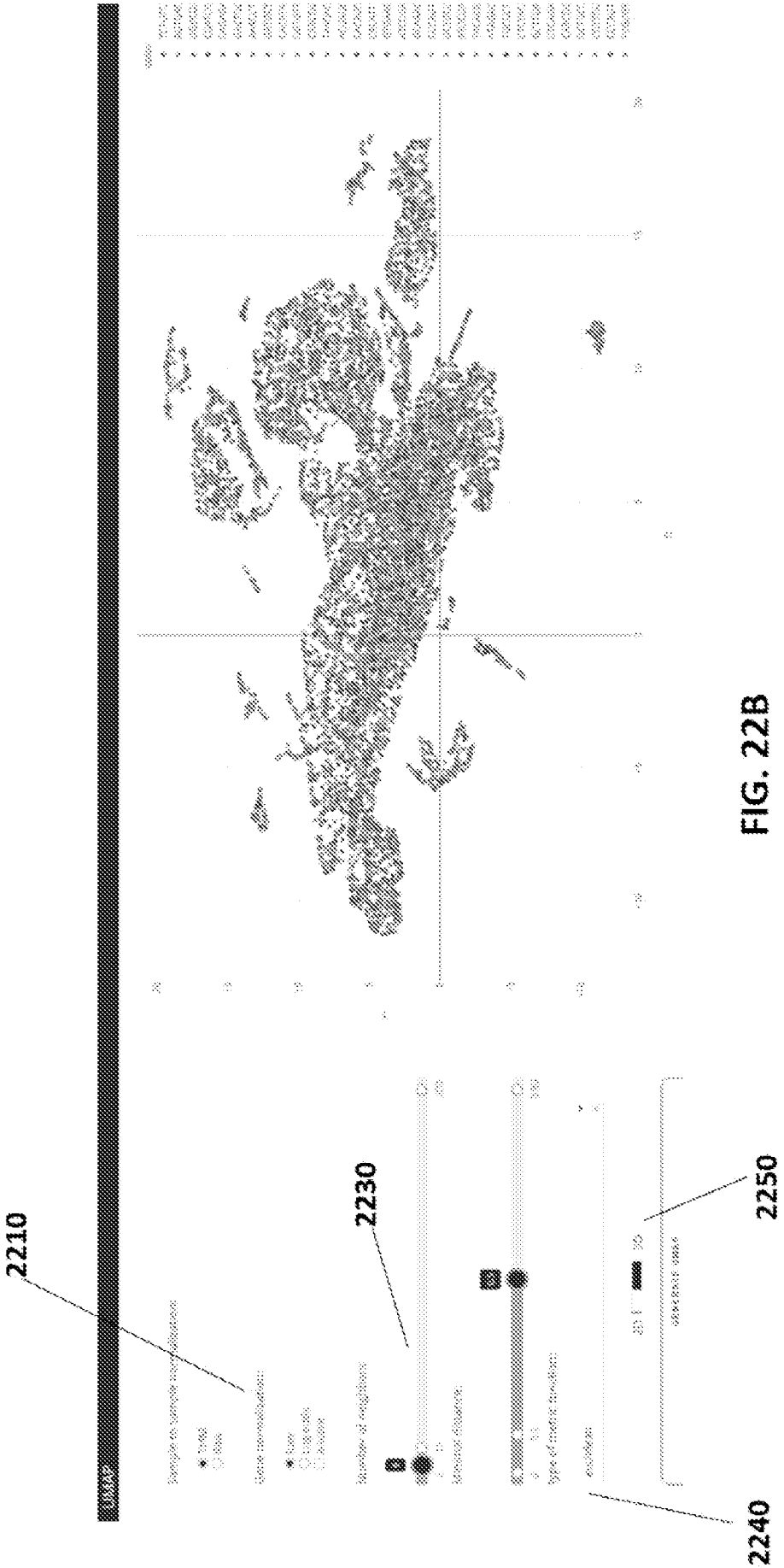


FIG. 22B

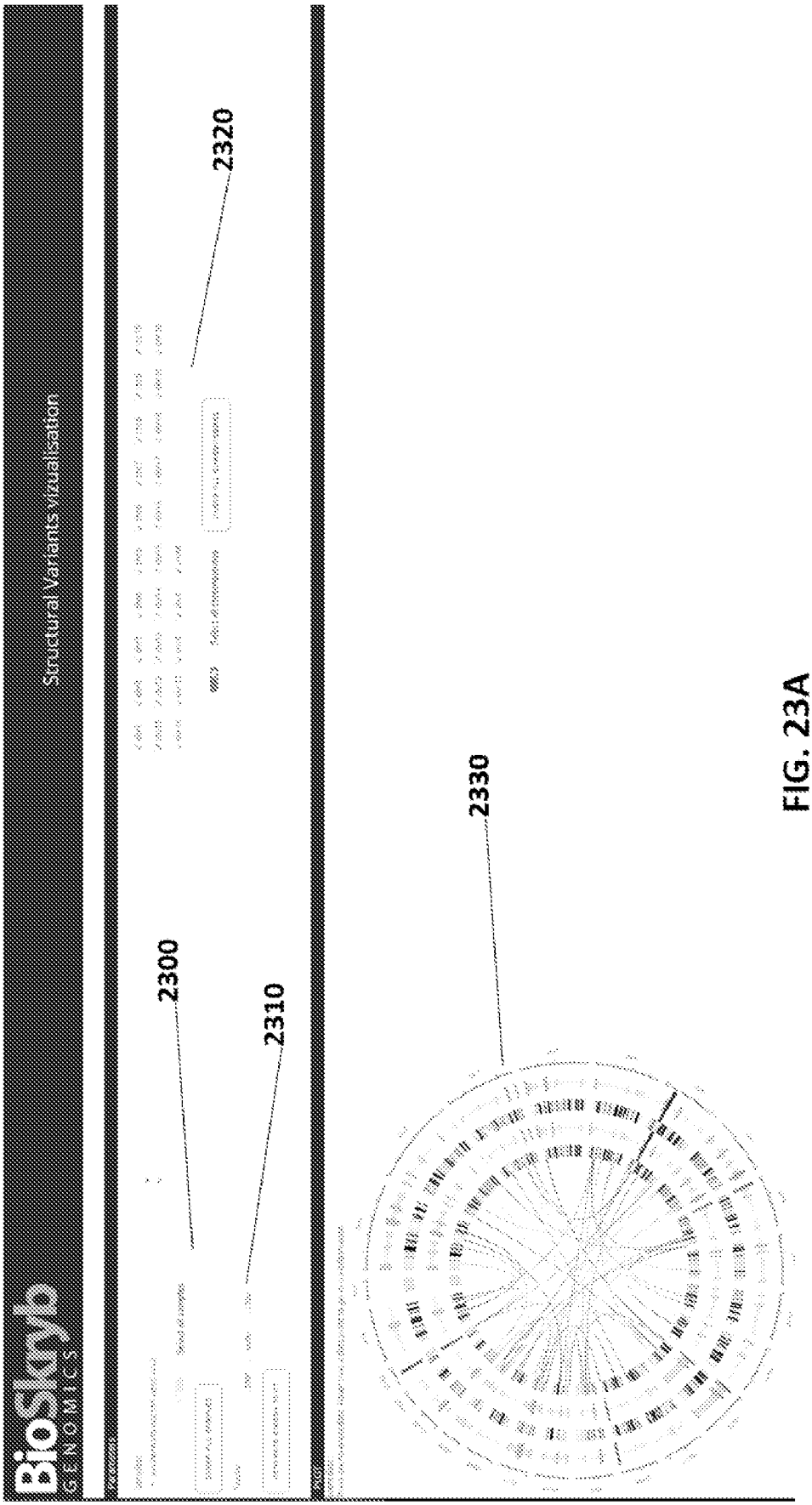


FIG. 23A

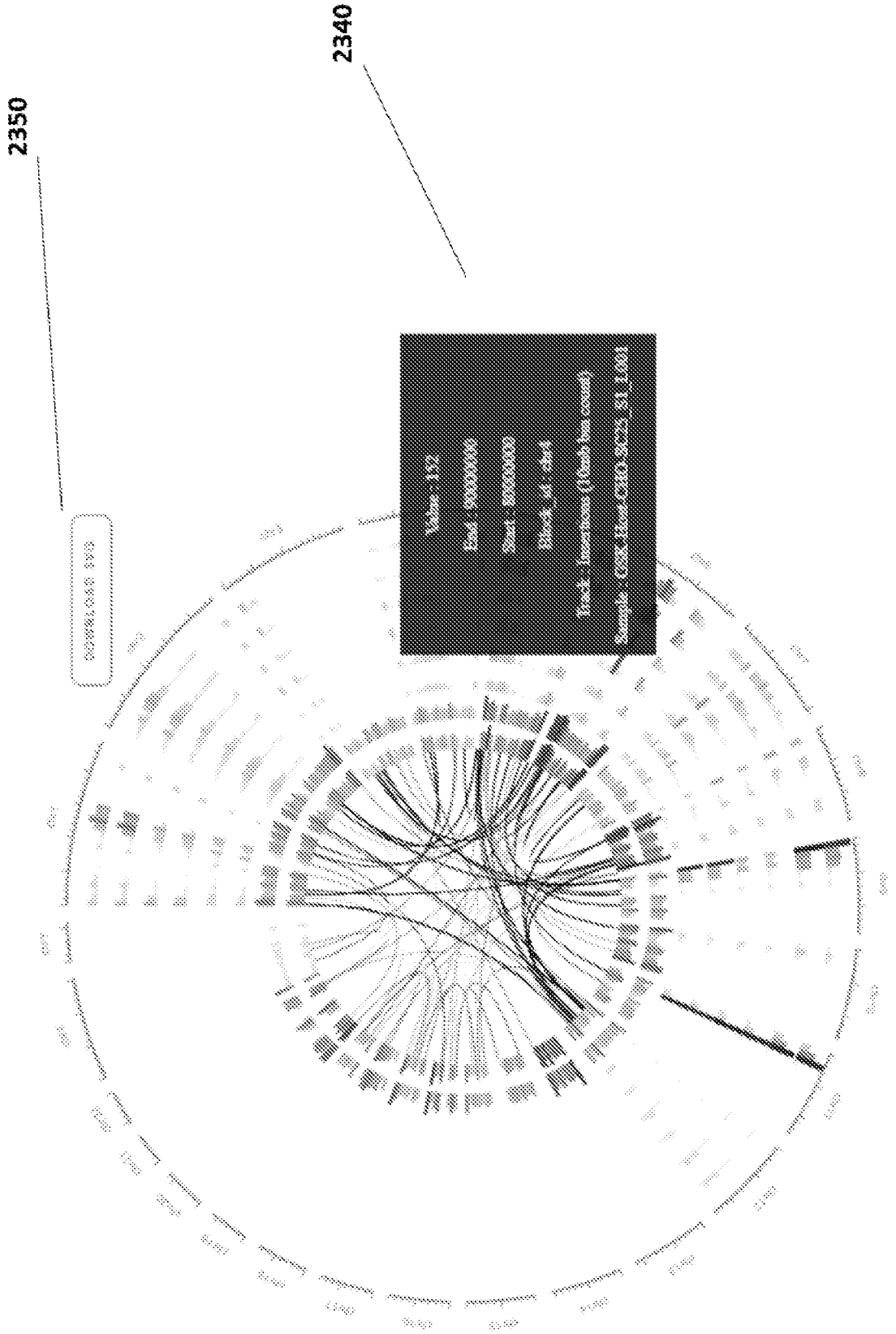


FIG. 23B